QUANTUM COGNITION, NEURAL OSCILLATORS, AND NEGATIVE PROBABILITIES

J. ACACIO DE BARROS* AND GARY OAS†

Abstract. This review paper has three main goals. First, to discuss a contextual neurophysiologically plausible model of neural oscillators that reproduces some of the features of quantum cognition. Second, to show that such model predicts contextual situations where quantum cognition is inadequate. Third, to present an extended probability theory that that not only can describe situations that are beyond quantum probability, but also provides an advantage in terms of contextual decision making.

1. Introduction

One common view is that humans are rational decision makers. What constitutes *rational* is in itself a matter of debate, but perhaps a common idea of rationality is the notion that, when making decisions, humans follow the prescriptions of classical logic. Where logical true or false values are replaced with uncertainty, we have to deal with beliefs, and not with certainty. One can argue that the rules of inference over beliefs, for a rational being, should be replaced by measures consistent with an underlying Boolean algebra of propositions. If this is the case, and under some reasonable assumptions, the rules of probability theory are derived [27, 59]. In other words, if one wishes to assign measures of belief in such a way that a decision maker, when faced with new evidence, acts in a way consistent with the rules of logic, one needs to use classical probability (CP).

In the 1980s, Tversky and Kahneman examined the heuristics of decision making with cleverly designed experiments where inferences required by CP were tested against actual human beliefs. In a series of results, they showed that in many situations humans did not follow CP, and later on developed a theory to described human decisions, prospect theory, which fitted experimental data better than the standard expected utility theory in economics [60][1]. Such was the importance of those results that Kahneman was awarded the Nobel Memorial Prize in Economics in 2002 (Tversky passed away in 1996).

In 2007, in a special session during the Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium at Stanford University, a group of researchers, among them Andrei Khrennikov, Emmanuel Haven, Jerome Busemeyer, Peter Bruza, and Patrick Suppes, met to discuss applications of the quantum mechanical formalism to the social sciences. The main idea put forth was that, in relation to the social sciences, quantum mechanics could go beyond an analogy of how to deal with complementary variables (in the sense of Bohr): the quantum mechanical formalism itself could better represent situations in which the CP was

[1]Expected utility theory relies heavily on CP [81, 7].

violated (such as Tversky and Kahneman's examples). The area of research spawning from this became known as Quantum Interactions, and the application of the quantum-mathematical formalism to psychology as Quantum Cognition (QC).

In this paper we put forth the following three theses. First, QC is about contextuality. By this we mean that we can think of decisions as experimental outcomes, and such outcomes depend on the experimental conditions (contexts). Second, that contextuality in QC may come from the inconsistency of (perhaps learned, for cognition) conditions. To support this, we provide two examples: a neural oscillator model that shows contextuality when incompatible events are activated, and a decision-making scenario where information based on subjective beliefs are inconsistent. Our third and final thesis is that such contextual effects may be better modeled by allowing non-observable probabilities to take negative values, and not by quantum probabilities that violate classical probability theory (CP). We support this by first showing that there are certain neural oscillator setups that result in responses that cannot be modeled in a natural way by the Hilbert space formalism of quantum mechanics (QM). Then, in another example, we not only show that the QC approach is inferior to negative probabilities (NP)[2], as we call our generalization of CP, but also argue that because of its inherent advantage with respect to Bayesian approaches, NP may be the mechanism of choice for actual biological systems dealing with contextual information.

We organize this paper in the following way. In Section (2) we introduce the idea of QC, and discuss the importance of interference for most of the discussions of violations of CP. In Section (3) we discuss QM, and reason that contextuality is the characteristic that makes its formalism the most relevant to QC. Keeping this in mind, we describe in Section (4) a neurophysiologically inspired neural oscillator model that presents the same contextuality observed in experiments used to support QC. We then show that neural oscillators can model certain decisions that are not compatible with the quantum formalism. Inspired by such model, we then present in Section (5) a theory of extended probabilities that describes the cases found in QC and provides some insight into contextual decision making. Our model seems to be computationally better than the quantum one and seems to offer better advice than the Bayesian approach. We finally end our paper with some remarks and suggestions for future research.

2. Elements of Quantum Cognition

In this section we describe some of the main characteristics of QC. Here we focus on QC models that rely on state interference[3]. We believe that the main features exhibited by these models are sufficient to make our main point. However, we should remark that our arguments and accounts do not immediately generalize to the use of quantum dynamics, but only to the description of the relationship between states and observables.

---

[2]Some readers may object to the use of negative probabilities, since probabilities come from the ratio of two non-negative numbers. We ask them to hold their concerns until Section 5, where we discuss NP in detail. However, at this point we emphasize that in our approach *no experimentally observable* event have NP.

[3]It is not our intent to give an exhaustive account of the field of QC. Readers interested in it should consult the excellent books available (e.g. [67, 21, 54]).

To understand how QM violates CP, and how this can be applied to cognition, let us look at an example. First, following Kolmogorov [71], we have the following definition.

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

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

The elements ω_i of Ω are called *elementary probability events* or simply *elementary events*[4].

Definition 1 implies that from elementary events and $\mathcal{F}$ we can create complex events. In a subjective interpretation the function p could be thought as a measure of rational belief [27, 58]. For example, a consequence of axioms K1–K2 is that, for two sets containing elementary events, A and B, if $A \subset B$, then $p\left(A\right) \leq p\left(B\right)$ follows. This property of CP is called *monotonicity*, and it is possible to show that if we relax the requirement of $\mathcal{F}$ being an algebra of events and instead allow it to be a quantum lattice, monotonicity is violated [56]. In other words, QC violates CP.

