Article

Indistinguishability and Negative Probabilities

J. Acacio de Barros [1] **and Federico Holik** [2]

[1] School of Humanities and Liberal Studies, San Francisco State University, San Francisco, CA, USA; barros@sfsu.edu
[2] Universidad de La Plata/CONICET, La Plata, Argentina; olentiev2@gmail.com
* Correspondence: barros@sfsu.edu; Tel.: +1-415-405-2674 (J.A.B.)

Version January 31, 2021 submitted to Entropy

Abstract: In this paper, we examined the connection between quantum systems' indistinguishability and signed (or negative) probabilities. We do so by first introducing a measure-theoretic definition of signed probabilities inspired by research in quantum contextuality. We then argue that ontological indistinguishability leads to the no-signaling condition and negative probabilities.

Keywords: indistinguishability; quantum ontology; negative probabilities; signed measure; quasi-set theory; contextuality

1. Introduction

The assignment of truth values to propositions asserting that a system's property has a definite value is problematic in quantum mechanics. Take the case of propositions about momentum and position for a quantum system. Heisenberg's uncertainty principle asserts that we cannot know the values of position and momentum simultaneously, at least not as precisely as one wants. This constraint brings the issue of whether systems have well-defined but unknowable values of position and momentum, or whether these are undefined. If the former, the probabilistic uncertainties appearing in quantum theory would have an epistemic character, being quantum properties the best description of what we can say about the system. If the latter, then what properties does the system have? For instance, when we measure a particle's momentum and find the value p, does it mean the particle[1] has momentum p? Moreover, is this value of momentum something that existed before the measurement? If not, then do measurements create properties? Do the experimenter, who chooses what to measure, set what properties a particle has? These questions become more problematic if we consider the Kochen-Specker theorem.

In their seminal paper, Kochen and Specker (KS) studied hidden-variable theories compatible with the quantum formalism and satisfying certain physically-motivated conditions. They proved that the values that these hidden variable theories assign to propositions about quantum systems must be contextual: the truth-value assigned to a given proposition will depend on the context in which it is considered. The idea for their proof is the following (see Section 2 for detail). Imagine we have a set of N binary observables $\mathcal{P} = \{P_1, P_2, ..., P_N\}$ corresponding to yes-no questions about a quantum particle. Each P_i is a Hermitian projection operator in a Hilbert space (in KS's paper a three dimensional one). As is well known, each P_i is associated with a proposition about the quantum system. KS constructed a set of such operators with the following characteristics. First, there were several subsets of three commuting operators, such that one and only one of them were true for this

---

[1] It is important to stress here that in this work, the word "particle" is used only for the sake of clarity of exposition, and that we are not compromised, in principle, with a particle ontology. Our considerations apply to whatever kind of entities the quantum systems might be (such as particles or fields, for example), provided they are indistinguishable.

set (i.e., they were orthogonal, and their sum was one). We can think of these subsets as a context, determined by the set of simultaneous propositions considered. These subsets had the additional feature that each $P_i \in \mathcal{P}$ appeared twice, one time for each of two possible contexts. By constructing an appropriate set $\mathcal{P}$, KS showed that the structure of quantum observables and their corresponding contexts did not allow the consistent assignment of truth values for each P_i that was the same for *all contexts*. Thus, in this sense, only *contextual* hidden variable theories are compatible with the quantum formalism. Furthermore, this contextuality exists for all quantum systems that are complex enough (more specifically, it holds for any Hilbert space of dimension greater than two).

Further study in hidden variable models led to the discovery of the so-called non-contextuality inequalities. These can be experimentally testable, opening an obvious field of research for discarding theories that deviate from experiments (and quantum theory). Examples of them are the KCBS inequalities in [1] and the GHZ inequalities in [2]. It was later shown that Bell and CHSH inequalities fall into this category. These inequalities' characteristic feature is that they put an upper bound on the correlations that a family of non-contextual hidden variable theories can model. Thus, an approach is non-contextual if the correlations predicted by it satisfy a specific bound. Since the correlations predicted by quantum theory do violate those inequalities, it is natural (and tempting) to say that quantum mechanics is contextual. Notice that this is a shift from the old quantum physics jargon, for which only hidden-variable theories could be considered as contextual or not.

Furthermore, in the last decades, this quantum theory feature has attracted a lot of interest due to its potential role in quantum information processing tasks. Thus, instead of being considered a negative characteristic, nowadays, physicists seeking to develop quantum technologies, consider contextuality a positive feature of quantum theory itself, which can be quantified, measured, and used as a resource. In this work, we will follow the current jargon, and refer to the feature of the quantum formalism discovered by Kochen and Specker as *quantum contextuality*. In other words, we will use expressions such as "quantum mechanics is contextual," "this theory (or state) has such amount of contextuality," and so on, to simply mean that outcomes of experiments are contextual.

There is yet another –less explored– feature of quantum mechanics that justifies the modern jargon. Propositions about quantum systems are linked to concrete experimental settings, which are selected by the experimenter. If we prepare a quantum system in a particular state and consider a proposition in a given context, we find empirically that the result of an experiment might not be the same should we repeat the test with the same state, but with the given proposition considered in a different context. This is phenomenologically given, and it is independent of any interpretation. Furthermore, one might avoid speaking about states at all, and only refer to preparations and testable quantities of physical systems and their correlations in a theory-independent way; still, it would be meaningful to determine whether experiments display contextuality or not, and this could be checked by observing probability distributions and non-contextuality inequalities objectively. If a system shows contextual correlations, we refer to this feature by saying that the system is *empirically contextual*. This notion of empirical contextuality is consistently defined, objectively testable, and it is model-independent (in the sense that they only assume very general features of probabilistic models).

Because of contextuality, one cannot represent quantum states with classical probabilities. Usually, one represents them by trace operators acting on a separable Hilbert space. But it seems possible to describe quantum states with extended probabilities. For example, the Wigner function takes a quantum state and transforms it into a classical phase space function. This function resembles a Kolmogorovian probability, but it may take negative values. Because it may be negative, it is considered a *quasi-probabilities*. Most approaches to quasi-probabilities rely on an underlying theory (such as quantum mechanics) whose states and observables are mapped to a classical phase space in which the states take the form of quasi-probabilities (see for example [3]).

In this work, we take an alternative approach and focus on two aspects of quantum contextuality. First, we rely on the notions of signed measurable space and measurement context to give a formal definition of negative probabilities that is general enough to cover all cases of interest in quantum contextuality (and hopefully also outside of physics). Classical probabilistic models are shown to be particular cases of our formulation, which is general enough to include contextual models, such as those coming from the quantum formalism. The approach presented here has many features in common with previous ones (see, for example, [4–6]). Still, it relies more directly upon the notions of compatible random variables (for which a joint probability distribution exists), and thus, it provides a straightforward extension of Kolmogorov's approach. Our signed probabilities are constructed as no-signaling, meaning that the quasi-probability distribution associated with a random variable is context-independent. This particular feature is particularly relevant in physics, given that all physical theories satisfy this condition.

The other focus of this article is on quantum indistinguishability. In previous works, we have discussed the connection between particle and property indistinguishability as related to contexts [7]. Here we show that property indistinguishability leads to the no-signaling condition. Since negative probabilities are necessary and sufficient for the description of no-signaling models, we argue that there is a connection between the principle of particle indistinguishability and negative probabilities. The assumption of indistinguishability for quantum particles leads to contextual and indistinguishable properties, which can, in turn, be naturally modeled using our definition of signed probabilities.

We organize this paper as follows. After reviewing elementary facts about contextuality in Section 2, in Section 3 we motivate and provide our definition of signed probabilities. In Section 4, we discuss the connection between quantum indistinguishability, negative probabilities, and the non-signaling condition. Finally, in Section 5, we end with some final remarks and conclusions.

2. Contextuality in Quantum Mechanics

Context is a term that comes from linguistics, especially from semantics and pragmatics [8]. For instance, in semantics, the truth-value of an utterance or written text may depend on the other statements or sentences that precede or follow it. Take the written sentence: "Alice sat by the bank to observe the people." Its truth-value varies depending on other comments that accompanied it: if it were preceded by "The river was calming and beautiful," its meaning would differ from if it were preceded by "The heist needed planning." For the case where "river" preceded the sentence, "bank" likely refers to the bank side of a river, whereas for the "heist" case, "bank" refers to a financial institution. Though this is a case where meaning changes, there are other examples in linguistics where meaning does not change, but truth-value does. We can think of those as examples of context-dependency, or contextuality, in linguistics [9].

Contextuality, as conceptually discussed above, is a central concept in the foundations of quantum mechanics. It is also the main driving difficulty in defining properties for quantum particles or systems. So, let us examine how contextuality appears in quantum mechanics by discussing the famous Kochen-Specker theorem [10]. Here we present a more straightforward proof involving only nine contexts [11].

We start with a four-dimensional Hilbert space, $\mathcal{H}$. According to the standard formalism of quantum mechanics, measurable properties are represented by Hermitian operators in $\mathcal{H}$ (known as observables). A quantum system is said to have a property if an experiment measuring it yields the same value all the time. In the formalism, this translates into having the system be in an eigenstate of the Hermitian operator. A particularly important subset of observables is projection operators, which correspond to 0- or 1-valued observables. We can think of these binary properties as truth-values: either the quantum system has the property (1), or it does not (0). To distinguish between general properties and those associated with projection operators, we call the latter testable propositions, or, in short, propositions. The distinction between testable propositions and properties is subtle and debated in the literature (see, e.g., [12,13]). Here we use the terminology that propositions are a particular type of observables, as discussed above.

A vector in $\mathcal{H}$ uniquely determines a projection operator. For example, the vector $|1,0,0,0\rangle \in \mathcal{H}$ corresponding to the column matrix with the first component as one and the others as zero determines the projector operator $\hat{P}_{1,0,0,0} \equiv |1,0,0,0\rangle\langle 1,0,0,0|$. Let us consider now the following set of equations.

$$\hat{P}_{0,0,0,1} + \hat{P}_{0,0,1,0} + \hat{P}_{1,1,0,0} + \hat{P}_{1,-1,0,0} = 1, \tag{1}$$
$$\hat{P}_{0,0,0,1} + \hat{P}_{0,1,0,0} + \hat{P}_{1,0,1,0} + \hat{P}_{1,0,-1,0} = 1, \tag{2}$$
$$\hat{P}_{1,-1,1,-1} + \hat{P}_{1,-1,-1,1} + \hat{P}_{1,1,0,0} + \hat{P}_{0,0,1,1} = 1, \tag{3}$$
$$\hat{P}_{1,-1,1,-1} + \hat{P}_{1,1,1,1} + \hat{P}_{1,0,-1,0} + \hat{P}_{0,1,0,-1} = 1, \tag{4}$$
$$\hat{P}_{0,0,1,0} + \hat{P}_{0,1,0,0} + \hat{P}_{1,0,0,1} + \hat{P}_{1,0,0,-1} = 1, \tag{5}$$
$$\hat{P}_{1,-1,-1,1} + \hat{P}_{1,1,1,1} + \hat{P}_{1,0,0,-1} + \hat{P}_{0,1,-1,0} = 1, \tag{6}$$
$$\hat{P}_{1,1,-1,1} + \hat{P}_{1,1,1,-1} + \hat{P}_{1,-1,0,0} + \hat{P}_{0,0,1,1} = 1, \tag{7}$$
$$\hat{P}_{1,1,-1,1} + \hat{P}_{-1,1,1,1} + \hat{P}_{1,0,1,0} + \hat{P}_{0,1,0,-1} = 1, \tag{8}$$
$$\hat{P}_{1,1,1,-1} + \hat{P}_{-1,1,1,1} + \hat{P}_{1,0,0,1} + \hat{P}_{0,1,-1,0} = 1. \tag{9}$$

Each equation above is numerically equal to one because all the vectors in each line form a complete and orthonormal basis for $\mathcal{H}$. This means that, for each equation (1)–(9), we have four true-false properties that are compatible, complete, and mutually exclusive. Therefore exactly one of them must be true, and the others zero, which means they all add to one.

An issue may be evident to some readers about (1)–(9): if we assign to each property a truth-value of zero or one we reach a contradiction. To see this contradiction, consider that each property $\hat{P}_i$ appears on the left hand side of (1)–(9) twice. Since $2\hat{P}_i$ is an even number, it follows that the sum of all the terms on the left-hand side of (1)–(9) must be even. However, we add the right-hand side of (1)–(9) we total nine, clearly not an even number, which is a mathematical contradiction.

The mathematical contradiction is a result of assuming that the truth-value of a property $\hat{P}_i$ is the same when it is co-measured with different properties. For example, $\hat{P}_{0,0,0,1}$ shows up in (1) but also in (2). However, the co-measured variables to $\hat{P}_{0,0,0,1}$ in (1) are all different from the ones in (2). In the example above, therefore, we have nine contexts, and each property shows up in exactly two of those contexts. If we allow, for example, $\hat{P}_{0,0,0,1}$ to have a different truth-value when co-measured with $\hat{P}_{0,0,1,0}$, $\hat{P}_{1,1,0,0}$, and $\hat{P}_{1,-1,0,0}$ (call it Context 1) from when it is co-measured with $\hat{P}_{0,0,0,1}$, $\hat{P}_{0,1,0,0}$, $\hat{P}_{1,0,1,0}$, and $\hat{P}_{1,0,-1,0}$ (Context 2), we reach no contradiction. It is in this sense that contextuality is claimed for quantum observables: the truth-value of a property varies with its context determined by the collection of co-measured properties.

The above example has some intriguing features. First, it is state-independent. This feature means that it does not matter how we prepare the quantum system; if we try to measure the properties on (1)–(9), they will change from context to context. Therefore contextuality is a property of the quantum-operator algebra. Second, what the KS theorem shows is a *logical contradiction* that arises from a context-independence assumption. This means that we do not need to involve probabilities in proving the contextuality of quantum properties.

However, probabilities are a fundamental aspect of quantum theory, and perhaps of any empirical theory. So, how could we formulate the KS theorem in terms of probability theory? The hint can be found on [14]: logical inconsistencies are but a special case of probability one events when a joint probability distribution does not exist that describes the outcomes of the experiments. To see this, let us consider the example of four two-valued properties, A, A', B, and B', who can only be observed in the following pairwise experimental arrangements: A with B; A with B'; A' with B; and A' with B'. If we assume that those properties are context-independent, then the combination of their values defined by

$$S = AB + AB' + A'B - A'B' \tag{10}$$

is always a number equal or less than two. The reader can verify the previous statement for all possible combinations, but as an example, if $A = 1$, $A' = -1$, $B = 1$, and $B' = 1$, $S = 1 + 1 - 1 + 1 = 2$. Since any combination of A, A', B, and B' yields a value of S that is 2 or less, it follows that convex combinations of S imply that

$$\langle S \rangle \leq 2, \tag{11}$$

where we are using the fact that the mean value of S, denoted $\langle S \rangle$, is a convex combination of each of its possible values. It follows, from (11) that if $S > 2$, there is no convex combination of the *logical* context-independent possibilities that yields the expected value of S. In other words, it is not possible to assign probabilities to the possible combinations of values of A, A', B, and B' consistent with $\langle S \rangle > 2$. This is why a joint probability distribution for A, A', B, and B' does not exist, although, of course, marginal probabilities do, since we can use the data tables to, say, compute the value of $\langle AB \rangle$.

We should point out that (11) is one of the CHSH inequalities [15]. By itself, as we saw above, a violation of (11) is sufficient to establish the non-existence of a joint probability distribution or contextuality for the observables in question. However, other inequalities need to be added to (11) to form a set of necessary and sufficient conditions for the contextuality of properties.