An important concept is that of a random variable, defined below.

Definition 2. Let $(\Omega, \mathcal{F}, p)$ be a probability space, and let Θ be a finite set, with $\mathcal{T}$ an algebra over this set. A *random variable* $\mathbf{X}$ is a measurable function $\mathbf{X} : \Omega \to \Theta$, i.e. for every $T \in \mathcal{T}$ we have $\mathbf{X}^{-1}\left(T\right) \in \mathcal{F}$.[5]

Intuitively, we can understand a random variable in the following way. To each value of Ω we assign a value in Θ, such that the function $\mathbf{X}$ determines a partition of the space Ω in different regions (consistent with $\mathcal{F}$, since $\mathbf{X}^{-1}\left(T\right) \in \mathcal{F}$). Such a partition attributes to each region of Ω a value in Θ. So, random variables can be seen as a way to represent possible outcomes of measurements that depend functionally on a probability event.

A simple example to illustrate random variables is the following. Let $\Omega = \{hh, ht, th, tt\}$ be the space of outcomes of tossing a coin twice in a row (h representing heads). If the coin is not biased, we have $p\left(hh\right) = p\left(ht\right) = p\left(th\right) = p\left(tt\right) = 1/4$. Let us say we now want to represent an experiment where whenever we get two values in a row (either heads or tails) the result is 1 (we could think of a one dollar payoff in a game), and -1 otherwise (one dollar lost). The ± 1-valued random variable $\mathbf{X}$ is the function $\mathbf{X} : \Omega \to \{-1, 1\}$ with outcomes $\mathbf{X}\left(hh\right) = \mathbf{X}\left(tt\right) = 1$ and $\mathbf{X}\left(ht\right) = \mathbf{X}\left(th\right) = -1$. From those functions we have the expected value of $\mathbf{X}$, given by

$$E\left(\mathbf{X}\right) = \sum_{\theta \in \Theta} \theta p\left(\mathbf{X} = \theta\right),$$

[4] It follows that any probability of an element of $\mathcal{F}$ is a real number in $[0, 1]$.

[5] Usually there are extra constraints for defining a random variable, but we avoid such technicalities by working with discrete Ω and Θ. The above definition is sufficient for our purposes.

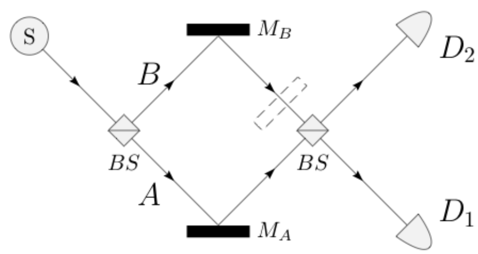

FIGURE 2.1. Mach-Zehnder interferometer. A single photon state is emitted from a source (S) and impinges on the first beam splitter (BS), where it has equal probability of being in arm A or B. Upon reflection at mirrors M_A, M_B, the two paths are recombined at the left beam splitter. The probabilities for detection at detectors D_1 and D_2 are dependent upon the phase relation between the two alternatives at the left beam splitter. In the ideal case, the probability for detection at D_1 is unity. The dashed box in path B represents the choice of inserting a barrier, thereby changing the phase relationship and thus the detection probabilities.

which in our example is

$$\begin{aligned} E(\mathbf{X}) &= (+1)\cdot p(\mathbf{X}=+1) + (-1)\cdot(\mathbf{X}=-1) \\ &= (+1)\cdot\left(\frac{1}{2}\right) + (-1)\cdot\left(\frac{1}{2}\right) = 0. \end{aligned}$$

The second moment is[6]

$$E(\mathbf{XY}) = \sum_{\theta,\phi\in\Theta} \theta\phi p(\mathbf{X}=\theta \& \mathbf{Y}=\phi).$$

A useful notation for ± 1-valued random variables is the following. Instead of writing $p(\mathbf{X}=+1)$, we write $p(x)$, and, instead of $p(\mathbf{X}=-1)$, we write $p(\overline{x})$.

A typical example of nonmonotonicity in QM is the two-slit experiment, whose main features can be seen in the Mach-Zehnder interferometer (MZI) of Fig. 2.1[7]. For this interferometer, imagine two different situations: situation 1, in which the arms of the interferometer are unobstructed and a detection is made on D_1 or D_2, and situation 2, in which a barrier is placed in arm B (represented by the dashed

[6]For ± 1-value random variables with zero expectation, it is easy to show that the moment $E(\mathbf{XY})$ has the same value as the correlation $\rho = E(\mathbf{XY})/(\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}})$.

[7]A more detailed discussion of the MZI in the context presented here can be found in [34, 35].

box in the figure). Following Feynman [47], a particle able to go through either arm of the interferometer has probability of detection 1 in D_1 (for an appropriate choice of lengths for the interferometer arms). However, if the particle is constrained to go through only one of the arms (because of a barrier in the other interferometer arm), the probability of detection in D_1 is 1/2. If we consider $\mathbf{D}_1 = 1$ as the value of $\mathbf{D}_1$ when there is a detection and -1 when not (similarly for $\mathbf{D}_2$), and if we have the ± 1-valued random variable $\mathbf{A} = 1$ (or $\mathbf{B} = 1$) as corresponding to going through A (or B), and -1 otherwise, we have

$$p\left(d_1\right) = p\left(d_1|a\right)p\left(a\right) + p\left(d_1|b\right)p\left(b\right), \tag{2.1}$$

where

$$p\left(x|y\right) \equiv p\left(x,y\right)/p\left(y\right) \tag{2.2}$$

is the *conditional probability* of x given y (for $p\left(y\right) \neq 0$). The expression (2.2) is known as Bayes's formula, and gives the definition of conditional probability in Kolmogorov's axiomatic framework[8]. From

$$p\left(d_1|a\right) = p\left(d_1|b\right) = p\left(a\right) = p\left(b\right) = \frac{1}{2} \tag{2.3}$$

the observed value of $p\left(d_1\right) = 1$ of situation 1 is incompatible with Eq. (2.1) of situation 2. Notice that Eq. (2.1) requires the existence of a joint distribution $p\left(x,y\right)$, and the derivation of (2.3) depends not only on such definition, but also on the additivity of probabilities from Kolmogorov's axioms.

The incompatibility between the observed probabilities for situation 1 and 2 comes from the assumption that the random variable $\mathbf{D}_1$ is the same for both situations, as this is a requirement of a joint probability distribution. However, the experimental conditions are different, and this assumption is somewhat silly: we have no reason to believe they should be the same, and indeed the data does not support this view. We call this impossibility to reconcile the probability distributions of a random variable under different experimental conditions *contextuality*, since each experiment provides an alternative context for the observation[9].

We now return to QC. As mentioned, experiments show that human decision making may not follow CP. For instance, in his classic work, Savage introduced a rational decision-making concept called the Sure Thing Principle (STP) [81]. The idea of the STP is simple:

> "A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant to the attractiveness of the purchase. So, to clarify the matter for himself, he asks whether he should buy if he knew that the Republican candidate were going to win, and decides that he would do so. Similarly, he considers whether he would buy if he knew that the Democratic candidate were going to win, and again finds

[8] In Kolmogorov's theory of probability, joint probabilities are primitives, whereas conditional probabilities are defined from the joints [66]. But other interpretations of probability, notably some subjective interpretations, consider conditional probabilities as more fundamental, and joint probabilities are derived from them. For some of such interpretations, probabilities are always conditional, and it may not even make sense to talk about joint probabilities [51].

[9] Physicists usually refer to contextuality as a particular concept related to hidden-variables in a Kochen-Specker-like situation, and would not call the MZI contextual. Here we take a comprehensive approach to contextuality, which we define mathematically below.

> that he would do so. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains, or will obtain, as we would ordinarily say. It is all too seldom that a decision can be arrived at on the basis of the principle used by this businessman, but, except possibly for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance." [81, pg. 21]

Formally, let us imagine that the ± 1-valued random variable $\mathbf{X}$ corresponds to buy if $\mathbf{X}=1$ (not buy if $\mathbf{X}=-1$), and let another ± 1-valued variable $\mathbf{A}$ be such that $\mathbf{A}=1$ is a Democrat win and $\mathbf{A}=-1$ is a Republican win. If $\mathbf{X}=1$ is preferred over $\mathbf{X}=-1$ when $\mathbf{A}=1$ and also when $\mathbf{A}=-1$, then $\mathbf{X}=1$ is always preferred, since $\mathbf{A}=1$ and $\mathbf{A}=-1$ exhausts all possibilities for $\mathbf{A}$. Savage calls this the *Sure-Thing Principle* (STP).

If we deal with propositions that are not certain, what a "rational" being should base his decisions for preferences of propositions (say, the proposition "buy a certain piece of property" or "$\mathbf{X}=1$") is represented in terms of probabilities. For example, given a set of propositions, say $\{P_1, P_2\}$, we can form more complex propositions by compounding them via the usual operators in propositional calculus, e.g. "$P_1 \& P_2$", "P_1 or P_2", "not P_1", etc. If we require that the rules of inference are such that the measures of belief assigned to propositions are consistent with this composition of propositions (e.g., if you assign a high belief for P_1, then "not P_1" should be assigned a low value), then the measures of belief follow the axioms of probability above [58, 51]. Such axioms imply Savage's STP.

To prove STP from CP, assume that

$$p(\mathbf{X}=1|\mathbf{A}=1) > p(\mathbf{X}=-1|\mathbf{A}=1).$$

This is interpreted as "$\mathbf{X}=1$ is preferred over $\mathbf{X}=-1$" if $\mathbf{A}=1$. If we also assume

$$p(\mathbf{X}=1|\mathbf{A}=-1) > p(\mathbf{X}=-1|\mathbf{A}=-1),$$

then, multiplying each inequality by $p\left(\mathbf{A}=1\right)$ and $p\left(\mathbf{A}=-1\right)$, and using the above notation,

$$p(x|a)p\left(a\right)+p(x|\overline{a})p\left(\overline{a}\right) > p(\overline{x}|a)p\left(a\right)+p(\overline{a}|\overline{a})p\left(\overline{a}\right).$$

From $p(\mathbf{A}=1\&\mathbf{A}=-1)=p\left(a\&\overline{a}\right)=1$ and the definition of conditional probabilities we have

$$P(\mathbf{X}=1) > P(\mathbf{X}=-1).$$

This, is clearly Savage's STP.

Savage's view of probabilities is *normative*, and not *descriptive*. A descriptive theory of decision under uncertainty tells us how actual human beings make decisions, whereas a normative theory tells us how they ought to make decisions (see [16]). In the words Boole, "probability I conceive as to be not so much expectation, as a rational ground for expectation" [15]. Therefore, we should think of probability theory not as a description of what humans actually think (or do), but instead as what humans *should* do faced with uncertain information.