The CHSH inequalities [15] are related to Bell's inequalities [16], and they can be used to show that quantum mechanics is a non-locally contextual theory, or simply non-local. This is done by starting with two spin-1/2 particles, A and B, in an entangled state

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|+-\rangle - |-+\rangle \right), \tag{12}$$

where $|+-\rangle$ is the state where particle A has spin $+1/2$ and B spin $-1/2$ and $|-+\rangle$ the other way around. It is easy to prove from (12) that the joint expectation of two spin measurements in directions θ_1 for A and θ_2 for B yield the following correlation:

$$E(\theta_1, \theta_2) = \sin(\theta_1 - \theta_2). \tag{13}$$

The reader can verify that for the combinations of measuring the spin of A at 0°and 45° and B at 22.5° and 67.5°, $\langle E \rangle = 2\sqrt{2} > 2$, which violates (11). So, quantum mechanics is not only contextual, but its contextuality manifests for observers that may be far apart from each other, such as the case of the two-particle example above. Contextuality appears in quantum mechanics from the structure of the Hilbert space and that it is present even for systems whose properties are space-like separated. This contextuality presents difficulties to the concept of property in quantum mechanics, as they would depend on the experimenter's choice of a measurement apparatus, as discussed above.

To summarize, in this section, we discussed the idea of contextuality both from an intuitive and formal perspective. We saw that contextuality is the impossibility of consistently assigning truth-values to the same testable proposition in different contexts. Equivalently, a similar assertion holds for observables: it is impossible to assign non-contextual values to all possible observables if some minimal functionality conditions are to be considered [10]. Alternatively, one can interpret contextuality as the proposition (or observable) changing from one context to another. These observations lead to a subtle (but fundamental) problem: do propositions (or observables) retain their identity when considered in different contexts? Let us be more explicit about this. In the scenario described above, consider the contexts AB and AB'. What is the status of observable A in contexts AB and AB'? Let us denote A_B and $A_{B'}$ to the observable A considered in contexts AB and AB', respectively. Usually, since quantum systems obey the no-signal condition, physicists tend to identify A_B and $A_{B'}$ (i.e., $A_B = A_{B'}$). However, this assumption is not trivial at all and has indeed been criticized. In some fields of research, A_B and $A_{B'}$ may not have the same distribution (as is the case in signaling theories) and, even if they have the same content, it should be dubious to identify them. Some authors have proposed that A_B and $A_{B'}$ should be considered different whenever a system manifests a strong degree of contextuality [17,18]. In previous works [7,19], we have proposed an alternative solution to the dichotomy $A_B = A_{B'}$ vs $A_B \neq A_{B'}$. Using a formal framework that allows dealing with collections of indistinguishable objects (see section 4 of this work), we have proposed that A_B and $A_{B'}$ can be thought of as indistinguishable (denoted by $A_B \equiv A_{B'}$). This point of view allows us to connect with contextuality one of the most fundamental features of quantum theory: quantum systems of the same kind are indistinguishable. More specifically, we show in [7,19] that the indistinguishability of particles leads to the indistinguishability of propositions and that this, in turn, gives place to contextuality. In the rest of this work, we elaborate on these ideas further and show a strong connection between the indistinguishability of testable propositions (or observables) and negative probabilities. To do this, we must first introduce a definition of negative probabilities that is useful for our purposes and general enough to cover all physical models of interest.

3. Negative Probabilities

Negative Probabilities (NP) have a long tradition in physics and find applications in different branches of quantum physics [20]. NP appeared in physics early in the 20th century in quantum mechanics, for example, in connection to the Klein-Gordon equation or Wigner's paper on the classical approximations for quantum statistical mechanics [21]. However, NP were considered an undesirable side effect of a defective model or theory. As such, theories yielding NP were discarded as having no physical interest. The first physicist to take NP seriously was Dirac, who used them as the basis for his interpretation of the theory of photons [22]. They also were discussed by Feynman, who thought they were a promising concept but could not find any use for them [23]. Nevertheless, their study helped understand the connection and differences between quantum and classical systems. In some fields –as is the case in quantum optics– they have even become a tool of everyday use [24]. Furthermore, they form the basis of many contextuality measures [25,26] and serve to characterize quantumness of states and theories [27]. Recent studies aim to understand the differences between the correlations originated in quantum theory and those that come from other plausible no-signaling generalized probabilistic models [28]. In this setting, negative probabilities are used to characterize different features of quantum mechanics [3,29]. Nowadays, NP have become a fundamental tool in quantum information theory and the development of quantum technologies. In particular, they play a significant role in the problem of quantum state estimation [30], the determination of quantum correlations and classicality of quantum states [31], and the study of quantum computers' speed-up [32,33].

In our discussion of NP, let us start with Wigner's work. In his 1932 paper [21], Wigner asked the following question: if we have an ensemble of N classical particles, what types of corrections would we have to introduce to their phase-space probability distributions such that their statistics coincided with the quantum one. For this purpose, he constructed what is now known as the Wigner distribution, given by

$$W(\mathbf{r},\mathbf{p}) = \frac{1}{(2\pi)^3}\int \psi^*\left(\mathbf{r}+\frac{\hbar}{2}\mathbf{s}\right)\psi\left(\mathbf{r}-\frac{\hbar}{2}\mathbf{s}\right)e^{i\mathbf{p}\cdot\mathbf{s}}d^3\mathbf{s}, \tag{14}$$

where $\mathbf{r}$ and $\mathbf{p}$ are the position and momentum, and $\mathbf{s}$ is an integration variable. A similar definition holds for arbitrary pairs of conjugate variables. It is easy to see that W behaves similarly to a joint probability distribution, in the sense that if we integrate W on either $\mathbf{r}$ or $\mathbf{p}$ we get the marginal probability distributions. For example,

$$\int W(\mathbf{r},\mathbf{p})d^3\mathbf{p} = |\psi(\mathbf{r})|^2. \tag{15}$$

However, as Wigner pointed out, W is not a proper joint probability distribution, as it can take negative values. For example, for the ground state of the harmonic oscillator, W is non-negative, but for the first excited state, it is negative in some regions of the phase space [34]. After Wigner, Dirac [35] used negative probabilities to try to solve the problem of infinities in quantum field theory. In his theory, negative probabilities were nothing more than an accounting tool for computing (non-negative) observable probabilities, and carried the same interpretation as the statement "having negative three apples.” This was similar to the interpretation suggested by Feynman in his article on negative probabilities [36]. For a review of the history of negative probabilities in physics, the interested reader is referred to [37]. More recently, negative probabilities have been used in foundations of quantum mechanics, and the interested reader is referred to references [6,38,39] and references therein. For possible interpretations of negative probabilities that are not based on a pragmatic bookkeeping, readers are referred to [5,40–43].

What are negative probabilities? Let us start with the standard probability theory. The currently accepted axioms for probability were laid down by Kolmogorov [44]. In his axioms, we start with a sample set Ω, which we can think of as possible states of the system of interest. For example, if we are interested in a die's outcomes, Ω could be the set $\{1,2,3,4,5,6\}$. We could, in principle, talk about the probabilities of the members of Ω. Still, Kolmogorov recognized that, in probability theory, we want to refer to logical combinations of possible states. To do so, he associated with Ω a σ-algebra $\mathcal{F}$ of its elements. Once we have Ω and $\mathcal{F}$, he define the probability p as a non-negative real-valued function $p : \mathcal{F} \rightarrow [0,1]$ satisfying the following properties.

K1. $p(\Omega) = 1$
K2. For every denumerable and disjoint family $\{A_i\}_{i\in\mathbb{N}}$, $p(\bigcup A_i) = \sum_i p(A_i)$.

It is easy to see, for simple examples, that Kolmogorov's definition captures the essence of probabilities first put forth by Pascal and then developed throughout the centuries (for a wonderful historical account of probability theory, see [45].).

However, as we saw in Section 2, it is not always possible to have a joint probability distribution that accounts for all experimental outcomes. There are different ways to approach this lack of a joint. One possibility is to notice that the algebra of observables is not Boolean, but follows a lattice structure that does not allow for certain Boolean operations (for example, the complement of a property may not exist) [46]. This is the quantum logic approach, and one could try to create a probability calculus over lattices, and not Boolean algebras. Of course, one such probability calculus is the Hilbert space formalism. Another approach could be to modify Kolmogorov's definition to allow for a new probability function, say p^*, to exist. For example, we could change K2 from an equality to an inequality, as is the case for upper and lower probabilities [47–49]. Another possibility is to keep the algebra intact, as well as K1 and K2, but change the requirement that p is non-negative, i.e., to allow for negative probabilities.

What are the axioms for negative probabilities? To give a straightforward description based on measure theory (obtaining thus a canonical generalization of Kolmogorov's approach), we rely on the notion of compatible random variables and signed measure spaces. In the rest of this section, we will try to motivate and write down a definition for negative probabilities in the spirit of Kolmogorov.

Let us start with a definition of random variables.

Definition 1. *Let $(\Omega, \mathcal{F}, p)$ be a probability space, and let $(M, \mathcal{M})$ be a Borel space with elements of M being real numbers, i.e. $\mathcal{M}$ is a σ-algebra over M. A (real-valued)* random variable $\mathbf{R}$ *is a measurable function $\mathbf{R} : \Omega \rightarrow M$, i.e. for all $m \in \mathcal{M}$, $\mathbf{R}^{-1}(m) \in \mathcal{F}$.*

Though the above definition may seem complicated, it is intuitive. What it says is that we can associate to partitions of the sample space Ω a particular real number. A simple example is the game of craps. Imagine we throw two dice and record their outcomes. A sample space for this example is $(1,1), (1,2), \ldots, (6,6)$, where each ordered pair corresponds to an outcome for each die. In a game of craps, often, what matters is the sum of the values and not the individual outcomes. For example, rolling a *seven out*, a sometimes desired outcome, is the result of one of the following outcomes: $(1,6)$, $(2,5)$, $(3,4)$, $(4,3)$, $(5,2)$, or $(6,1)$. A random variable yielding the sum of the thrown dice would associate to all those outcomes the value 7. As defined, random variables are a way to model outcomes of experiments or observations that are stochastic, i.e., that have certain randomness associated with them.

If we look back at our examples in Section 2, we can see that random variables may express contextuality. For example, let us consider the four two-valued properties A, A', B, and B'. Since they could be used to describe yes/no properties, let us think of each of them as a ± 1-valued random variables in a given a probability space $(\Omega, \mathcal{F}, p)$, e.g. $\mathbf{A} : \Omega \rightarrow 1, -1$. In terms of random variables, (10) would be rewritten simply as

$$\mathbf{S} = \mathbf{AB} + \mathbf{AB}' + \mathbf{A}'\mathbf{B} - \mathbf{A}'\mathbf{B}'. \tag{16}$$

Since it follows from standard probability theory that

$$\langle \mathbf{S} \rangle = \langle \mathbf{AB} \rangle + \langle \mathbf{AB}' \rangle + \langle \mathbf{A}'\mathbf{B} \rangle - \langle \mathbf{A}'\mathbf{B}' \rangle \leq 2, \tag{17}$$

any violation of this inequality would imply that no (standard) probability space exists that allow for the correlations observed in those random variables. Equation (17) is one of the well-known CHSH inequalities, which are necessary and sufficient conditions for the existence of a joint probability distribution [15,50]. However, for this example, it is trivial to construct four different probability spaces for each experimental situation, i.e. A and B, A' and B, A and B', and A' and B'. The impossibility is to find a single probability space that yields all four correlations that are experimentally observed in quantum theory. And this is how random variables can help us define negative probabilities. We can relax the non-negativity assumption as long as we guarantee that all observable properties do not result in negative probabilities[2]. This motivates the following definitions.

Definition 2. *Let Ω be a sample space and $\mathcal{F}$ a σ-algebra over Ω. A* signed measure *is a function $\mu : \mathcal{F} \rightarrow \mathbb{R}$ such that*

$$\mu(\emptyset) = 0 \tag{18}$$

and for every denumerable and disjoint family $\{A_i\}_{i \in \mathbb{N}}$

$$\mu(\bigcup_i A_i) = \sum_i \mu(A_i) \tag{19}$$

The triple $(\Omega, \mathcal{F}, \mu)$ is called a signed measure space *[53].*

Signed measure spaces expand the idea of measures (not probabilities), to the negative domain. However, it should be clear to the reader that signed measures are a generalization of probability measures, one we will use to define negative probabilities.

Definition 3. *Let $(\Omega, \mathcal{F}, \mu)$ be a signed measure space, and let $(M, \mathcal{M})$ be a Borel space with elements of M being real numbers, i.e. $\mathcal{M}$ is a σ-algebra over M. A (real-valued)* extended random variable $\mathbf{R}^*$ *is a measurable function $\mathbf{R}^* : \Omega \rightarrow M$, i.e. for all $m \in \mathcal{M}$, $(\mathbf{R}^*)^{-1}(m) \in \mathcal{F}$.*

Notice that extended random variables are not at all equivalent to random variables, except in special cases when μ is a probability measure.

Definition 4. *Let $\{R_i^*\}$, $i = 1, \ldots, n$, be a collection of extended random variables defined on a signed measure space $(\Omega, \mathcal{F}, \mu)$. A* μ-induced context *is a subset $C_j^\mu = \{R_k^*\}_{k \in N_j}$, $N_j \subset \{1, \ldots, n\}$, for which there exists a sub-σ-algebra $\mathcal{F}_j$ of $\mathcal{F}$ such that, by defining $p_j^\mu(F) := \mu(F)$ for all $F \in \mathcal{F}_j$, the triad $(\Omega, \mathcal{F}_j, p_j^\mu)$ becomes a probability space, and $R_{i_k}^*$ is a random variable with respect to it, for all $k \in \{1, ..., n_j\}$.*

---

[2] We point out that in the context of weak measures, negative probabilities may be "measurable," but we will not discuss this issue or its meanings here. Instead, we refer the interested reader to references [51,52].

Some observations are in order. First, the notion of context given by Definition 4 depends on the chosen measure μ. Since we are grounding our definitions on measure theory, the available mathematical tools are a set Ω, a collection $\mathcal{F}$ of subsets of it (forming a Boolean algebra), and a signed measure μ. The dependence on μ makes our definition of context *measure dependent*. We aim to represent each possible state of the system under study by a normalized signed measure. A concrete probabilistic model for a system is determined when all its possible states are specified. Once this is done, the contexts of the theory can be unambiguously determined as follows. We denote by $\mathcal{S}$ to the collection of all possible states of a system, described as signed measurable spaces. In order to obtain a consistent theory (such as a classical or quantum probability theory), we assume that all states have associated the same outcome set Ω and the same σ-algebra $\mathcal{F}$ and that they are normalized. It is useful to put this in terms of a definition.

Definition 5. *Let Ω be a set and $\mathcal{F}$ a σ-algebra of subsets of Ω. A* family of signed probabilistic models *for $(\Omega, \mathcal{F})$ is a collection $\mathcal{S}_{(\Omega,\mathcal{F})}$ of signed measures on $(\Omega, \mathcal{F})$ such that, for all $\mu \in \mathcal{S}_{(\Omega,\mathcal{F})}$, $\mu(\Omega) = 1$. Any $\mu \in \mathcal{S}_{(\Omega,\mathcal{F})}$ is called a* state of the model.

The above definition is analogous to that of states in a classical probabilistic model, the sole difference being that we allow the states to take negative values. In order to describe the observables of physical theories, we need each extended random variable to be consistently defined with regard to all possible states $\mathcal{S}_{(\Omega,\mathcal{F})}$ in the following sense. Considered as a function $R_i^* : \Omega \longrightarrow \mathbb{R}$, we must have that each extended random variable must satisfy $(R_i^*)^{-1}(\Delta) \in \mathcal{F}$, for every Borel set $\Delta \subseteq \mathbb{R}$ (this means that the R_i^*'s are measurable functions with regard to all possible $\mu \in \mathcal{S}_{(\Omega,\mathcal{F})}$). This condition grants that the extended random variables are well defined for all $\mu \in \mathcal{S}_{(\Omega,\mathcal{F})}$. With these definitions, we are ready to provide a state-independent definition of context.