According to such views, STP should hold if agents are making rational decisions. However, as mentioned, actual human decision makers do not follow the STP [90, 82]. For example, in [90] students were told about a game of chance, to be played in two steps. The first step, not voluntary, players had a 50% probability of winning

\$200 and 50% of loosing \$100. The second step allowed a choice of whether or not to gamble a second time, with the same odds and payoffs. When told that that they had won the first bet, 69% of subjects accepted the second gamble, and when told they lost 59% also accepted. If we think of the random variables **A** and **X** as

$$\begin{aligned}
\mathbf{A} = +1 &\quad\leftrightarrow\quad \text{"Won first bet,"}\\
\mathbf{A} = -1 &\quad\leftrightarrow\quad \text{"Lost first bet,"}\\
\mathbf{X} = +1 &\quad\leftrightarrow\quad \text{"Accept second gamble,"}\\
\mathbf{X} = -1 &\quad\leftrightarrow\quad \text{"Reject second gamble,"}
\end{aligned}$$

and

$$\begin{aligned}
P(x|a) = 0.69 > P(\overline{x}|a) = 0.31,\\
P(x|\overline{a}) = 0.59 > P(\overline{x}|\overline{a}) = 0.41,
\end{aligned}$$

then "Accept second gamble" is preferred over "Reject second gamble" regardless of **A**. However, the decision **X** was asked later on the semester, but participants were not told whether they won in the first step (they did not know **A**). Under the unknown condition, 64% of students rejected the second gamble, and

$$P(x) = 1 - 0.64 = 0.36 < P(\overline{x}) = 0.64,$$

a clear violation of the STP.

Violations of CP by human decision makers are one of the main driving force behind QC. For instance, the non-monotonicity of probabilities in the MZI yield results that are very similar to violations of the STP. In the MZI, let us say that the statement "detector D_1 is preferred over D_2" corresponds to a higher probability of detecting a particle in D_1 instead of D_2, and let us represent such statement in terms of the random variable **X**, where

$$p\left(x\right) > p\left(\overline{x}\right) \tag{2.4}$$

corresponds to the previous statement. We can also represent the which-path information by a ± 1-valued random variable **A**, where $\mathbf{A} = 1$ corresponds to the particle going through A and $\mathbf{A} = -1$ to the particle going through B. Clearly, in this case, we have[10]

$$p\left(x|a\right) = p\left(x|\overline{a}\right) = p\left(\overline{x}|a\right) = p\left(\overline{x}|\overline{a}\right) = \frac{1}{2}. \tag{2.5}$$

Similarly to violations of STP, when a or $\overline{a}$, we have no reason to prefer x over $\overline{x}$ or *vice versa*, but CP imply, from (2.5), that $p\left(x\right) = p\left(\overline{x}\right)$, in dissonance with (2.4).

Thus, non-monotonic violations of CP, such as the STP, can be reproduced by quantum interference, as in the MZI. Thus, it should not come as a surprise that the typical QC model relies on interference. For example, Busemeyer, Wang, and Townsend [20] used quantum interference to model the disjunction effect observed by Tversky and Shafir [82, 90]. In their model, they showed that a quantum process with interference yielded better fit to experimental observations than a classical Markov model. There are several other uses of the quantum formalism to cognitive sciences, such as modeling the conjunction effect [79, 18], order effects [89, 11, 91, 19] (see also [68]), and the "guppy" effect [3, 4, 5]. We refer the interested reader to the excellent available reviews, such as [67, 21, 54, 8].

[10] Equal probabilities in (2.5) are not necessary, as biases in the interferometer could modify the conditional probabilities.

To summarize, CP fails to properly describe actual human cognition, and we need to generalize it to account for such cases. A generalization of CP was revealed with the creation of a mathematical formalism to deal with order and context effects in Physics: the quantum formalism. QC tries to use the mathematics of quantum mechanics to describe systems that may have similar contextual effects to the Physical ones.

3. Contextuality in Classical and Quantum Systems

We saw that contextual effects, such as the ones present in the MZI, may give rise to outcomes of experiments that are not consistent with CP. In this section, we explore the idea of contextuality as the connection to violations of CP in QC.

Let us start with the role of contextuality in QC[11]. QM was laid out about 100 years ago. However, it comes as a surprise to many that there is no consensus as to what this theory actually represents. Of course, researchers agree with the theory's predictions, but there is substantial disagreement as to what the theory has to say about the physical systems it models. For example, is the theory about what the system actually is (ontological) or about what we can tell about the system (epistemological)?

The sources of such disagreements are in the many consequences of quantum mechanics that are irreconcilable with the classical views of nature. Following [36], we stress three main characteristics of QM that are not part of classical mechanics: non-determinism, contextuality, and non-locality. To single out contextuality as the relevant aspect to QC, let us analyze each separately.