Definition 6. *Consider a family of signed probability models $\mathcal{S}_{(\Omega,\mathcal{F})}$. Let $\{R_i^*\}$, $i = 1, \ldots, n$, be a collection of extended random variables defined on $\mathcal{S}_{(\Omega,\mathcal{F})}$. A* general context *is a subset $C_j = \{R_k^*\}_{k \in N_j}$, $N_j \subset \{1, \ldots, n\}$ of those extended random variables, for which there exists a sub-σ-algebra $\mathcal{F}_j$ of $\mathcal{F}$ satisfying that, for all $\mu \in \mathcal{S}$, by defining $p_j^{\mu}(F) := \mu(F)$ for all $F \in \mathcal{F}_j$, the triad $(\Omega, \mathcal{F}_j, p_j^{\mu})$ becomes a probability space, and $R_{i_k}^*$ is a random variable with respect to it, for all $k \in \{1, ..., n_j\}$.*

Using the definition of general context, we can naturally introduce the notion of *signed probability space* as follows.

Definition 7. *A* signed probability space*, also called here* negative probability space*, is a signed measure space $(\Omega, \mathcal{F}, \mu)$ endowed with a non-empty set of contexts $C = \{C_j^{\mu}\}$ (in the sense of Definition 4), such that $\mu(\Omega) = 1$. The measure μ in this space is a* signed probability *or* negative probability.

In other words, a signed probability space is a signed measure space for which there exist contexts, and these contexts give place to well defined probabilistic scenarios.

Proposition 1. *If a state $\mu \in \mathcal{S}_{(\Omega,\mathcal{F})}$ of an extended probabilistic model admits a non-empty set of contexts, then, it defines a signed probability space.*

Proof. If $\mu \in \mathcal{S}_{(\Omega,\mathcal{F})}$ is a state, then, μ is a signed measure on $(\Omega, \mathcal{F})$ such that $\mu(\Omega) = 1$. Thus, the existence of a non empty family of contexts for $(\Omega, \mathcal{F}, \mu)$, makes it satisfy Definition 7. □

After the above Definitions, it is important to make the following remarks.

Proposition 2. *If $(\Omega, \mathcal{F}, p)$ is a probability space, then it is also a signed probability space.*

Proof. Any $(\Omega, \mathcal{F}, p)$ satisfying Kolmogorov's axioms also satisfies the axioms of signed measure in Definition 2. Given that p is normalized, it is also a state with respect to the pair $(\Omega, \mathcal{F})$. Any collection of random variables defined on $(\Omega, \mathcal{F}, p)$, induces a context satisfying Definition 4 (by taking sub-σ-algebra as $\mathcal{F}$ itself). Thus, the states of classical probabilistic systems can be described as a particular case of signed probabilities. □

The states of the extended probability model of quantum theory are just the quantum states' images under the Wigner transform. Any context of a quantum system – understood in the usual sense of a family of commuting observables – can be described in our approach by a collection of extended random variables.

Definitions 4, 6, and 7 are inspired in the following properties of the Wigner distribution function. For simplicity, suppose that we have a phase space $\Omega = \{(x, p) \in \mathbb{R} \times \mathbb{R}\} = \Omega_1 \times \Omega_2$ (i.e., we are taking $\Omega_1 = \mathbb{R} = \Omega_2$). Let $\mathcal{F}$ be the collection of Borel subsets of Ω. Then, we have that the quasi-probability of obtaining a system in the set $F \in \mathcal{F}$ is given by $\mu(F) := \int\int_F W(x, p)dxdp$, where $W(x, p)$ is the Wigner distribution function. Indeed, this distribution defines a normalized signed measurable space $(\Omega, \mathcal{F}, \mu)$. To obtain the marginal measures, we must do as follows. Let $\mathcal{F}_1$ be the subalgebra of $\mathcal{F}$ formed by all elements of the form $\Delta \times \Omega_2$, where Δ ranges over any possible Borel set of the real line. Define $W(x) := \int_{\Omega_2} W(x, p)dp$ and $p_1^{\mu}(\Delta \times \Omega_2) := \int_{\Delta} W(x)dx = \int_{\Delta}\int_{\Omega_2} W(x, p)dxdp = \mu(\Delta \times \Omega_2)$. While μ is not in general a positive measure, p_1 always is, and $(\Omega, \mathcal{F}_1, p_1^{\mu})$ is indeed Kolmogorovian. It also coincides numerically with the probabilities for position context computed from the quantum formalism. A similar Kolmogorovian measure $(\Omega, \mathcal{F}_2, p_2^{\mu})$ can be obtained in an analogous way for the momentum context. Further comments are in order:

- Suppose that a random variable belongs to two different general contexts C_i and C_j (according to Definition 6). For each $\mu \in \mathcal{S}$, the condition $p_j^{\mu}(F) := \mu(F)$ in Definition 6 implies that $p_i^{\mu}(F) = \mu(F) = p_j^{\mu}(F)$, for all events F associated to this random variable. In other words, the probability of a proposition is independent of the context in which it is tested. This implies that the probability distribution assigned to an observable will be independent of the other observables with which it is co-measured. This condition is nothing but the generalized version of the *no-signaling condition* in physics (we will further discuss this below). It means that the probability of a given event (or more generally, the probability distribution of a given random variable) will not depend on the context in which it is considered. Thus, according to Definition 6, all negative probabilities that we consider satisfy the no-signaling condition.
- In definition 6, for each μ, all measurable functions defined over the probability space $(\Omega, \mathcal{F}_j, p_j^{\mu})$ define legitimate observables in the classical sense. These observables are all compatible. It is in this sense that the C_j's define contexts. If we mix an observable from context i with other taken from context j, there is no reason to assume that there will exist a joint (Kolmogorovian) probability distribution for them, because μ is not necessarily positive definite. For example, the proposition "the observable f_i (taken from context C_i) possesses its value in the interval $\Delta \in \mathcal{F}_i$ and the observable g_j (from context j) possesses its value in the set $\Gamma \in \mathcal{F}_j$", has a quasi-probability given by $\mu(\Delta \times \Gamma)$. These observables are not necessarily compatible because, by construction, we allow this quantity to be negative. Being negative, this probability cannot be observed in any measurement context.

Each context represents a real empirical scenario, where probabilities and observable quantities are suitably defined. In general, given a set of random variables, it is not necessarily true that a joint probability distribution (understood in the Kolmogorovian sense) exists for all variables. However, for random variables describing physical measurements in different contexts, a negative probability distribution can always be constructed. Definition 7 includes those cases.

A typical practical situation is the following. Suppose that a collection of contexts $\{C_j\}$ is given and that there is more than one signed probability space in which those contexts are defined. Among all possible signed probability spaces compatible with a family of contexts, which one should we chose? To help us understand this question, we should define compatible signed probability spaces.

Definition 8. *A family of signed probability spaces is compatible if their collection of contexts is the same.*

Given a family of contexts $F = \{C_j\}$, call $\mathcal{S}(F)$ the maximal set of compatible signed probability spaces that have F as its collection of contexts. Which signed probability space should we take among all possible in $\mathcal{S}(F)$? The problem of the existence of a "minimal one" is subtle and will be treated elsewhere. Instead, we give here the following definition, which is useful in many circumstances. We also restrict to finite sets in order to simplify the analysis.

Definition 9. *Let* $\mathbf{\Omega}_i = (\Omega_i, \mathcal{F}_i, \mu_i)$*,* $i \in I$*, be a compatible collection of signed probability spaces. For each* $\mathbf{\Omega}_i$*, let* $M_i = \sum_{\omega \in \Omega_i} |\mu_i(\omega)|$*. Then* $\mathbf{\Omega}_k$ *is a* minimal signed (or negative) probability space *if* $M_k = \min\{M_i | i \in I\}$ *when it exists.*

From now on, we will use the notation p^* for negative probabilities, p for regular probabilities, and μ for measures that are not necessarily probabilities (signed or not). With this notation in mind, we can write the following results [6].

Proposition 3. *Let* $\mathbf{\Omega} = (\Omega, \mathcal{F}, p^*)$ *be a minimum signed probability space. If* $M = \sum_{\omega \in \Omega} |p^*(\omega)| = 1$*, then* $\mathbf{\Omega}$ *is also a probability space. Alternatively, if* $\mathbf{\Omega}$ *is a probability space, then it is also a minimum signed probability space, with* $M = 1$*.*

Proof. Since, by Definition 9, we have $\sum_{\omega \in \Omega} p^*(\omega) = 1$, it follows that $\sum_{\omega \in \Omega} |p^*(\omega)| = 1$ implies $p^*(\omega)$ is non-negative for all $\omega \in \Omega$. Given that negative probabilities satisfy all of Kolmogorov's axioms except the non-negativity one, it follows that p^* is a probability, if $M = 1$. Alternatively, for non-negative p^* that add to one, it is immediate that the sum of their absolute value also add to one. See reference [6] for details. □

The above Proposition suggests that the L1 norm plays an essential role in whether a probability distribution exists or not for a set of correlations and random variables. This motivates the following definition.

Definition 10. *Let* $\mathbf{\Omega} = (\Omega, \mathcal{F}, p^*)$ *be a minimal signed probability space. The quantity* δ*, defined as* $\delta = \sum_{\omega \in \Omega} |p^*(\omega)| - 1$ *is called the* contextuality index of $\mathbf{\Omega}$ *or, in short,* contextuality index.

The contextuality index provides a measure of contextuality for a set of experimental outcomes associated to observations of a system. This is at the core of the following proposition, but is also suggested by the previous one.

Proposition 4. *A collection of no-signaling extended random variables on a minimal signed probability space is contextual if and only if the contextuality index* δ *is greater than zero.*

Proof. If we assume that the random variables are contextual, this means that there is no non-negative joint probability distribution that explains all the correlations for the random variables. But since they are no-signaling, from [6] it follows that there is a negative probability consistent with the correlations. Since, by definition, $\sum_{\omega \in \Omega} p^*(\omega) = 1$, and some of the $p^*(\omega) < 0$, it follows that $\sum_{\omega \in \Omega} |p^*(\omega)| > 1$, and therefore $\delta \neq 0$. Also, from the definition of negative probabilities, it follows that δ cannot be less than zero, and we have that $\delta > 0$. Now, let us assume that $\delta > 0$. Since δ is the lowest possible value for the L1 norm minus one, this implies that there is no non-negative joint, which also implies contextuality. For a more detailed proof using a different definition of negative probabilities, see [6]. □

Another straightforward consequence of the definition of negative probabilities is that, for each context C_i, the extended random variables are equivalent to regular random variables. This equivalency should not come as a surprise since, for each context, we have a complete data table involving all possible experimental outcomes. We also point out that if there exists a context C_i such that $\Omega_i = \Omega$, then p^* is a probability.

Let us now examine some examples. Let R_1, R_2, and R_3 be three extended random variables defined over a negative probability space, and assume that $C_1 = (R_1, R_2)$ and $C_2 = (R_1, R_3)$ define two different measurement contexts. Then, it follows from Definition 9 that $p^*(R_1 = \alpha) = \sum_{\beta_i} p^*(R_1 = \alpha | R_2 = \beta_i) p^*(R_2 = \beta_i)$ and $p^*(R_1 = \alpha) = \sum_{\beta_i} p^*(R_1 = \alpha | R_3 = \beta_i) p^*(R_3 = \beta_i)$, where α and β_i are the possible values the random variables can take. In other words, the (pseudo) probability distribution of a random variable defined over a negative probability space cannot depend on whether it is co-observed with one or another random variable [38,39,54]. As remarked above, this property is known in the physics literature as the "no-signaling condition" [55]. Alternatively, if experimental observations of a quantity show its probability distributions as independent of other co-observable variables, then it follows that there always exist a negative probability with extended random variables that model the experimental outcomes. In other words, the existence of extended random variables on a negative probability space is a necessary and sufficient condition for the non-signaling condition to hold [38,39,54].

The equivalence between negative probabilities and non-signaling is one reason why negative probabilities may be a useful tool for exploring the quantum world. Additionally, other properties of quantum systems are well described by negative probabilities. For example, in reference [56], many of the principles attempted to describe quantum mechanics were represented in terms of negative probabilities. It was shown there that negative probabilities provided an elegant and straightforward way to express them.

At this point, it is illustrative to consider the example of two photons, A and B, in the singlet state with z-polarization either ± 1, given by (12). We saw in Section 2 that no probability distribution exists that can account for the quantum correlations, because quantum mechanics violates (11). However, let us see how we can build a negative probability distribution for the above example. First, we point out that for the above case, the smallest Ω we can use, without loss of generality [57], is given by

$$\Omega = \{\omega_{\bar{a}\bar{a}'\bar{b}\bar{b}'}, \omega_{\bar{a}\bar{a}'\bar{b}b'}, \omega_{\bar{a}\bar{a}'b\bar{b}'}, \omega_{\bar{a}\bar{a}'bb'}, \omega_{\bar{a}a'\bar{b}\bar{b}'}, \ldots, \omega_{aa'b\bar{b}'}, \omega_{aa'bb'}\}, \tag{20}$$

where $\omega_{aa'bb'}$ corresponds to the outcome $A = a$, $A' = a'$, $B = b$, and $B' = b'$. It should be clear that Ω generates a σ-algebra $\mathcal{F}$, formed by all its subsets (i.e., $\mathcal{F} = \mathcal{P}(\Omega)$). Accordingly, the random variables can be defined easily from Ω. For example, A would be the random variable defined as the following function.

$$A(\omega) = \begin{cases} +1 & \text{if } \omega \in \{\omega_{a\bar{a}'\bar{b}\bar{b}'}, \omega_{a\bar{a}'\bar{b}b'}, \omega_{a\bar{a}'b\bar{b}'}, \omega_{a\bar{a}'bb'}, \omega_{aa'\bar{b}\bar{b}'}, \omega_{aa'\bar{b}b'}, \omega_{aa'b\bar{b}'}, \omega_{aa'bb'}\} \\ -1 & \text{if } \omega \in \{\omega_{\bar{a}\bar{a}'\bar{b}\bar{b}'}, \omega_{\bar{a}\bar{a}'\bar{b}b'}, \omega_{\bar{a}\bar{a}'b\bar{b}'}, \omega_{\bar{a}\bar{a}'bb'}, \omega_{\bar{a}a'\bar{b}\bar{b}'}, \omega_{a\bar{a}'b\bar{b}'}, \omega_{\bar{a}a'b\bar{b}'}, \omega_{\bar{a}a'bb'}\} \end{cases}. \tag{21}$$

Alternatively, A' is given by

$$A'(\omega) = \begin{cases} +1 & \text{if } \omega \in \{\omega_{\bar{a}a'\bar{b}\bar{b}'}, \omega_{\bar{a}a'\bar{b}b'}, \omega_{\bar{a}a'b\bar{b}'}, \omega_{\bar{a}a'bb'}, \omega_{aa'\bar{b}\bar{b}'}, \omega_{aa'\bar{b}b'}, \omega_{aa'b\bar{b}'}, \omega_{aa'bb'}\} \\ -1 & \text{if } \omega \in \{\omega_{\bar{a}\bar{a}'\bar{b}\bar{b}'}, \omega_{\bar{a}\bar{a}'\bar{b}b'}, \omega_{\bar{a}\bar{a}'b\bar{b}'}, \omega_{\bar{a}\bar{a}'bb'}, \omega_{a\bar{a}'\bar{b}\bar{b}'}, \omega_{a\bar{a}'\bar{b}b'}, \omega_{a\bar{a}'b\bar{b}'}, \omega_{a\bar{a}'bb'}\} \end{cases}, \tag{22}$$

and similarly for B and B'. On the other hand, given that A and B are compatible in the two photons model, there exists a context that contains both. This means that there exists an observable (A, B), that gives the joint outcomes (i, j) $(i, j = \pm 1)$ of performing a simultaneous measure of both A and B. It is defined by

$$(A,B)(\omega) = \begin{cases} (+1,+1) & \text{if } \omega \in \{\omega_{aa'bb'}, \omega_{a\bar{a}'bb'}, \omega_{aa'b\bar{b}'}, \omega_{a\bar{a}'b\bar{b}'}\} \\ (-1,+1) & \text{if } \omega \in \{\omega_{\bar{a}a'bb'}, \omega_{\bar{a}\bar{a}'bb'}, \omega_{\bar{a}a'b\bar{b}'}, \omega_{\bar{a}\bar{a}'b\bar{b}'}\} \\ (+1,-1) & \text{if } \omega \in \{\omega_{aa'\bar{b}b'}, \omega_{aa'\bar{b}\bar{b}'}, \omega_{a\bar{a}'\bar{b}b'}, \omega_{a\bar{a}'\bar{b}\bar{b}'}\} \\ (-1,-1) & \text{if } \omega \in \{\omega_{\bar{a}a'\bar{b}b'}, \omega_{\bar{a}\bar{a}'\bar{b}b'}, \omega_{\bar{a}a'\bar{b}\bar{b}'}, \omega_{\bar{a}\bar{a}'\bar{b}\bar{b}'}\} \end{cases} . \tag{23}$$

Let us see how the context defined by AB defines a probability space, and how this space relates to Ω and $\mathcal{F}$. Notice first that all possible propositions associated to (A,B) (which have the form "A has value i and B has value j", for $i,j = \pm 1$), are represented by the subsets of Ω listed in equation (23). By computing all possible unions, intersections and complements of these subsets, a Boolean subalgebra $\mathcal{F}_{(A,B)}$ of $\mathcal{F}$ is generated. Now, in a two photons state, A and B are of course compatible, and there exists a probability assignment (defined by a quantum state of the compound system) $\mu_{(A,B)}$ such that the triad $(\Omega, \mathcal{F}_{(A,B)}, \mu_{(A,B)})$ is a classical probability space. If we now consider a global probability assignment $(\Omega, \mathcal{F}, \mu)$ (satisfying definition 7), if it is a valid extension, we must have that $\mu(F) = \mu_{(A,B)}(F)$, for all $F \in \mathcal{F}_{(A,B)}$.

Another interesting observable is given by the product of outcomes of A and B. Let us denote it by AB. It is defined by

$$AB(\omega) = \begin{cases} 1 & \text{if } \omega \in \{\omega_{aa'bb'}, \omega_{a\bar{a}'bb'}, \omega_{aa'b\bar{b}'}, \omega_{a\bar{a}'b\bar{b}'}, \omega_{\bar{a}a'\bar{b}b'}, \omega_{\bar{a}\bar{a}'\bar{b}b'}, \omega_{\bar{a}a'\bar{b}\bar{b}'}, \omega_{\bar{a}\bar{a}'\bar{b}\bar{b}'}\} \\ -1 & \text{if } \omega \in \{\omega_{\bar{a}a'bb'}, \omega_{\bar{a}\bar{a}'bb'}, \omega_{\bar{a}a'b\bar{b}'}, \omega_{\bar{a}\bar{a}'b\bar{b}'}, \omega_{aa'\bar{b}b'}, \omega_{aa'\bar{b}\bar{b}'}, \omega_{a\bar{a}'\bar{b}b'}, \omega_{a\bar{a}'\bar{b}\bar{b}'}\} \end{cases} . \tag{24}$$

We obtain again a Boolean subalgebra $\mathcal{F}_{AB}$ of $\mathcal{F}$. Similar constructions can be made for $A'B$, AB', AA', BB', (A,A'), (A,B'), and so on. What are the differences between those observables that mix incompatible observables (such as AA') with respect to those which do not (such as AB)? If we write down the details for AA', we obtain

$$AA'(\omega) = \begin{cases} 1 & \text{if } \omega \in \{\omega_{aa'bb'}, \omega_{aa'b\bar{b}'}, \omega_{aa'\bar{b}b'}, \omega_{aa'\bar{b}\bar{b}'}, \omega_{\bar{a}\bar{a}'\bar{b}b'}, \omega_{\bar{a}\bar{a}'\bar{b}\bar{b}'}, \omega_{\bar{a}\bar{a}'bb'}, \omega_{\bar{a}\bar{a}'b\bar{b}'}\} \\ -1 & \text{if } \omega \in \{\omega_{\bar{a}a'bb'}, \omega_{\bar{a}a'b\bar{b}'}, \omega_{\bar{a}a'\bar{b}b'}, \omega_{\bar{a}a'\bar{b}\bar{b}'}, \omega_{a\bar{a}'\bar{b}b'}, \omega_{a\bar{a}'\bar{b}\bar{b}'}, \omega_{a\bar{a}'bb'}, \omega_{a\bar{a}'b\bar{b}'}\} \end{cases} . \tag{25}$$

We get again a Boolean subalgebra $\mathcal{F}_{AA'}$ for AA'. Notice first that $\mathcal{F}_{AA'} \neq \mathcal{F}_{AB}$. Second, if we want to define probabilities for the outcomes of AA', we have to consider the measures defined by the model we are considering, here a two photons system. In this case, the states are determined by the Born rule. We know that if a collection of observables is commutative, a quantum state assigns them a positive probability. Thus, any legitimate quantum state will assign positive probabilities for all the events in the Boolean algebras $\mathcal{F}_{AB}$, $\mathcal{F}_{AB'}$, $\mathcal{F}_{A'B}$ and $\mathcal{F}_{A'B'}$. What happens with the events in $\mathcal{F}_{AA'}$ and $\mathcal{F}_{BB'}$? The non-negativity of the probabilities assigned by quantum states to the propositions associated with those algebras is no longer granted. This will become clear with the examples discussed in the following Section (see Proposition 7).

Quantum mechanics tells us that, in addition to the correlations in (13), the observable expectations also satisfy the following:

$$\langle A \rangle = \langle A' \rangle = \langle B \rangle = \langle B' \rangle = 0. \tag{26}$$

If we now impose (13) and (26) to the probabilities, from the definition of the random variables set above, we would have at once that the probabilities of ω_i would have to satisfy the following set of linear equations.

$$\begin{aligned} p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} + p_{\bar{a}\bar{a}'\bar{b}b'} + p_{\bar{a}\bar{a}'b\bar{b}'} + p_{\bar{a}\bar{a}'bb'} + p_{\bar{a}a'\bar{b}\bar{b}'} + p_{\bar{a}a'\bar{b}b'} + p_{\bar{a}a'b\bar{b}'} + p_{\bar{a}a'bb'} & \\ +p_{a\bar{a}'\bar{b}\bar{b}'} + p_{a\bar{a}'\bar{b}b'} + p_{a\bar{a}'b\bar{b}'} + p_{a\bar{a}'bb'} + p_{aa'\bar{b}\bar{b}'} + p_{aa'\bar{b}b'} + p_{aa'b\bar{b}'} + p_{aa'bb'} &= 1, \end{aligned} \tag{27}$$

$$\begin{aligned} -p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} - p_{\bar{a}\bar{a}'\bar{b}b'} - p_{\bar{a}\bar{a}'b\bar{b}'} - p_{\bar{a}\bar{a}'bb'} - p_{\bar{a}a'\bar{b}\bar{b}'} - p_{\bar{a}a'\bar{b}b'} - p_{\bar{a}a'b\bar{b}'} - p_{\bar{a}a'bb'} & \\ +p_{a\bar{a}'\bar{b}\bar{b}'} + p_{a\bar{a}'\bar{b}b'} + p_{a\bar{a}'b\bar{b}'} + p_{a\bar{a}'bb'} + p_{aa'\bar{b}\bar{b}'} + p_{aa'\bar{b}b'} + p_{aa'b\bar{b}'} + p_{aa'bb'} &= 0, \end{aligned} \tag{28}$$

$$\begin{aligned} -p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} - p_{\bar{a}\bar{a}'\bar{b}b'} - p_{\bar{a}\bar{a}'b\bar{b}'} - p_{\bar{a}\bar{a}'bb'} + p_{\bar{a}a'\bar{b}\bar{b}'} + p_{\bar{a}a'\bar{b}b'} + p_{\bar{a}a'b\bar{b}'} + p_{\bar{a}a'bb'} & \\ -p_{a\bar{a}'\bar{b}\bar{b}'} - p_{a\bar{a}'\bar{b}b'} - p_{a\bar{a}'b\bar{b}'} - p_{a\bar{a}'bb'} + p_{aa'\bar{b}\bar{b}'} + p_{aa'\bar{b}b'} + p_{aa'b\bar{b}'} + p_{aa'bb'} & = 0, \end{aligned} \tag{29}$$

$$\begin{aligned} -p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} - p_{\bar{a}\bar{a}'\bar{b}b'} + p_{\bar{a}\bar{a}'b\bar{b}'} + p_{\bar{a}\bar{a}'bb'} - p_{\bar{a}a'\bar{b}\bar{b}'} - p_{\bar{a}a'\bar{b}b'} + p_{\bar{a}a'b\bar{b}'} + p_{\bar{a}a'bb'} & \\ -p_{a\bar{a}'\bar{b}\bar{b}'} - p_{a\bar{a}'\bar{b}b'} + p_{a\bar{a}'b\bar{b}'} + p_{a\bar{a}'bb'} - p_{aa'\bar{b}\bar{b}'} - p_{aa'\bar{b}b'} + p_{aa'b\bar{b}'} + p_{aa'bb'} & = 0, \end{aligned} \tag{30}$$

$$\begin{aligned} -p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} + p_{\bar{a}\bar{a}'\bar{b}b'} - p_{\bar{a}\bar{a}'b\bar{b}'} + p_{\bar{a}\bar{a}'bb'} - p_{\bar{a}a'\bar{b}\bar{b}'} + p_{\bar{a}a'\bar{b}b'} - p_{\bar{a}a'b\bar{b}'} + p_{\bar{a}a'bb'} & \\ -p_{a\bar{a}'\bar{b}\bar{b}'} + p_{a\bar{a}'\bar{b}b'} - p_{a\bar{a}'b\bar{b}'} + p_{a\bar{a}'bb'} - p_{aa'\bar{b}\bar{b}'} + p_{aa'\bar{b}b'} - p_{aa'b\bar{b}'} + p_{aa'bb'} & = 0, \end{aligned} \tag{31}$$

$$\begin{aligned} p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} + p_{\bar{a}\bar{a}'\bar{b}b'} - p_{\bar{a}\bar{a}'b\bar{b}'} - p_{\bar{a}\bar{a}'bb'} + p_{\bar{a}a'\bar{b}\bar{b}'} + p_{\bar{a}a'\bar{b}b'} - p_{\bar{a}a'b\bar{b}'} - p_{\bar{a}a'bb'} & \\ -p_{a\bar{a}'\bar{b}\bar{b}'} - p_{a\bar{a}'\bar{b}b'} + p_{a\bar{a}'b\bar{b}'} + p_{a\bar{a}'bb'} - p_{aa'\bar{b}\bar{b}'} - p_{aa'\bar{b}b'} + p_{aa'b\bar{b}'} + p_{aa'bb'} & = \frac{1}{\sqrt{2}}, \end{aligned} \tag{32}$$

$$\begin{aligned} p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} - p_{\bar{a}\bar{a}'\bar{b}b'} + p_{\bar{a}\bar{a}'b\bar{b}'} - p_{\bar{a}\bar{a}'bb'} + p_{\bar{a}a'\bar{b}\bar{b}'} - p_{\bar{a}a'\bar{b}b'} + p_{\bar{a}a'b\bar{b}'} - p_{\bar{a}a'bb'} & \\ -p_{a\bar{a}'\bar{b}\bar{b}'} + p_{a\bar{a}'\bar{b}b'} - p_{a\bar{a}'b\bar{b}'} + p_{a\bar{a}'bb'} - p_{aa'\bar{b}\bar{b}'} + p_{aa'\bar{b}b'} - p_{aa'b\bar{b}'} + p_{aa'bb'} & = \frac{1}{\sqrt{2}}, \end{aligned} \tag{33}$$

$$\begin{aligned} p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} + p_{\bar{a}\bar{a}'\bar{b}b'} - p_{\bar{a}\bar{a}'b\bar{b}'} - p_{\bar{a}\bar{a}'bb'} - p_{\bar{a}a'\bar{b}\bar{b}'} - p_{\bar{a}a'\bar{b}b'} + p_{\bar{a}a'b\bar{b}'} + p_{\bar{a}a'bb'} & \\ +p_{a\bar{a}'\bar{b}\bar{b}'} + p_{a\bar{a}'\bar{b}b'} - p_{a\bar{a}'b\bar{b}'} - p_{a\bar{a}'bb'} - p_{aa'\bar{b}\bar{b}'} - p_{aa'\bar{b}b'} + p_{aa'b\bar{b}'} + p_{aa'bb'} & = \frac{1}{\sqrt{2}}, \end{aligned} \tag{34}$$

$$\begin{aligned} p_{\bar{a}\bar{a}'\bar{b}\bar{b}'} - p_{\bar{a}\bar{a}'\bar{b}b'} + p_{\bar{a}\bar{a}'b\bar{b}'} - p_{\bar{a}\bar{a}'bb'} - p_{\bar{a}a'\bar{b}\bar{b}'} + p_{\bar{a}a'\bar{b}b'} - p_{\bar{a}a'b\bar{b}'} + p_{\bar{a}a'bb'} & \\ +p_{a\bar{a}'\bar{b}\bar{b}'} - p_{a\bar{a}'\bar{b}b'} + p_{a\bar{a}'b\bar{b}'} - p_{a\bar{a}'bb'} - p_{aa'\bar{b}\bar{b}'} + p_{aa'\bar{b}b'} - p_{aa'b\bar{b}'} + p_{aa'bb'} & = -\frac{1}{\sqrt{2}}, \end{aligned} \tag{35}$$

where we are using the simplifying notation that $p_{aa'bb'} = p^*(\omega_{aa'bb'})$, $p_{aa'b\bar{b}'} = p^*(\omega_{aa'b\bar{b}'})$, and so on. Notice that equation (27) corresponds to the condition $\mu(\Omega) = 1$ in Definition 7. Equations (28)–(31) represent the expectations in (26). Finally, equations (32)–(35) are the expectations computed using (13).

Equations (27)–(35) form a set of nine linearly independent equations. However, to completely determine the probabilities of each the 16 elementary events $\omega_i \in \Omega$, one needs a total of 16 equations. Thus, the problem is under-determined. However, it is possible to write a general solution to (27)–(35) that will have seven undetermined parameters, and it is straightforward to show that at least one of the p_{ω_i}'s are negative for all possible solutions. But if one compute the marginal expectations for each of the experimental contexts, one would observe that for contexts $C_1 = (A, B)$, $C_2 = (A, B')$, $C_3 = (A', B)$, and $C_4 = (A', B)$ all the marginal probabilities are non-negative. What we mean is that the marginal probabilities observed in, say, C_1, i.e. $p^*(A = \pm 1, B = \pm 1)$, are all non-negative. This comes from the constraints in (27)–(35). An explicit solution to (27)–(35) is lengthy and cumbersome but can be obtained easily. The interested reader can either examine a solution given in reference [6] or compute it themselves.

We now prove a general relationship between quantum mechanics and negative probabilities.