Determinism, in its simplest form, comes from the idea that a past state of a system determines its future state. This is certainly true in classical mechanics, where given the state of a particle at time t_0 and the forces acting on it, the state of such particle at times $t > t_0$ are determined by

$$\frac{d^2\mathbf{r}\left(t\right)}{dt^2} = \frac{\mathbf{F}\left(t,\mathbf{r},\dot{\mathbf{r}}\right)}{m},$$

where $\mathbf{r}\left(t\right)$ is the position of the particle at time t, m its mass, and $\mathbf{F}\left(t,\mathbf{r},\dot{\mathbf{r}}\right)$ the force acting on the force. From this, it follows that the state of the system at $t \geq t_0$ is completely determined by the particle's position $\mathbf{r}\left(t_0\right)$ and velocity $\dot{\mathbf{r}}\left(t_0\right)$. Determinism is also true for classical electromagnetic theory, and seems to even be consistent with thermodynamics, via the Kinetic Theory of Gases. However, already in the late 1800's, with the discovery of the radioactive decay, some physicists started to realize that some nuclear processes seemed inconsistent with the idea of determinism [77]. As knowledge about microscopic systems increased, and QM developed, physicists realized that quantum systems seemed to be different from classical ones, as they did not always allow for predictable outcomes from the state of the system; not only was the state based on less information (as position and velocity could not be simultaneously measured), but it also had an intrinsically probabilistic connection to measurement outcomes (Born's rule). Thus, QM seems to violate determinism.

However, we can argue that quantum non-determinism is not necessary to QC for two reasons. First, as we argued extensively elsewhere [84], the discrimination

[11] For a quick review of quantum mechanics, see Ref. [55] in this issue.

between determinism and predictability is difficult, and all we can say is that certain cognitive processes may not be predictable. So, positing a non-deterministic underlying process is unnecessary. Second, cognitive models already make extensive use of stochastic processes without resorting to QM [17]. So, the use of QM in cognitive modeling seems unnecessary.

The second feature of QM we analyze is non-locality. In 1932, Einstein, Podoslky, and Rosen (EPR) published a seminal paper [45] where they examined the effects of a measurement on an entangled quantum system, e.g. a system comprised of two particles, 1 and 2. In this paper, EPR argued that if an interaction happens with particle 1, such interaction cannot in any way instantaneously affect particle 2, if the particles have a large spatial separation. In Bohm's version for spin half particles, the EPR state is given by

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2\right), \tag{3.1}$$

where $|+\rangle_i$ ($|-\rangle_i$) corresponds to an eigenvector of spin in direction $\hat{\mathbf{z}}$ with eigenvalue $+1$ (-1) for particle i (we use here units where $\hbar/2 = 1$). As we see from (3.1), if we measure the spin in the direction $\hat{\mathbf{z}}$ for particle 1 and obtain $+1$ (or -1), then we "know" for sure the result of a spin-z measurement for particle 2. Thus, according to EPR, since we cannot have any instantaneous influence of 1 in 2, a measurement in 1 yields information about 2 without disturbing it. EPR then went on and argued that such result would imply that the description of nature given by quantum mechanics was incomplete, as clearly we could know something about 2 without directly measuring it. In a surprising result, John Bell [12, 13] showed that EPR's view that a measurement in 1 did not disturb 2 was inconsistent with the experimental predictions of QM. Therefore, QM seems to allow for some superluminal influence[12]. This characteristic of QM is known as non-locality.

Aspect and collaborators provided evidence for quantum non-locality in the 1980s [10, 9], when they showed that a set of inequalities (known as CHSH inequalities, after reference [25]) were violated. For non-signaling systems[13], the CHSH inequalities are necessary and sufficient conditions for the existence of a joint probability distribution [48], which are also equivalent to the the existence of a local (realistic) theory, meaning that their violation implies non-locality. We point out that to show non-locality, Aspect's experiment had to show correlations between spacelike separated measurements.

With Aspect's experiment in mind, we ask ourselves whether non-locality is relevant to cognition. Given the brain's radius is of order of 10^{-1} m, any events within the brain would have to be correlated within a time window of 10^{-10} s for them to be separated by a spacelike interval. Since there are no cognitive processes

[12]Bell's results and the actual claims about superluminal influences are conceptually very subtle, and it would go beyond the scope of this article to carefully explain them. We refer the interested reader to Bell's excellent papers in [14].

[13]In physics, the non-signaling condition is the statement that no matter, energy, or information (i.e. signal) can be sent between two spacelike separated events. It is a restriction imposed by relativity theory. In practice, this condition simply states that marginal probabilities for one observer cannot change when a second, far away, observer changes the choice of measurement, such that the choices and measurements are spacelike separated. Other terms for this property of marginal probabilities have been used (e.g. "parameter independence," "marginal selectivity," among others).

that can be measured within such intervals of time, no empirical evidence of non-locality in the brain should be expected. Furthermore, since cognitive processes are several orders of magnitude slower than 10^{-10} s, one could never reject classical mechanisms that explain such influences. Therefore non-locality should not be pertinent for QC.

We are then left with the idea of contextuality. Contextuality in QM was discussed explicitly by Kochen and Specker in [70], but its roots appeared early on in the realization that experiments did not actually reveal the outcomes of a pre-existing quantity, but instead create them. In Peres's example [78], three measurement directions, $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, and $\hat{\mathbf{e}}_3$, for spin $1/2$ present problems if assumed that measurements reveal the component of the spin in a given direction, i.e. if we imagine that the particle has some unknown spin μ and that measuring it in direction $\hat{\mathbf{e}}_1$ reveals the component of μ in this direction (i.e., $\hat{\mathbf{e}}_1 \cdot \mu$). Since each spin measurement only yields either $+1$ or -1, if we choose our directions such that $\hat{\mathbf{e}}_1 + \hat{\mathbf{e}}_2 + \hat{\mathbf{e}}_3 = 0$, we have

$$\mu \cdot (\hat{\mathbf{e}}_1 + \hat{\mathbf{e}}_2 + \hat{\mathbf{e}}_3) = \mu \cdot \hat{\mathbf{e}}_1 + \mu \cdot \hat{\mathbf{e}}_2 + \mu \cdot \hat{\mathbf{e}}_3,$$

which would yield a contradiction, since the left hand side is zero (by our choice of directions) and the right hand side is either ± 3, or ± 1, but never zero. This problem is resolved when we realize that an experiment to measure $\hat{\mathbf{e}}_1$ is incompatible with an experiment to measure $\hat{\mathbf{e}}_2$ or $\hat{\mathbf{e}}_3$ (spins operators do not commute), and that the contradiction comes from assuming that the values of the spin components do not change when we change the experiment.