Proposition 5. *Let $\mathcal{Q}$ be the collection of complete sets of simultaneously observable one-dimensional projection operators on a Hilbert space $\mathcal{H}$, i.e., for each $Q_i \in \mathcal{Q}$ there are $N = \dim \mathcal{H}$ commuting projection operators such that $\sum_{\hat{P}_j \in Q_i} \hat{P}_j = \hat{1}$. Let p be a measure over elements of Q_i given by Born's rule. Let also $\{R_i^*\}$ be a collection of extended dichotomous random variables on a signed measure space $(\Omega, \mathcal{F}, \mu)$, such that for each Q_i there is a context C_i such that for all $\hat{P}_j \in Q_i$ there is a 1-1 equivalent element of C_i with the same marginal probability distributions, i.e., within a context C_i the expectations of R_j^* and $\hat{P}_j$ are the same, as well as any other higher moments in combination with other variables in the same context. Then μ is a negative probability space that represents all contexts C_i.*

Proof. To prove that μ is negative probability space, we just need to show that $\mu(\Omega) = 1$. In order do so, let us notice that each extended random variable R_i^* defines a partition of the sample space Ω corresponding to each of their values (similarly to what we had in Equations (21)–(25)). For each combination of extended random variables, there is a corresponding partition. In particular, for a given projection operator, say, $\hat{P}_1$, by assumption, there exists a two-valued extended random variable R_1^*. The two outcomes, $R_1^* = 1$ and $R_1^* = -1$, define a partition of Ω, formed by two subsets that we denote by F_1 and F_{-1}, such that $F_1 \cap F_{-1} = \emptyset$ and $F_1 \cup F_{-1} = \Omega$. Since the measure μ assigns to those subsets the same probabilities as the Born's rule, we must have $1 = \langle \hat{P}_1 \rangle + \langle \hat{1} - \hat{P}_1 \rangle = \mu(F_1) + \mu(F_{-1}) = \mu(\Omega)$. Thus, μ is normalized, and defines a negative probability. $\square$

In the following section we present, in Propositions 6 and 7, examples of how this result applies in simple but important cases. We end with this section with a final comment. The requirement that p^* minimizes the L1 norm (see Definition 9) provides us with a number δ that is greater than or equal to zero. If it is zero, the random variables are not contextual, and proper a joint probability distribution exists. However, the correlations for the Bell-EPR case do not allow for a proper joint [58]. The fact that δ is not zero provides a way for measuring how contextual (or, in this case, because it is contextual-at-a-distance, how non-local) a system of random variables is. The more δ departs from 0, the more contextual it is [17,25,59,60].

In this section, we showed a generalized probability theory that includes negative (or signed) probabilities. This theory is well suited for describing quantum systems, as it is compatible with the no-signaling condition. Furthermore, negative probabilities have advantages with other alternative extended probability theories. For example, upper and lower probabilities can also be used to describe quantum contextuality [47,48]. However, because upper and lower probabilities involve inequalities, their computation is challenging and cumbersome. Additionally, the main appeal for upper and lower probabilities is that they have an interpretation. For instance, monotonic upper and lower probabilities can be interpreted within Dempster-Shaffer theory (they call them plausibility and belief, respectively) [61]. However, this interpretation fails in quantum theory, where upper and lower probabilities are non-monotonic, and Dempster-Shaffer's reasoning does not apply anymore.

Unlike upper probabilities, negative probabilities can be easily computed, as shown in the example above. Furthermore, one can use negative probabilities as a contextual calculus for conflicting subjective contextual information even outside of physics [62–65]. So, the use of negative probabilities for quantum systems seem worth exploring.

However, a question often asked is this: what is the meaning of an event having a negative probability? First, we point out that, in our definition, negative probability events are never observed: negative probabilities exist for the unobserved joint events. This is similar to the use of negative numbers to count physical objects, e.g. apples in a fruit stand. Of course, the concept of a negative number of apples is absurd: one could never observe -3 apples. This is emphasized by DeMorgan's comment about negative numbers [66]: "[the student] must reject the definition still sometimes given of the quantity $-a$, that it is less than nothing. It is astonishing that the human intellect should ever have tolerated such an absurdity as the idea of a quantity less than nothing; above all, that the notion should have outlived the belief in judicial astrology and the existence of witches, either of which is ten thousand times more possible." Even though the meaning may be problematic for DeMorgan, the use of negative numbers to track operations of future sales and purchases of apples does not need to be; a negative number of apples makes sense, but only as an accounting trick that helps us figure out the observable (non-negative) final number of apples. We do not *need* an interpretation of negative numbers of apples. In this sense, an interpretation of negative probabilities is as unnecessary as an interpretation of negative numbers of apples.

Nevertheless, there are many different interpretations of negative probabilities for non-monotonic systems (see [5,40,41,43,67,68]). For example, Khrennikov proposes that negative probabilities are associated with sequences that violate von Mises's principle of stability, which states that probabilities are about well-behaved sequences whose mean converge to a certain number [41]. By focusing on infinite sequences that do not converge using the standard real-number metric, Khrennikov showed that such sequences converge using p-adic numbers, with negative probabilities being associated to such sequences that violated the principle of stability. Another approach is that of Abramsky and Brandenburger [5]. They proposed to use negative probabilities to describe a data table where events could themselves be signed. In their interpretation, the joint event of, say, three random variables being +1, would also carry an additional bit, a sign. Two events could then cancel each other if their signs were different, and negative probabilities manifest those two types of events. As mentioned in the previous paragraph, another way to think about negative probabilities is the pragmatic view: negative probabilities are a useful tool for computing quantum probabilities. This view does not demand an interpretation, and it was the way that both Feynman and Dirac thought about negative probabilities [35,36]. In this paper, we are proposing that, at least in quantum physics, negative probabilities can be interpreted as a miscounting and mislabeling of a data table because quantum particles, and some propositions about them, are indistinguishable.

4. Indistinguishability in Quantum Mechanics and Mathematics

Compound quantum systems can be prepared in entangled states that violate non-contextuality inequalities. An example we saw was the state in (12), whose correlations (26) lead to a violation of (11). However, there is a different physical effect associated with compound quantum systems involving particles of *the same kind*. To write the state of the compound system, we must invoke the symmetrization postulate. This postulate asserts that the state of a compound quantum system of identical particles must be symmetric under permutation of the particles if the particles are Bosons and anti-symmetric if they are Fermions.

Suppose that we have two Fermions, one of them prepared in the state $|a\rangle$ and the other in the state $|b\rangle$. Then, after applying the symmetrization postulate, the state of the compound system is given by

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|a\rangle \otimes |b\rangle - |b\rangle \otimes |a\rangle \right) . \tag{36}$$

A similar procedure should be used to construct the state of two Bosons by using a plus instead of a minus sign, thus yielding a symmetric state.

The implications of the symmetrization postulate (SP) are of significant importance for quantum theory. Pauli's exclusion principle and also the so-called quantum statistics (Einstein-Bose and Fermi-Dirac statistics) follow from the SP. This feature of the quantum formalism is particularly relevant for the study of the properties of indistinguishable particles in quantum information theory [69–71]. Furthermore, the peculiar properties of compound systems of identical particles lead to heated debates in the literature about the interpretation of quantum mechanics. A remarkable position was that of E. Schrödinger, who claimed that elementary particles are not individuals, given that the theory gives no means to identify them [72,73]. An even more extraordinary view was that suggested by Wheeler, who once told Feynman that all electrons have the same properties because they are all the same electron [74]. We do not necessarily agree with Schrödinger or Wheeler, but we emphasize a broad agreement among physicists that two electrons are indistinguishable at some fundamental level.

Researchers discussed the indistinguishability of elementary particles in connection to indistinguishability in logic and mathematics. Indeed, to deal with genuinely indistinguishable entities, the quasi-set theory was developed as a set-theoretical framework in which the classical laws of identity do not apply for specific elements of the theory (see, for example, [75–77]). This formalism was used in [78,79] to reconstruct the Fock-space formulation of quantum mechanics avoiding any particle labeling (see [80] for an alternative approach). The axioms of quasi-set theory are chosen so that it is possible to form collections of indistinguishable entities, violating Leibniz's principle of identity of indiscernibles [75]. In this theory, the identity symbol "$=$" cannot be applied to all its elements. Instead, a weaker equivalence relation "$\equiv$" is used to describe a situation where an element x is indistinguishable from another element y, and it is formally represented by $x \equiv y$. This corresponds to the idea that x and y represent indistinguishable quantum objects.

Quasi-set theory assumes that a cardinal can be assigned to these collections so that every quasi-set has a definite number of elements. The indistinguishable elements of a quasi-set cannot be identified by names, counted, or ordered. In this sense, the standard set-theory rules do not apply for all elements of the theory. Quasi-sets having indistinguishable elements are thought of as representing collections of quantum objects of the same kind, i.e., indistinguishable objects. Another essential feature of quasi-set theory is that it contains a copy of Zermelo-Fraenkel set theory to develop standard mathematics within it.

Quasi-set theory allows us to formally describe collections of indiscernible objects without resorting to any mathematical tricks. The connection between indistinguishability and contextuality was studied recently. In [19], we have shown that the possibility of identifying particles in different contexts lies at the core of the Kochen-Specker contradiction. In [7], we studied how the assumption of the indistinguishability of properties allows one to understand the occurrence of contextual random variables.

The connection between particle indistinguishability and indistinguishability of properties is essential here. So, let us examine how it comes about. In the quantum formalism, a testable proposition about an object is formally represented by a projection operator. Given an observable A, consider the proposition "the value of A lies in the interval Δ" (that we write compactly as $P^A(\Delta)$). By using the spectral theorem, $P^A(\Delta)$ can be mathematically represented by an orthogonal projection $\hat{P}^A(\Delta)$ (notice that the "hat" distinguishes the mathematical object from the proposition it represents). We aim to represent quantum properties related to the particles and describe expressions such as “a particle has a certain property.”

It is instructive to illustrate the connection between quantum indistinguishability and the identification of propositions with the same content, but in different contexts, by considering a quasi-pair concept in quasi-set theory. The quasi-pair $\langle [x], P^A(\Delta) \rangle$ can be used to describe one quanta possessing the property $P^A(\Delta)$ (see also the discussion presented in [7]), where the $[x]$ is the collection of all possible indistinguishable elements from x. Thus, $\langle [x], P^A(\Delta) \rangle$, can be interpreted as: “a quantum object satisfies that the value of A lies in Δ". Notice that we refer to a quantum object, without specifying which one it is (because, according to the spirit of quasi-set theory, they are indiscernible). The classical analog of this proposition could make explicit reference to the particle identity (as, for example, in "particle e_1 satisfies that the value of A lies in Δ"). Moreover, we could use standard set theory and write $\langle \{e_1\}, P^A(\Delta) \rangle$ (notice that, in the last pair, we are using the standard singleton $\{e_1\}$, which is formed by the sole individual e_1). However, this is impossible if we assume that quantum particles are indistinguishable, and we use quasi-set theory. If we now take another quanta y such that $y \equiv x$, and consider the proposition $\langle [y], P^A(\Delta) \rangle$, using the rules of quasi-set theory, we obtain $\langle [x], P^A(\Delta) \rangle \equiv \langle [y], P^A(\Delta) \rangle$. This can be interpreted as follows: *indistinguishability of particles leads to the identification of propositions among different contexts*. Each time we consider different instances of a proposition about a quantum system, the propositions associated with these instances are indistinguishable, and thus, they can be identified. Notice that a proportion’s instantiation has the form “a quantum object’s value of A lies in Δ.” If we now have an instantiation of an equivalent assertion, but considered in a different context, given that we cannot refer to the identity of the quanta involved, we have no means to distinguish the propositions either. Assuming the axioms given in [75], indistinguishable quasi-sets are identical (but have in mind that, in this framework, identity is a *derived notion*). It is in this sense that indistinguishable propositions can be identified.

The above discussion is particularly relevant for the problem mentioned at the end of Section 2. Given the random variables A_B and $A_{B'}$ discussed in Section 2 (that have the same content), we have two options: either $A_B = A_{B'}$, or $A_B \neq A_{B'}$. Assuming that quanta are indistinguishable and describing propositions using quasi-set theory (as above), when all propositions associated to A_B have indistinguishable counterparts in those associated to $A_{B'}$, we obtain that $A_B \equiv A_{B'}$ (i.e., they can be identified as random variables). The assumption of quanta indistinguishability, together with the use of quasi-set theory, serves as a justification for identifying those random variables (see [19] and [7] for a related discussion).

Let us now use the above framework to connect particle indistinguishability with non-signaling. Let **A** and **B** represent two agents, Alice and Bob, that aim to communicate with each other. For **A** to send a signal to **B**, they need to appeal to some physical mechanism that can be generally described by sharing a physical system that induces observable correlations between what they observe on it.

Suppose that they can measure different observables on their respective sides. We denote by A, A', etc., the observables for **A**, and B, B', etc. for **B**). Given A and A', we assume that they are complementary, i.e., that if Alice selects A, she cannot at the same time select A'; similarly for Bob's B and B'. However, because Alice and Bob are observing different parts of the communication device, we assume that any of the observables for **A** are always compatible with whatever choice Bob makes in **B**. The idea of a communication device is that Alice can affect Bob's observations of B or B' by changing her settings from observing A to A' (or vice versa), .

Let us assume now that Alice and Bob construct a device that works. In other words, they figured out a way to communicate between themselves using some (unknown to us) mechanism where Alice's choices affect Bob's observations. However, Alice and Bob now make a new proposal: they want to see if their device works with indistinguishable quantum particles. This proposal means that whenever we have the contexts (A, B) and (A, B'), the properties associated with A in context B are indistinguishable from those of A in context B'. Under these assumptions, we should have that, for each property, the probability of obtaining $P^A(\Delta)$ in context B is the same as the probability of obtaining $P^A(\Delta)$ in context B'. If they were not the same, Alice could use these probabilities to attach an "identity card" to some particles in B but not to others. This would be a way of distinguishing indistinguishable particles.

The above conclusion leads to the following conditions:

$$\sum_b p(P^A(a), P^B(b)|A, B) = \sum_b p(P^A(a), P^{B'}(b)|A, B') = p(P^A(a)|A) \tag{37}$$

and

$$\sum_a p(P^A(a), P^B(b)|A, B) = \sum_a p(P^{A'}(a), P^B(b)|A', B) = p(P^B(b)|B). \tag{38}$$

Equations (37) and (38) are no-signaling conditions [55]. Thus, the assumption of indistinguishability of properties leads to the no-signaling condition: whatever Alice does to "her particle" cannot affect what Bob infers about "his particle," because this would mean attaching an identity card to Alice's and Bob's particles. This condition is extreme, and is specific to physical theories, in particular quantum mechanics, and should not hold in other domains (such as cognition; see, for example [63,65,81]).

To summarize, quantum particles are indistinguishable, and this indistinguishability leads to the indistinguishability of properties. However, we showed that property indistinguishability implies that communication devices such as those discussed by [82] cannot work. If we could use the correlations in entangled systems to send a signal between Alice and Bob, such devices could distinguish particles.

Let us consider two examples that illustrate how the following chain of implications works.

$$\text{Indistinguishability} \Longrightarrow \text{No-signal} \Longrightarrow \text{Negative probabilities}$$

We illustrate the above idea with Propositions 6 and 7. Below we go through the proof of Propositions 6 and 7, but we stress that the proofs are all based on the idea put forth above, namely that indistinguishability implies no-signaling, and therefore negative probabilities. Let us first clarify the notation. Consider three dichotomous random variables forming jointly measurable pairs $X - Y$, $X - Z$, and $Y - Z$. We denote by X_Y the random variable X in the context $X - Y$, with a similar interpretation for X_Z, Y_X, Y_Z, Z_X, and Z_Y. Then, we have the following proposition, whose proofs follow the above idea that indistinguishability implies no-signaling, which implies negative probabilities.

Proposition 6. *For jointly measurable pairs $X - Y$, $X - Z$ and $Y - Z$ of dichotomous random variables, if the indistinguishability relations $X_Y \equiv X_Z$, $Y_X \equiv Y_Z$, and $Z_X \equiv Z_Y$ are satisfied, there exists a signed probability space (i.e., satisfying Definition 7), for which each pair of jointly measurable variables is a context (satisfying Definition 4).*

Proof. Suppose that we have three dichotomous random variables, X, Y and Z. Assume that $X - Y$, $X - Z$ and $Y - Z$, are jointly measurable. In the context $X - Y$, we have different elementary events, which are given by $X = 1$ and $Y = 1$, $X = -1$ and $Y = 1$, $X = 1$ and $Y = -1$, and $X = -1$ and $Y = -1$. Let us denote these results by xy, $x\bar{y}$, $\bar{x}y$ and $\bar{x}\bar{y}$, respectively. Combined propositions are given by sets like $\{xy, x\bar{y}\}$ (representing the proposition "xy or $x\bar{y}$"), $\{x\bar{y}, \bar{x}y, \bar{x}\bar{y}\}$ (representing "not xy"), and so on. If we define $X - Y := \{xy, x\bar{y}, \bar{x}y, \bar{x}\bar{y}\}$, the complete Boolean algebra is given by $\mathcal{P}(X - Y)$ (that we denote by $\mathcal{B}_{X;Y}$) and can be represented by the diagram in Figure 1.

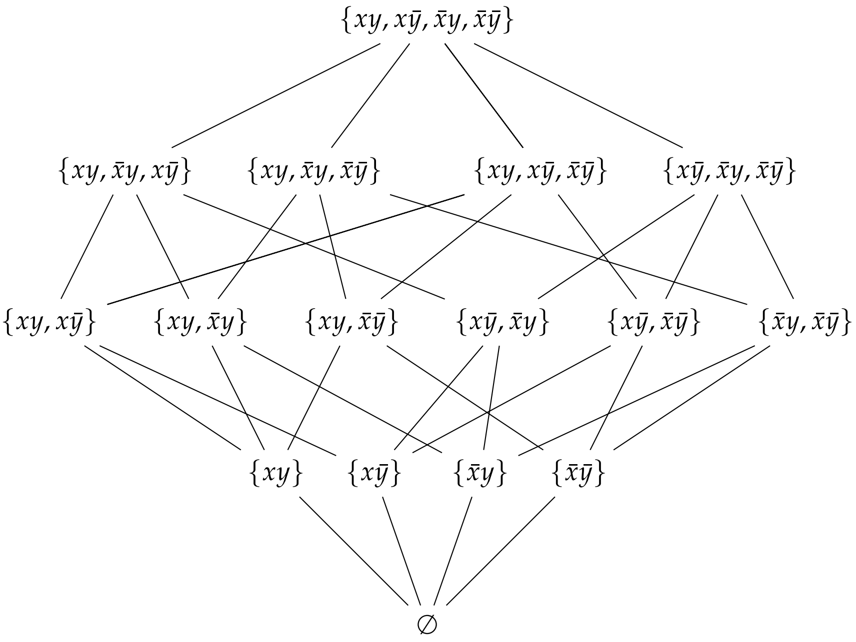

Figure 1. Hasse diagram of the $X - Y$ Boolean algebra.

Analogous Boolean algebras $\mathcal{B}_{X;Z}$ and $\mathcal{B}_{Y;Z}$ hold for $X - Z$ and $Y - Z$, which are given by all possible subsets of $\{xz, x\bar{z}, \bar{x}z, \bar{x}\bar{z}\}$ and $\{yz, y\bar{z}, \bar{y}z, \bar{y}\bar{z}\}$, respectively. The random variable X can be considered in the context $X - Y$ (we denote this random variable by X_Y). The proposition "$X = 1$ in the context Y, disregarding the value of Y", is represented by the proposition $\{xy, x\bar{y}\}$. Its negation, is given by $\{\bar{x}y, \bar{x}\bar{y}\}$. It is easy to check that the set $\mathcal{B}_{X_Y} := \{\emptyset, \{xy, x\bar{y}\}, \{\bar{x}y, \bar{x}\bar{y}\}, \{xy, x\bar{y}, \bar{x}y, \bar{x}\bar{y}\}\}$ forms a Boolean subalgebra of $\mathcal{B}_{X;Y}$. And we also have an isomorphism of Boolean algebras between $\mathcal{B}_{X_Y}$ and $\mathcal{P}(\{x, \bar{x}\}) := \mathcal{B}_X$. Thus, we have that the random variable X considered in context Y defines a sub-Boolean algebra of $\mathcal{B}_{X;Y}$. The same happens for Y_X, and X_Z with regard to $\mathcal{B}_{X;Z}$, Y_Z with regard to $\mathcal{B}_{Y;Z}$, etc. We certainly have that $\mathcal{B}_{X_Y}$ is isomorphic to $\mathcal{B}_{X_Z}$, $\mathcal{B}_{Y_X}$ is isomorphic to $\mathcal{B}_{Y_Z}$, etc. Should we identify those random variables? As remarked in the Introduction, this is a crucial problem in probability theory and statistics. In quantum physics, we usually do that, but this is not necessarily so in other fields of research.

As discussed above, we assume that object's indistinguishability implies the identification of properties. Thus, we assume that X_Y and X_Z can be identified as random variables. This means that, given the isomorphism between $\mathcal{B}_{X_Y}$ and $\mathcal{B}_{X_Z}$, for each proposition $F_1 \in \mathcal{B}_{X_Z}$, we have $F_2 \in \mathcal{B}_{X_Z}$ such that its content is the same, and that it has the same probability of occurrence. As an example of this, consider the sets $F_1 = \{xy, x\bar{y}\}$ (that corresponds to the assertion "$X = 1$ in context $X - Y$") and $F_2 = \{xz, x\bar{z}\}$ (that corresponds to the assertion "$X = 1$ in context $X - Z$"). As sets, they are different. But we can identify F_1 and F_2 in the following sense: for any (classical) probability assignments $(X - Y, \mathcal{B}_{X;Y}, p_{X;Y})$ and $(X - Z, \mathcal{B}_{X;Z}, p_{X;Z})$, we must have that $p_{X;Y}(F_1) = p_{X;Z}(F_2)$ (i.e., the probabilities are numerically identical for propositions taken from different contexts).

Up to now, we have the following situation. We have three different Boolean algebras of propositions, $\mathcal{B}_{X;Y}$, $\mathcal{B}_{X;Z}$ and $\mathcal{B}_{Y;Z}$. $\mathcal{B}_{X;Y}$ contains $\mathcal{B}_{X_Y}$ and $\mathcal{B}_{Y_X}$ as Boolean subalgebras (and the same happens for $\mathcal{B}_{X;Z}$ and $\mathcal{B}_{X_Z}$ and $\mathcal{B}_{Z_X}$ and $\mathcal{B}_{Y;Z}$ and $\mathcal{B}_{Y_Z}$ and $\mathcal{B}_{Z_Y}$). Furthermore, we have that, due to the indistinguishability postulate, all probability assignments $(X-Y, \mathcal{B}_{X;Y}, p_{X;Y})$, $(X-Z, \mathcal{B}_{X;Z}, p_{X;Z})$ and $(Y-Z, \mathcal{B}_{Y;Z}, p_{Y;Z})$, must be compatible with regard to indistinguishable propositions. Is there a Boolean algebra containing all the propositions in $\mathcal{B}_{X;Y}$, $\mathcal{B}_{X;Z}$ and $\mathcal{B}_{Y;Z}$? Can we find a global probability assignment compatible with $p_{X;Y}$ $p_{X;Z}$ and $p_{Y;Z}$? In the following, we show how to build that required Boolean algebra, and how to build a signed probability assignment for arbitrary (but positive) $p_{X;Y}$ $p_{X;Z}$ and $p_{Y;Z}$.

Define $X-Y-Z := \{xyz, \bar{x}yz, x\bar{y}z, xy\bar{z}, \bar{x}\bar{y}z, \bar{x}y\bar{z}, x\bar{y}\bar{z}, \bar{x}\bar{y}\bar{z}\}$ and $\mathcal{B}_{X;Y;Z} := \mathcal{P}(X-Y-Z)$. We need to recover $\mathcal{B}_{X;Y}$, $\mathcal{B}_{X;Z}$ and $\mathcal{B}_{Y;Z}$ as subalgebras of $\mathcal{B}_{X-Y-Z}$. In order to do so, define $(X-Y)_Z := \{\{xyz, xy\bar{z}\}, \{x\bar{y}z, x\bar{y}\bar{z}\}, \{\bar{x}yz, \bar{x}y\bar{z}\}, \{\bar{x}\bar{y}z, \bar{x}\bar{y}\bar{z}\}\}$ and $\mathcal{B}_{(X-Y)_Z} := \mathcal{P}((X-Y)_Z)$. It is obvious that $\mathcal{B}_{(X-Y)_Z}$ is isomorphic to $\mathcal{B}_{X;Y}$. We can also define $\mathcal{B}_{(X-Z)_Y}$ and $\mathcal{B}_{(Y-Z)_X}$ in an analogous way, and obtain algebras isomorphic to $\mathcal{B}_{X;Z}$ and $\mathcal{B}_{Y;Z}$, respectively. Similarly, if we consider $\mathcal{B}_{X_{Y-Z}} := \{\emptyset, \{xyz, x\bar{y}z, xy\bar{z}, x\bar{y}\bar{z}\}, \{\bar{x}yz, \bar{x}\bar{y}z, \bar{x}y\bar{z}, \bar{x}\bar{y}\bar{z}\}, \mathbf{1}\}$, we obtain a Boolean subalgebra of $\mathcal{B}_{X;Y;Z}$ which is isomorphic to $\mathcal{B}_{X_Y}$. Indeed, $\mathcal{B}_{X_{Y-Z}}$ is isomorphic to $\mathcal{B}_{X_Y}$ and $\mathcal{B}_{X_Z}$, reflecting the fact that those random variables were identified by the relation "$\equiv$".

It is possible now to define a signed probability space $(X-Y-Z, \mathcal{B}_{X;Y;Z}, p_{X;Y;Z})$ satisfying Definition 9 as follows. Let $p_{X;Y;Z}(F) := p_{X;Y}(F)$, whenever $F \in \mathcal{B}_{(X-Y)_Z}$, $p_{X;Y;Z}(F) := p_{X;Z}(F)$, whenever $F \in \mathcal{B}_{(X-Z)_Y}$, and $p_{X;Y;Z}(F) := p_{Y;Z}(F)$, whenever $F \in \mathcal{B}_{(Y-Z)_X}$. We must also impose that $\sum_{\omega \in X-Y-Z} p_{X;Y;Z}(\omega) = 1$. Let us now build $p_{X;Y;Z}$ explicitly. In order to shorten the notation, in some parts we write $p_{X;Y;Z}(xyz) := p_{xyz}$, $p_{X;Y;Z}(\bar{x}yz) := p_{\bar{x}yz}$, $p_{X;Y;Z}(x\bar{y}z) := p_{x\bar{y}z}$, and so on. The first constrain that we impose is normalization:

$$p_{xyz} + p_{\bar{x}yz} + p_{x\bar{y}z} + p_{xy\bar{z}} + p_{x\bar{y}\bar{z}} + p_{\bar{x}y\bar{z}} + p_{\bar{x}\bar{y}z} + p_{\bar{x}\bar{y}\bar{z}} = 1 \tag{39}$$

Notice that Equation (39) imposes the following normalization conditions on $p_{X;Y}$, $p_{X;Z}$ and $p_{Y;Z}$:

$$p_{X;Y}(xy) + p_{X;Y}(\bar{x}y) + p_{X;Y}(x\bar{y}) + p_{X;Y}(\bar{x}\bar{y}) = 1 \tag{40a}$$

$$p_{X;Y}(xz) + p_{X;Y}(\bar{x}z) + p_{X;Y}(x\bar{z}) + p_{X;Y}(\bar{x}\bar{z}) = 1 \tag{40b}$$

$$p_{Y;Z}(yz) + p_{Y;Z}(\bar{y}z) + p_{Y;Z}(y\bar{z}) + p_{Y;Z}(\bar{y}\bar{z}) = 1 \tag{40c}$$

The context $X-Y$ imposes the following constrains on $p_{X;Y;Z}$. First, notice that $p_{X;Y}$ is fixed by the following: $\langle X \rangle$, $\langle Y \rangle$ and $\langle XY \rangle$, and the normalization condition (40a). In therms of $p_{X;Y;Z}$, this can be expressed as:

$$\begin{aligned} &p_{X;Y;Z}(xyz) - p_{X;Y;Z}(\bar{x}yz) + p_{X;Y;Z}(x\bar{y}z) + p_{X;Y;Z}(xy\bar{z}) + p_{X;Y;Z}(x\bar{y}\bar{z}) - \\ &p_{X;Y;Z}(\bar{x}y\bar{z}) - p_{X;Y;Z}(\bar{x}\bar{y}z) - p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) = \langle X \rangle \end{aligned} \tag{41a}$$

$$\begin{aligned} &p_{X;Y;Z}(xyz) + p_{X;Y;Z}(\bar{x}yz) - p_{X;Y;Z}(x\bar{y}z) + p_{X;Y;Z}(xy\bar{z}) - p_{X;Y;Z}(x\bar{y}\bar{z}) + \\ &p_{X;Y;Z}(\bar{x}y\bar{z}) - p_{X;Y;Z}(\bar{x}\bar{y}z) - p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) = \langle Y \rangle \end{aligned} \tag{41b}$$

$$\begin{aligned} &p_{X;Y;Z}(xyz) - p_{X;Y;Z}(\bar{x}yz) - p_{X;Y;Z}(x\bar{y}z) + p_{X;Y;Z}(xy\bar{z}) - p_{X;Y;Z}(x\bar{y}\bar{z}) - \\ &p_{X;Y;Z}(\bar{x}y\bar{z}) + p_{X;Y;Z}(\bar{x}\bar{y}z) + p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) = \langle XY \rangle \end{aligned} \tag{41c}$$

Similarly, for the context $X-Z$, besides equations (41a) and (40b) for the mean value of X, we have:

$$p_{X;Y;Z}(xyz) + p_{X;Y;Z}(\bar{x}yz) + p_{X;Y;Z}(x\bar{y}z) - p_{X;Y;Z}(xy\bar{z}) - p_{X;Y;Z}(x\bar{y}\bar{z}) - p_{X;Y;Z}(\bar{x}y\bar{z}) + p_{X;Y;Z}(\bar{x}\bar{y}z) - p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) = \langle Z \rangle \tag{42a}$$

$$p_{X;Y;Z}(xyz) - p_{X;Y;Z}(\bar{x}yz) + p_{X;Y;Z}(x\bar{y}z) - p_{X;Y;Z}(xy\bar{z}) - p_{X;Y;Z}(x\bar{y}\bar{z}) + p_{X;Y;Z}(\bar{x}y\bar{z}) - p_{X;Y;Z}(\bar{x}\bar{y}z) + p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) = \langle XZ \rangle \tag{42b}$$

Finally, for the context $Y - Z$, besides equation (40c) and the mean values of Y and Z (given by (41b) and (42a), respectively), we have

$$p_{X;Y;Z}(xyz) + p_{X;Y;Z}(\bar{x}yz) - p_{X;Y;Z}(x\bar{y}z) - p_{X;Y;Z}(xy\bar{z}) + p_{X;Y;Z}(x\bar{y}\bar{z}) - p_{X;Y;Z}(\bar{x}y\bar{z}) - p_{X;Y;Z}(\bar{x}\bar{y}z) + p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) = \langle YZ \rangle \tag{43a}$$