We say a set of experimental outcomes are contextual if their values change under different conditions (see [43, 44, 41, 32]). To illustrate this, imagine three ± 1-valued random variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ recorded under such conditions that we never observe all three simultaneously, but only in pairs (e.g. $\mathbf{X}$ and $\mathbf{Y}$ but not $\mathbf{Z}$, or $\mathbf{Y}$ and $\mathbf{Z}$ but not $\mathbf{X}$). For simplicity assume that their expectations are all zero, $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$, and that they are perfectly anti-correlated, $E(\mathbf{XY}) = E(\mathbf{XZ}) = E(\mathbf{YZ}) = -1$. Now, the assumption that a variable is the same under different experimental conditions leads to a contradiction. To see this, start with a hypothetical $\mathbf{X} = 1$ and $\mathbf{Y} = -1$ on a trial. The second correlation gives $\mathbf{Z} = -1$, but the third correlation gives $\mathbf{Z} = 1$. Clearly, $\mathbf{Z}$ when measured with $\mathbf{X}$ is different from $\mathbf{Z}$ measured with $\mathbf{Y}$, and this system is contextual.

We formalize contextuality following Dzhafarav and Kujala. Let us assume that variables are *a priori* contextual, and instead of calling them $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, we include a label to describe context. For the three correlation experimental conditions, we have the following six variables: $\mathbf{X}_\mathbf{Y}$, $\mathbf{Y}_\mathbf{X}$,$\mathbf{X}_\mathbf{Z}$, $\mathbf{Z}_\mathbf{X}$, $\mathbf{Y}_\mathbf{Z}$, and $\mathbf{Z}_\mathbf{Y}$. For these variables, the observed correlations are $E(\mathbf{X}_\mathbf{Y}\mathbf{Y}_\mathbf{X}) = E(\mathbf{X}_\mathbf{Z}\mathbf{Z}_\mathbf{X}) = E(\mathbf{Y}_\mathbf{Z}\mathbf{Z}_\mathbf{Y}) = -1$, and it is straightforward to confirm that no contradiction arises from this expanded set of random variables. So, we now have a clear definition of contextuality: our system of three random variables is non-contextual if and only if it is possible to find a probability distribution consistent with the observed correlations and expectations such that $P(\mathbf{X}_\mathbf{Y} = \mathbf{X}_\mathbf{Z}) = P(\mathbf{Y}_\mathbf{X} = \mathbf{Y}_\mathbf{Z}) = P(\mathbf{Z}_\mathbf{Y} = \mathbf{Z}_\mathbf{X}) = 1$. In other words, a system is non-contextual if the values of the random variables do not depend on the experimental contexts, and contextual otherwise. The notion of non-contextuality (and contextuality) can be easily extended to more variables, and we refer the reader to reference [40].

The most famous case of a contextual quantum system was presented in Kochen and Specker's seminal paper [70], where they provided a set of yes-no questions that if answered in accord with quantum mechanical predictions lead to inconsistencies, similar to our example above (though with 118 questions, instead of only three). As mentioned above, the CHSH inequalities are equivalent to the existence of a joint probability distribution. Its violation by quantum mechanics means that Bell-type quantum systems are contextual. However, they present a special type of contextuality, where the contexts for the variables are set by the choice the experimenters make in a spacelike separated interval (thus the non-local character of QM). Furthermore, because they are probabilistic, unlike Kochen-Specker, they are not often discussed as examples of contextuality[14], though they clearly are, if we think of contextuality as above. As such, QM provides other types of contextuality which do not require non-locality, such as is the case with the MZI or with order effects.

Can contextuality be a feature of cognitive systems? Absolutely. As we saw in the examples from QC above, the cases where CP fails to describe all situations where different contexts were used to probe an answer (say, a known versus unknown context, in the violation of the STP). Furthermore, as we show in the next section, such contextual outcomes can be modeled in a very classical way.

To summarize, in this Section we discussed reasons for using QM in cognitive models. Among those reasons, we argued that only stochasticity and contextuality are relevant. To support this, in the next section we provide a neural model that fits the same outcomes as quantum cognitive models, but also provide cases where outcomes are contextual but yet not describable by quantum mechanics[15].

4. A neural model of Quantum Cognition

In the previous sections, we discussed the features of QM relevant to QC. We argued that contextuality is the most probable feature relevant to social systems. In this section we present, in a hopefully intuitive way, a *classic* neural oscillator model that replicates some of the characteristics of QC [85]. Our goal is that such model might shed some light into the limitations of using QM to model cognition.

Our model relies on neurophysiological evidence that suggests cognitive processes as an activity involving large collections of synchronizing neurons. This is corroborated by EEG experiments showing the EEG data as a good representation of language or visual imagery [**?**]. In this Section we follow [33], and readers interested in more technical details are referred to [85].