Notice that the mean values of X, Y and Z are imposed only once. This is possible only because we have made the identifications $X_Y \equiv X_Z$, $Z_Y \equiv Z_X$ and $Y_X \equiv Y_Z$. Equations (39), (41), (42), and (43), constitute a set of seven compatible equations for $p_{X;Y;Z}$. As is well known,eight independent equations are needed to define $p_{X;Y;Z}$. Thus, there are infinitely many solutions that satisfy our indistinguishability conditions for contexts. Each one of these solutions, by construction, satisfy our definition of signed probability given in (9). There is one parameter free for determining $p_{X;Y;Z}$, namely, the mean value $\langle XYZ \rangle$. In order to study the space of solutions, let us write down the matrix form of the set of equations (39), (41), (42), and (43):

$$\begin{bmatrix} 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p_{X;Y;Z}(xyz) \\ p_{X;Y;Z}(\bar{x}yz) \\ p_{X;Y;Z}(x\bar{y}z) \\ p_{X;Y;Z}(xy\bar{z}) \\ p_{X;Y;Z}(x\bar{y}\bar{z}) \\ p_{X;Y;Z}(\bar{x}y\bar{z}) \\ p_{X;Y;Z}(\bar{x}\bar{y}z) \\ p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) \end{bmatrix} = \begin{bmatrix} \langle X \rangle \\ \langle Y \rangle \\ \langle Z \rangle \\ \langle XY \rangle \\ \langle XZ \rangle \\ \langle YZ \rangle \\ 1 \end{bmatrix}$$

The solutions are given by

$$p_{X;Y;Z}(xyz) = \frac{1}{4}\left(1 + \langle XY \rangle + \langle XZ \rangle + \langle YZ \rangle\right) - \alpha, \tag{44a}$$

$$p_{X;Y;Z}(\bar{x}yz) = \frac{1}{4}\left(\langle Y \rangle + \langle Z \rangle - \langle XY \rangle - \langle YZ \rangle\right) + \alpha, \tag{44b}$$

$$p_{X;Y;Z}(x\bar{y}z) = \frac{1}{4}\left(\langle X \rangle + \langle Z \rangle - \langle XY \rangle - \langle YZ \rangle\right) + \alpha, \tag{44c}$$

$$p_{X;Y;Z}(xy\bar{z}) = \frac{1}{4}\left(\langle X \rangle + \langle Y \rangle - \langle XZ \rangle - \langle YZ \rangle\right) + \alpha, \tag{44d}$$

$$p_{X;Y;Z}(x\bar{y}\bar{z}) = \frac{1}{4}\left(1 - \langle Y \rangle - \langle Z \rangle + \langle YZ \rangle\right) - \alpha, \tag{44e}$$

$$p_{X;Y;Z}(\bar{x}y\bar{z}) = \frac{1}{4}\left(1 - \langle X \rangle - \langle Z \rangle + \langle XZ \rangle\right) - \alpha, \tag{44f}$$

$$p_{X;Y;Z}(\bar{x}\bar{y}z) = \frac{1}{4}\left(1 - \langle X \rangle - \langle Y \rangle + \langle XY \rangle\right) - \alpha, \tag{44g}$$

$$p_{X;Y;Z}(\bar{x}\bar{y}\bar{z}) = \alpha, \tag{44h}$$

where α is a free parameter. It is immediate from the above solutions that for some correlations, e.g., $\langle XY\rangle = \langle XZ\rangle = \langle YZ\rangle = -1$ no non-negative solutions exist.

□

We use a similar notation as before (but with four jointly measurable pairs) in the following Proposition.

Proposition 7. *For jointly measurable pairs $X - Z$, $X - W$, $Y - Z$ and $Y - W$ of dichotomous random variables, if the indistinguishability relations $X_Z \equiv X_W$, $Y_Z \equiv Y_W$, $Z_X \equiv Z_Y$ and $W_X \equiv W_Y$ are satisfied, there exists a signed probability space (i.e., satisfying Definition 7), for which each pair is a context (satisfying Definition 4).*

Proof. Now, let us work out the example with four dychotomic random variables X, Y, Z and W. This example is relevant in the Alice and Bob scenario. Let us assume that $X - Z$, $X - W$ and $Y - Z$ and $Y - W$ form jointly measurable quantities. Proceeding as before, we impose the indistinguishability conditions $X_Z \equiv X_W$, $Y_Z \equiv Y_W$, $Z_X \equiv Z_Y$ and $W_X \equiv W_Y$. Again, we will have the Boolean algebras $\mathcal{B}_{X;Z}$, $\mathcal{B}_{X;W}$, $\mathcal{B}_{X;Z}$, $\mathcal{B}_{Y;Z}$, $\mathcal{B}_{Y;W}$, $\mathcal{B}_{X_Z}$, $\mathcal{B}_{X_W}$, and so on. In order to build a Boolean algebra containing all these algebras as subalgebras, consider $X;Y;Z;W := \{xyzw, \bar{x}yzw, x\bar{y}zw, xy\bar{z}w, xyz\bar{w}, \bar{x}\bar{y}zw, \bar{x}y\bar{z}w, \bar{x}yz\bar{w}, x\bar{y}\bar{z}w, x\bar{y}z\bar{w}, xy\bar{z}\bar{w}, \bar{x}\bar{y}\bar{z}w, x\bar{y}\bar{z}\bar{w}, \bar{x}y\bar{z}\bar{w}, \bar{x}\bar{y}z\bar{w}, \bar{x}\bar{y}\bar{z}\bar{w}\}$ and $\mathcal{B}_{X;Y;Z} := \mathcal{P}(X;Y;Z;W)$. It is straightforward to check that the algebras associated to all jointly measurable variables are subalgebras of $\mathcal{B}_{X;Y;Z}$. Let us work out an example. In order to get a subalgebra of $\mathcal{B}_{X;Y;Z}$ isomorphic to $\mathcal{B}_{X;Z}$, consider the set:

$$\begin{aligned}\mathcal{P}\,(\{\{xyzw, x\bar{y}zw, xyz\bar{w}, x\bar{y}z\bar{w}\}, \{\bar{x}yzw, \bar{x}\bar{y}zw, \bar{x}yz\bar{w}, \bar{x}\bar{y}z\bar{w}\},\\ \{xy\bar{z}w, x\bar{y}\bar{z}w, xy\bar{z}\bar{w}, x\bar{y}\bar{z}\bar{w}\}, \{\bar{x}y\bar{z}w, \bar{x}\bar{y}\bar{z}w, \bar{x}y\bar{z}\bar{w}, \bar{x}\bar{y}\bar{z}\bar{w}\}\})\end{aligned} \tag{45}$$

Proceeding similarly, we can show that all the desired algebras can be considered as subalgebras of $\mathcal{B}_{X;Y;Z}$. Now, we assume as before that there exist joint probability spaces $(X;Z, \mathcal{B}_{X;Z}, p_{X;Z})$, $(X;W, \mathcal{B}_{X;W}, p_{X;W})$, $(Y;Z, \mathcal{B}_{Y;Z}, p_{Y;Z})$ and $(Y;W, \mathcal{B}_{Y;W}, p_{Y;W})$. As before, $(X;Z, \mathcal{B}_{X;Z}, p_{X;Z})$ is solely determined by the normalization condition and the values of $\langle X\rangle$, $\langle Z\rangle$ and $\langle XZ\rangle$ (and similar parameters for the other jointly measurable variables). In order to get a global probability, let us proceed us before, by imposing these conditions on $p_{X;Y;Z;W}$. Given that the equations are cumbersome, we just write the matrix equations, which are:

$$\begin{pmatrix}
1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\
1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \\
1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix} \times
\begin{bmatrix}
p(xyzw) \\ p(\bar{x}yzw) \\ p(x\bar{y}zw) \\ p(xy\bar{z}w) \\ p(xyz\bar{w}) \\ p(\bar{x}\bar{y}zw) \\ p(\bar{x}y\bar{z}w) \\ p(\bar{x}yz\bar{w}) \\ p(x\bar{y}\bar{z}w) \\ p(xy\bar{z}\bar{w}) \\ p(x\bar{y}z\bar{w}) \\ p(x\bar{y}\bar{z}\bar{w}) \\ p(\bar{x}y\bar{z}\bar{w}) \\ p(\bar{x}\bar{y}z\bar{w}) \\ p(\bar{x}\bar{y}\bar{z}w) \\ p(\bar{x}\bar{y}\bar{z}\bar{w})
\end{bmatrix}
=
\begin{bmatrix}
\langle X \rangle \\ \langle Y \rangle \\ \langle Z \rangle \\ \langle W \rangle \\ \langle XZ \rangle \\ \langle XW \rangle \\ \langle YZ \rangle \\ \langle YW \rangle \\ 1
\end{bmatrix} \tag{46}$$

Each row above corresponds to a linearly independent equation, and therefore the above equations are compatible. Since there are fewer equations than variables, there are infinitely many solutions satisfying our definitions of negative probability and contexts (with seven arbitrary parameters). An explicit solution is shown in the Appendix. □

The above procedure can be extended to an arbitrary set of dichotomous random variables. Compatible equations are obtained each time we add equations that respect the indistinguishability condition between different random variables.

5. Conclusions

In this work, we have put forth the following argument. We started by pointing out a well-known and robust connection between contextual theories (such as quantum mechanics) and signed (or negative) probabilities. To generalize this connection, we presented a definition of signed probabilities that relies solely on the notions of signed measurable space and measurement contexts. As expected from previous results, the signed probabilities defined here satisfy the no-signaling condition. With a formal definition of negative probabilities, we followed previous works' reasoning line on indistinguishability and contextuality. We discussed how the assumption of (ontic) particle indistinguishability leads to the following conclusion. Some of the particle testable propositions can be identified among different contexts. This characteristic, in turn, implies the non-signaling condition. Our findings suggest that, in the quantum domain, there is a robust connection between indistinguishability assumptions and the existence of signed probabilities. To generalize this connection, we presented a definition of signed probabilities that rely on the notions of signed measurable space and measurement contexts, extending Kolmogorov's approach naturally.

It should be clear why negative probabilities are suitable to describe the states of indistinguishable entities. Negative probabilities are necessary and sufficient for no-signaling, and the identification of testable propositions imply no-signaling. Additionally, indistinguishable particles and propositions may lead to contradictions if we assume that their underlying logic is classic. However, as shown in [19] and [7], such contradictions rely on counterfactual reasoning that assumes the classical theory of identity for particles and properties. Therefore, in this situation, we can interpret negative probabilities as the consequence of imposing on indistinguishable particles a classical way of counting, i.e., a Boolean algebra. When doing so, we need to allow for negative counts to correct for the over-counting of different but indistinguishable particles. This different accounting for events is, in a certain sense, similar to Abramsky and Brandenburger's operational interpretation of negative probabilities [5]. However, contrary to their interpretation, here we propose that this accounting comes from an error in identifying properties, which is due to a fundamental ontological property of particles: they are indistinguishable.

Paul Dirac was the first to use negative probabilities in physics. He used them to deal with the problem of infinities in quantum field theory [35]. Later, Richard Feynman tried to use negative probabilities in quantum mechanics [36]. It is fair to say that, though such influential physicists worked with them, negative probabilities remain outside of mainstream physics. The reason is likely not about a lack of meaning for the concept of negative probabilities, as we saw multiple references proposing different interpretations. Perhaps the main reason is that, albeit interesting and easy to compute, negative probabilities did not produce yet any exciting insights into quantum mechanics. We hope that with a well-defined concept of negative probabilities and a connection to a clear ontology inspired by quantum mechanics, negative probabilities can yield new understanding about the quantum world.

Author Contributions: All authors contributed equally for this paper.

Funding: F.H. was partially funded by the project "Per un'estensione semantica della Logica Computazionale Quantistica- Impatto teorico e ricadute implementative" , Regione Autonoma della Sardegna, (RAS: RASSR40341), L.R. 7/2017, annualità 2017- Fondo di Sviluppo e Coesione (FSC) 2014-2020.

Acknowledgments: Both authors thank Professors Décio Krause and Pawel Kurzynski for discussions. We also thank the anonymous referees for comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

1. Klyachko, A.A.; Can, M.A.; Binicioğlu, S.; Shumovsky, A.S. Simple Test for Hidden Variables in Spin-1 Systems. *Phys. Rev. Lett.* **2008**, *101*, 020403. doi:10.1103/PhysRevLett.101.020403.
2. Acacio de Barros, J.; Suppes, P. Inequalities for Dealing with Detector Inefficiencies in Greenberger-Horne-Zeilinger–Type Experiments. *Phys. Rev. Lett.* **2000**, *84*, 793–797. doi:10.1103/PhysRevLett.84.793.
3. Singer, M.; Stulpe, W. Phase-space representations of general statistical physical theories. *Journal of Mathematical Physics* **1992**, *33*, 131–142, [https://doi.org/10.1063/1.529975]. doi:10.1063/1.529975.
4. Abramsky, S.; Brandenburger, A. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics* **2011**, *13*, 113036. doi:10.1088/1367-2630/13/11/113036.
5. Abramsky, S.; Brandenburger, A. An Operational Interpretation of Negative Probabilities and No-Signalling Models. In *Horizons of the Mind. A Tribute to Prakash Panangaden*; van Breugel, F.; Kashefi, E.; Palamidessi, C.; Rutten, J., Eds.; Number 8464 in Lecture Notes in Computer Science, Springer Int. Pub., 2014; pp. 59–75.
6. de Barros, J.A.; Kujala, J.V.; Oas, G. Negative probabilities and contextuality. *Journal of Mathematical Psychology* **2016**, *74*, 34–45. doi:10.1016/j.jmp.2016.04.014.
7. de Barros, J.A.; Holik, F.; Krause, D. Indistinguishability and the origins of contextuality in physics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2019**. doi:10.1098/rsta.2019.0150.
8. Stalnaker, R. *Context and Content: Essays on Intentionality in Speech and Thought*; Clarendon Press: Oxford ; New York, 1999.