In our model, the mathematical behavioral stimulus-response theory (SR theory) is described by synchronized neural oscillators[16]. SR theory is one of the most successful behavioral theories, mainly because it can be mathematically formalized as a simple set of axioms. In terms of random variables $\mathbf{Z}$, $\mathbf{S}$, $\mathbf{R}$, and $\mathbf{E}$, with $\mathbf{Z} : \Omega \to E^{|S|}$, $\mathbf{S} : \Omega \to S$, $\mathbf{R} : \Omega \to R$, and $\mathbf{E} : \Omega \to E$, where S is the set of stimuli, R the set of responses, and E the set of reinforcements, a trial in SR theory has

[14] There are exceptions, such as the works of Cabello [23, 22, 24, 52].

[15] Here we mean non describable in the sense discussed in [69]; see also [28, 31].

[16] It is beyond the scope of this article to give a full fledged account of SR theory, and here we only attempt to describe it in an intuitive way. Readers interested in a mathematical treatment of this theory are referred to [83, 86].

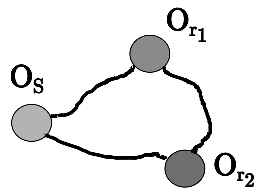

FIGURE 4.1. Schematic representation of the SR oscillator model for two possible responses, 1 or 2, represented by the synchronization of the stimulus oscillator O_s with the response oscillators O_{r_1} or O_{r_2}. Each circle corresponds to groups of neurons synchronized among themselves, and the lines to connections between each group of neurons.

the following structure:

$$\mathbf{Z}_n \rightarrow \mathbf{S}_n \rightarrow \mathbf{R}_n \rightarrow \mathbf{E}_n \rightarrow \mathbf{Z}_{n+1}. \tag{4.1}$$

Intuitively, a trial n starts with the subject having a given state of conditioning $\mathbf{Z}_n$. Then, a stimulus $s \in S$ is sampled ($\mathbf{S}_n$), and a response $\mathbf{R}_n$ is given according to the sate of conditioning (or randomly, if no conditioning is associated to s). After a response, a reinforcement $\mathbf{E}_n$ event occurs, informing the subject of the correct answer, and this may result (with probability c) to a change in conditioning to this reinforced event, thus leading to a new state of conditioning $\mathbf{Z}_{n+1}$. In other words, learning happens with repeated reinforcement in a probabilistic way by changes in the state of conditioning.

To obtain SR theory in terms of neurons, a distal stimulus is represented in the brain by a set of synchronized neurons, and similarly for responses. Collections of neurons synchronize in phase because of their excitatory connections, and synchronize out of phase because of inhibitory connections [33]. Because we are talking about ensembles of neurons (perhaps thousands), each set stimulus/responses can be described in a first approximation by a periodic function, which for simplicity we assume to be a cosine function. Thus, the basic unit in our model is an oscillator

$$O(t) = A(t) \cos \omega t, \tag{4.2}$$

where $\omega = \omega(t)$ is its time-dependent frequency. Since ω is a function of time, $O(t)$ is determined by the argument of the cosine, i.e. by $\varphi(t) = \omega(t) \cdot t$. Thus, we rewrite this simple oscillator as $O(t) = A(t) \cos \varphi(t)$, and call $\varphi(t)$ the phase of $O(t)$. Firing neurons spike with same amplitude but varying frequencies. Therefore, a collection of firing neurons can be approximately described by $A(t) = A_0$ and $\varphi(t)$, and in our model we assume interactions that affect only the phase.

So, let $O_s(t)$ be a stimulus oscillator given by

$$O_s(t) = A \cos(\omega_0 t) = A \cos(\varphi_s(t)), \tag{4.3}$$

and let

$$O_{r_1}(t) = A \cos(\omega_0 t + \delta\phi_1) = A \cos(\varphi_{r_1}(t)), \tag{4.4}$$

$$O_{r_2}(t) = A \cos(\omega_0 t + \delta\phi_2) = A \cos(\varphi_{r_2}(t)), \tag{4.5}$$

be the two response oscillators (Figure 4.1) .

To describe synchronization, we start with $O_s(t)$ and $O_{r_1}(t)$. When uncoupled their natural frequencies ω_s and ω_{r_1} are constant. From equation (4.2) their uncoupled dynamics satisfy

$$\frac{d\varphi_s}{dt} = \omega_s, \tag{4.6}$$

$$\frac{d\varphi_{r_1}}{dt} = \omega_{r_1}. \tag{4.7}$$

If weakly coupled, their interaction does not affect the sinusoidal character of $O_s(t)$ and $O_{r_1}(t)$, but affects their relative phases, and (4.6) and (4.7) need to include an interaction term. Such term reflects the tendency of phases to either move closer to each other for excitatory synapses or move apart from each other for inhibitory synapses. Then, in first approximation, we have

$$\frac{d\varphi_s}{dt} = \omega_s - k_{s,r_1}\sin\left(\varphi_s - \varphi_{r_1}\right), \tag{4.8}$$

$$\frac{d\varphi_{r_1}}{dt} = \omega_{r_1} - k_{r_1,s}\sin\left(\varphi_{r_1} - \varphi_s\right), \tag{4.9}$$

where k_{ij} are the couplings. To understand where synchronization comes from, let us define

$$\begin{aligned} \varphi'_s &= \varphi_s - \omega_s t, \\ \varphi'_{r_1} &= \varphi_{r_1} - \omega_{r_1} t. \end{aligned}$$

Substituting in (4.8) and (4.9), we have

$$\frac{d\varphi'_s}{dt} = -k_{s,r_1}\sin\left(\left(\varphi'_s - \varphi'_{r_1}\right) + \left(\omega_s - \omega_{r_1}\right)t\right), \tag{4.10}$$

$$\frac{d\varphi'_{r_1}}{dt} = -k_{r_1,s}\sin\left(\left(\varphi'_{r_1} - \varphi'_s\right) - \left(\omega_s - \omega_{r_1}\right)t\right). \tag{4.11}$$

Equations (4.10) and (4.11) have fixed points[17] when

$$\varphi'_{r_1} - \varphi'_s = \delta\omega\, t,$$

or,

$$\varphi_{r_1} = \varphi_s.$$

In other words, (4.8) and (4.9) are stationary when synchronized.