9. de Barros, J.A.; Montemayor, C.; De Assis, L.P.G.; Skokowski, P.; Perry, J. Rationality, belief, and the possible-worlds semantics. *Submitted.* **2019**.
10. Kochen, S.; Specker, E.P. The Problem of Hidden Variables in Quantum Mechanics. *Journal of Mathematics and Mechanics* **1967**, *17*, 59–87.
11. Cabello, A. Simple Explanation of the Quantum Violation of a Fundamental Inequality. *Physical Review Letters* **2013**, *110*, 060402. doi:10.1103/PhysRevLett.110.060402.
12. Randall, C.; Foulis, D. Properties and operational propositions in quantum mechanics. *Foundations of Physics* **1983**, *13*, 843–857.
13. Da Costa, N.; Lombardi, O.; Lastiri, M. A modal ontology of properties for quantum mechanics. *Synthese* **2013**, *190*, 3671–3693.
14. de Barros, J.A.; Suppes, P. Inequalities for Dealing with Detector Inefficiencies in Greenberger-Horne-Zeilinger Type Experiments. *Physical Review Letters* **2000**, *84*, 793–797.
15. Clauser, J.; Horne, M.; Shimony, A.; Holt, R. Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters* **1969**, *23*, 880–884. doi:10.1103/PhysRevLett.23.880.
16. Bell, J. On the Problem of Hidden Variables in Quantum Mechanics. *Rev. Mod. Phys.* **1966**, *38*, 447–452.
17. Dzhafarov, E.N.; Kujala, J.V. Contextuality-by-Default 2.0: Systems with Binary Random Variables. In *Quantum Interaction: 10th International Conference, QI 2016*; de Barros, J.A.; Coecke, B.; Pothos, E., Eds.; Springer International Publishing, 2017; Vol. 10106, *Lecture Notes in Computer Science*. arXiv: 1604.04799.
18. Dzhafarov, E.N. The Contextuality-by-Default View of the Sheaf-Theoretic Approach to Contextuality. *arXiv:1906.02718 [quant-ph]* **2019**. arXiv: 1906.02718.
19. de Barros, J.A.; Holik, F.; Krause, D. Contextuality and Indistinguishability. *Entropy* **2017**, *19*, 435. doi:10.3390/e19090435.
20. Hillery, M.; O'Connell, R.; Scully, M.; Wigner, E. Distribution functions in physics: Fundamentals. *Physics Reports* **1984**, *106*, 121 – 167. doi:https://doi.org/10.1016/0370-1573(84)90160-1.
21. Wigner, E. On the Quantum Correction For Thermodynamic Equilibrium. *Physical Review* **1932**, *40*, 749–759. doi:10.1103/PhysRev.40.749.
22. Dirac, P.A.M. Bakerian Lecture - The physical interpretation of quantum mechanics. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1942**, *180*, 1–40, [https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1942.0023]. doi:10.1098/rspa.1942.0023.
23. Feynman, R.P. Negative Probability. In *Quantum Implications: Essays in Honour of David Bohm*; Routledge & Kegan Paul Ltd, London & New York, 1987; pp. 235–248.
24. Cahill, K.E.; Glauber, R.J. Density Operators and Quasiprobability Distributions. *Phys. Rev.* **1969**, *177*, 1882–1902. doi:10.1103/PhysRev.177.1882.
25. de Barros, J.A.; Dzhafarov, E.N.; Kujala, J.V.; Oas, G. Measuring Observable Quantum Contextuality. In *Quantum Interaction*; Atmanspacher, H.; Filk, T.; Pothos, E., Eds.; Number 9535 in Lecture Notes in Computer Science, Springer International Publishing, 2015; pp. 36–47. doi:10.1007/978-3-319-28675-4_4.
26. Kujala, J.V.; Dzhafarov, E.N. Measures of contextuality and non-contextuality. *Philosophical Transactions of the Royal Society A* **2019**, *377*, 20190149.
27. Abramsky, S.; Brandenburger, A. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics* **2011**, *13*, 113036. doi:10.1088/1367-2630/13/11/113036.
28. Al-Safi, S.W.; Short, A.J. Simulating all Nonsignaling Correlations via Classical or Quantum Theory with Negative Probabilities. *Phys. Rev. Lett.* **2013**, *111*, 170403. doi:10.1103/PhysRevLett.111.170403.
29. Spekkens, R.W. Negativity and Contextuality are Equivalent Notions of Nonclassicality. *Phys. Rev. Lett.* **2008**, *101*, 020401. doi:10.1103/PhysRevLett.101.020401.
30. Leonhardt, U. Discrete Wigner function and quantum-state tomography. *Phys. Rev. A* **1996**, *53*, 2998–3013. doi:10.1103/PhysRevA.53.2998.
31. Cormick, C.; Galvão, E.F.; Gottesman, D.; Paz, J.P.; Pittenger, A.O. Classicality in discrete Wigner functions. *Phys. Rev. A* **2006**, *73*, 012301. doi:10.1103/PhysRevA.73.012301.
32. Veitch, V.; Ferrie, C.; Gross, D.; Emerson, J. Negative quasi-probability as a resource for quantum computation. *New Journal of Physics* **2012**, *14*, 113011. doi:10.1088/1367-2630/14/11/113011.
33. Galvão, E.F. Discrete Wigner functions and quantum computational speedup. *Phys. Rev. A* **2005**, *71*, 042302. doi:10.1103/PhysRevA.71.042302.
34. Suppes, P. Probability concepts in quantum mechanics. *Philosophy of Science* **1961**, *28*, 378–389.

35. Dirac, P. Bakerian Lecture. The Physical Interpretation of Quantum Mechanics. *Proceedings of the Royal Society of London B* **1942**, *A180*, 1–40.
36. Feynman, R. Negative probability. In *Quantum implications: essays in honour of David Bohm*; Hiley, B.; Peat, F., Eds.; Routledge: London and New York, 1987; pp. 235–248.
37. MÃŒckenheim, G. A review of extended probabilities. *Physics Reports* **1986**, *133*, 337–401.
38. Oas, G.; de Barros, J.A.; Carvalhaes, C. Exploring non-signalling polytopes with negative probability. *Physica Scripta* **2014**, *T163*, 014034.
39. Al-Safi, S.; Short, A. Simulating all Nonsignaling Correlations via Classical or Quantum Theory with Negative Probabilities. *Physical Review Letters* **2013**, *111*, 170403. doi:10.1103/PhysRevLett.111.170403.
40. Burgin, M. Interpretations of Negative Probabilities. *arXiv:1008.1287 [physics, physics:quant-ph]* **2010**. arXiv: 1008.1287.
41. Khrennikov, A. *Interpretations of Probability*; Walter de Gruyter, 2009.
42. SzÃ©kely, G.J. Half of a coin: negative probabilities. *Wilmott Magazine* **2005**, *50*, 66–68.
43. Machado, T.; A, J. Fractional Coins and Fractional Derivatives. *Abstract and Applied Analysis* **2013**, *2013*. doi:10.1155/2013/205097.
44. Kolmogorov, A. *Foundations of the theory of probability*, 2nd ed.; Chelsea Publishing Co.: Oxford, England, 1956.
45. Galavotti, M.C. *Philosophical introduction to probability*; Vol. 167, *CSLI Lecture Notes*, CSLI Publications: Stanford, CA, 2005.
46. Foulis, D.J.; Randall, C.H. Operational Statistics. I. Basic Concepts. *Journal of Mathematical Physics* **1972**, *13*, 1667–1675. doi:10.1063/1.1665890.
47. Suppes, P.; Zanotti, M. Existence of hidden variables having only upper probabilities. *Foundations of Physics* **1991**, *21*, 1479–1499.
48. de Barros, J.A.; Suppes, P. Probabilistic Inequalities and Upper Probabilities in Quantum Mechanical Entanglement. *Manuscrito* **2010**, *33*, 55–71.
49. Holik, F.; Saenz, M.; Plastino, A. A discussion on the origin of quantum probabilities. *Annals of Physics* **2014**, *340*, 293–310. doi:10.1016/j.aop.2013.11.005.
50. Clauser, J.; Horne, M. Experimental consequences of objective local theories. *Physical Review D* **1974**, *10*, 526–535. doi:10.1103/PhysRevD.10.526.
51. Sokolovski, D. Weak values, “negative probability,” and the uncertainty principle. *Physical Review A* **2007**, *76*, 042125.
52. Hosoya, A.; Shikano, Y. Strange weak values. *Journal of Physics A: Mathematical and Theoretical* **2010**, *43*, 385307. doi:10.1088/1751-8113/43/38/385307.
53. Halmos, P. *Measure Theory*; Springer-Verlag: New York, NY, 1974.
54. Abramsky, S.; Hardy, L. Logical Bell inequalities. *Physical Review A* **2012**, *85*, 062114. doi:10.1103/PhysRevA.85.062114.
55. Popescu, S.; Rohrlich, D. Quantum nonlocality as an axiom. *Foundations of Physics* **1994**, *24*, 379–385. doi:10.1007/BF02058098.
56. Oas, G.; de Barros, J.A. A Survey of Physical Principles Attempting to Define Quantum Mechanics. In *Contextuality From Quantum Physics to Psychology*; Dzhafarov, E.; Zhang, R.; Jordan, S.M., Eds.; World Scientific, 2015.
57. Suppes, P.; Zanotti, M. When are probabilistic explanations possible? *Synthese* **1981**, *48*, 191–199.
58. Fine, A. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics* **1982**, *23*, 1306–1310. doi:10.1063/1.525514.
59. de Barros, J.A.; Dzhafarov, E.; Kujala, J.; Oas, G. Unifying Two Methods of Measuring Quantum Contextuality. *arXiv:1406.3088 [quant-ph]* **2014**. arXiv: 1406.3088.
60. Dzhafarov, E.; Kujala, J. Generalizing Bell-type and Leggett-Garg-type Inequalities to Systems with Signaling. *arXiv:1407.2886 [quant-ph]* **2014**.
61. Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. *The Annals of Mathematical Statistics* **1967**, *38*, 325–339. doi:10.1214/aoms/1177698950.
62. de Barros, J.A. Decision Making for Inconsistent Expert Judgments Using Negative Probabilities. In *Quantum Interaction: 7th International Conference*; Atmanspacher, H.; Haven, E.; Kitto, K.; Raine, D., Eds.; Lecture Notes in Computer Science, Springer: Berlin/Heidelberg, 2014; pp. 257–269.

63. de Barros, J.A. Beyond the Quantum Formalism: Consequences of a Neural-Oscillator Model to Quantum Cognition. In *Advances in Cognitive Neurodynamics (IV)*; LiljenstrÃ¶m, H., Ed.; Advances in Cognitive Neurodynamics, Springer Netherlands, 2015; pp. 401–404.
64. de Barros, J.A.; Oas, G. Negative probabilities and counter-factual reasoning in quantum cognition. *Physica Scripta* **2014**, *T163*, 014008.
65. de Barros, J.A.; Oas, G. Quantum Cognition, Neural Oscillators, and Negative Probabilities. In *The Palgrave Handbook of quantum models in social science: applications and grand challenges*; Haven, E.; Khrennikov, A., Eds.; Palgrave MacMillan, 2015.
66. DeMorgan, A. *On the Study and Difficulties of Mathematics*; Open Court Publishing Company, 1910.
67. Burgin, M. An Introduction to Symmetric Inflated Probabilities. Quantum Interaction. Springer, Cham, 2016, Lecture Notes in Computer Science, pp. 206–223. doi:10.1007/978-3-319-52289-0_17.
68. Zurek, W.H. Environment-induced superselection rules. *Physical Review D* **1982**, *26*, 1862–1880. doi:10.1103/PhysRevD.26.1862.
69. Lo Franco, R.; Compagno, G. Indistinguishability of Elementary Systems as a Resource for Quantum Information Processing. *Phys. Rev. Lett.* **2018**, *120*, 240403. doi:10.1103/PhysRevLett.120.240403.
70. Bose, S.; Home, D. Generic Entanglement Generation, Quantum Statistics, and Complementarity. *Phys. Rev. Lett.* **2002**, *88*, 050401. doi:10.1103/PhysRevLett.88.050401.
71. Bose, S.; Home, D. Duality in Entanglement Enabling a Test of Quantum Indistinguishability Unaffected by Interactions. *Phys. Rev. Lett.* **2013**, *110*, 140404. doi:10.1103/PhysRevLett.110.140404.
72. Schrödinger, E. 'What is an elementary particle?', reprinted in Castellani, E. *Interpreting bodies: classical and quantum objects in modern physics* **1950**.
73. Schroedinger, E. *Science and humanism*; Cambridge University Press: Cambridge, UK, 1952.
74. Feynman, R.P. The development of the space-time view of quantum electrodynamics. *Nobel Lectures, Physics* **1963**, *1970*, 155–178.
75. Krause, D. On a quasi-set theory. *Notre Dame Journal of Formal Logic* **1992**, *33*, 402–411. doi:10.1305/ndjfl/1093634404.
76. Krause, D.; Sant'Anna, A.S.; Volkov, A.G. Quasi-set theory for bosons and fermions: Quantum distributions. *Foundations of Physics Letters* **1999**, *12*, 51–66. doi:10.1023/A:1021678721611.
77. Domenech, G.; Holik, F. A Discussion on Particle Number and Quantum Indistinguishability. *Foundations of Physics* **2007**, *37*, 855–878. doi:10.1007/s10701-007-9129-5.
78. Domenech, G.; Holik, F.; Krause, D. Q-spaces and the Foundations of Quantum Mechanics. *Foundations of Physics* **2008**, *38*, 969–994. doi:10.1007/s10701-008-9246-9.
79. Domenech, G.; Holik, F.; Kniznik, L.; Krause, D. No Labeling Quantum Mechanics of Indiscernible Particles. *International Journal of Theoretical Physics* **2010**, *49*, 3085–3091. doi:10.1007/s10773-009-0220-x.
80. Compagno, G.; Castellini, A.; Lo Franco, R. Dealing with indistinguishable particles and their entanglement. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *376*, 20170317, [https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2017.0317]. doi:10.1098/rsta.2017.0317.
81. Cervantes, V.H.; Dzhafarov, E.N. Snow queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. *Decision* **2018**, *5*, 193–204. doi:10.1037/dec0000095.
82. Dieks, D. Communication by EPR devices. *Physics Letters A* **1982**, *92*, 271–272. doi:10.1016/0375-9601(82)90084-6.

Appendix

Here we write down an explicit solution for the Alice-Bob system of equations (46). Since (46) has 16 variables but nine equations, the solution will have seven arbitrary parameters, α_i, $i = 1, \dots, 7$. It is straightforward to compute that a general solution for (46) is the following:

$$p(xyzw) = \alpha_1, \tag{47a}$$

$$p(xyz\bar{w}) = \alpha_2, \tag{47b}$$

$$p(xy\bar{z}w) = \alpha_3, \tag{47c}$$

$$p(xy\bar{z}\bar{w}) = \alpha_4, \tag{47d}$$

$$p(x\bar{y}zw) = \alpha_5, \tag{47e}$$

$$p(x\bar{y}z\bar{w}) = \frac{1}{4}\left(1 + \langle XZ \rangle\right) - \alpha_1 - \alpha_2 - \alpha_5, \tag{47f}$$

$$p(x\bar{y}\bar{z}w) = \frac{1}{4}\left(1 + \langle XW \rangle\right) - \alpha_1 - \alpha_3 - \alpha_5, \tag{47g}$$

$$p(x\bar{y}\bar{z}\bar{w}) = -\frac{1}{4}\left(\langle XZ \rangle + \langle XW \rangle\right) + \alpha_1 - \alpha_4 + \alpha_5, \tag{47h}$$

$$p(\bar{x}yzw) = \frac{1}{4}\left(1 + \langle YW \rangle\right) - \alpha_1 - \alpha_3 + \alpha_6, \tag{47i}$$

$$p(\bar{x}yz\bar{w}) = \frac{1}{4}\left(\langle YZ \rangle - \langle YW \rangle\right) - \alpha_2 + \alpha_6 + \alpha_3, \tag{47j}$$

$$p(\bar{x}y\bar{z}w) = \alpha_6, \tag{47k}$$

$$p(\bar{x}y\bar{z}\bar{w}) = \frac{1}{4}\left(1 - \langle YZ \rangle\right) - \alpha_3 - \alpha_4 - \alpha_6, \tag{47l}$$

$$p(\bar{x}\bar{y}zw) = -\frac{1}{4}\left(\langle YW \rangle + \langle XW \rangle\right) + \alpha_1 + \alpha_3 - \alpha_7, \tag{47m}$$

$$p(\bar{x}\bar{y}z\bar{w}) = \frac{1}{4}\left(-\langle XZ \rangle + \langle XW \rangle - \langle YZ \rangle + \langle YW \rangle\right) + \alpha_2 - \alpha_3 + \alpha_7, \tag{47n}$$

$$p(\bar{x}\bar{y}\bar{z}w) = \alpha_7, \tag{47o}$$

$$p(\bar{x}\bar{y}\bar{z}\bar{w}) = \frac{1}{4}\left(\langle XZ \rangle + \langle YZ \rangle\right) + \alpha_3 + \alpha_4 - \alpha_7. \tag{47p}$$

It is straightforward to see that for correlations violating the CHSH form of Bell's inequalities, the above solutions cannot be in the interval $[0,1]$, and are therefore not standard probabilities. For example, for the PR-box correlation of $-\langle XZ \rangle = \langle XW \rangle = \langle YZ \rangle = \langle YW \rangle = -1$, it follows that $p(x\bar{y}\bar{z}w) = -(\alpha_1 + \alpha_3 + \alpha_5)$, which implies that $\alpha_1 = \alpha_3 = \alpha_5 = 0$ for it to be non-negative. This implies, similarly, that $\alpha_7 = 0$ from $p(\bar{x}yzw)$, α_2 from $p(\bar{x}yz\bar{w})$, α_4 from $p(x\bar{y}\bar{z}\bar{w})$, α_6 from $p(\bar{x}\bar{y}\bar{z}\bar{w})$, and α_7 from $p(\bar{x}yzw)$. But since α_i must be zero for $i = 1, \ldots, 7$, it follows that $p(\bar{x}\bar{y}z\bar{w}) = -1/2$, a negative value. Thus, as expected, the PR box maximally violating the CHSH does not have a non-negative joint probability distribution but has a negative probability. Similar results can be obtained for other PR boxes as well as for the QM correlations for the Alice-Bob experiment.