The system (4.8) and (4.9) can be extended to N oscillators, and become

$$\frac{d\varphi_i}{dt} = \omega_i - \sum_{j\neq i} k_{ij}\sin\left(\varphi_i - \varphi_j\right). \tag{4.12}$$

Equations (4.12) are known as Kuramoto equations [72], and they are often used to describe synchronizing systems. Their advantage come from two main points. They can be exactly solved under symmetry assumptions in the limit of large N, providing insight into the nature of emerging synchronization. Second, sets of weakly-coupled oscillating systems can be roughly described by Kuramoto-like equations [57]. In our model, we assume Kuramoto's equations are a good approximation for the dynamics of coupled sets of neural oscillators.

From the oscillators' mathematical description, we can describe how SR theory is modeled by them. The main idea is straightfoward. Once a distal stimulus is presented, an associated ensemble of neurons in activated the brain. Neurons

[17] A fixed point is a point where all derivatives are zero. Their are important points because they represent stationary solutions for the dynamical system. Fixed points can have stationary solutions that are either stable or unstable.

in this ensemble synchronize, and we describe this highly complex system by its average phase. We think of this synchronization as an activation of the stimulus representation in the brain.

Once the stimulus is activated, it may elicit a response by activating synaptically coupled oscillators (in a mechanism that may lead to spreading activation [26]). Similarly to stimuli, responses are represented by ensembles of synchronized neurons. Among the active responses, selection of a particular one is done by the relative phase synchronization between the stimulus oscillator and the selected response. This phase synchronization is determined by the couplings between the stimulus and response oscillators, and the couplings are related to the state of conditioning in SR theory.

The simplest model utilizes three oscillators as introduced above. Once activated, the rate of firings within each response oscillator is due to their own dynamics and also the firings of O_s. Thus, it is reasonable to assume that they interfere, with interference meaning higher coherence when in phase and lower coherence when out of phase. Mathematically, we have, for equal amplitude oscillators, equations (4.3)–(4.5). As with physical oscillators, the mean intensity is a measure of the excitation carried by the oscillations, and at response 1 it is

$$\begin{aligned} I_1 &= \left\langle \left(O_s(t) + O_{r_1}(t)\right)^2 \right\rangle_t \\ &= \left\langle O_s(t)^2 \right\rangle_t + \left\langle O_{r_1}(t)^2 \right\rangle_t + \left\langle 2 O_s(t) O_{r_1}(t) \right\rangle_t , \end{aligned}$$

where $\langle f(t) \rangle_{t_0}$ is the time average of $f(t)$ defined by $\langle f(t) \rangle_{t_0} = \frac{1}{\Delta T} \int_{t_0}^{t_0+\Delta T} f(t)\, dt$ ($\Delta T \gg 1/\omega_0$). We have at once

$$I_1 = A^2 \left(1 + \cos\left(\delta\phi_1\right)\right),$$

and similarly

$$I_2 = A^2 \left(1 + \cos\left(\delta\phi_2\right)\right).$$

Therefore, the intensity for r_1 or r_2 depends on the phase difference between the SR oscillators.

Since I_1 and I_2 are competing responses, the maximum contrast between them happens when one of their relative phases (with respect to the stimulus oscillator) is zero while the other is π. It is standard to normalize the difference $I_1 - I_2$ by the total intensity,

$$b = \frac{I_1 - I_2}{I_1 + I_2}. \tag{4.13}$$

taking values between -1 and 1. The quantity b is called the contrast.

The contrast provides a way to think about a continuum of responses between r_1 and r_2. Assume

$$\delta\phi_1 = \delta\phi_2 + \pi \equiv \delta\phi, \tag{4.14}$$

which yields

$$I_1 = A^2 \left(1 + \cos\left(\delta\phi\right)\right), \tag{4.15}$$

and

$$I_2 = A^2 \left(1 - \cos\left(\delta\phi\right)\right). \tag{4.16}$$

Then, to determine b all we need is $\delta\phi$, as

$$b = \cos\left(\delta\phi\right), \tag{4.17}$$

$0 \leq \delta\varphi \leq \pi$. So, $\delta\phi$ codes a continuum of responses between -1 and 1 or any arbitrary interval (ζ_1, ζ_2) upon rescaling.

The above discussion presents only some aspects of our oscillator model, which was designed to reproduce SR theory. It goes beyond the scope of this paper to describe fully this model, particularly because learning, one of the central features of SR theory, is not relevant to our current purposes of showing quantum-like characteristics in neural oscillators. However, we examine in more detail two of the mathematical components of the oscillator SR model that are relevant to us here: sampling and response.

When a stimulus s_n is sampled, a collection of neurons fire synchronously, corresponding to the activation of a neural oscillator, O_{s_n}. In consonance with SR theory, we assume the activation of s_n in a way that is consistent with the random variable $\mathbf{S}_n$. In other words, from a set of s_n oscillators, we activate only one oscillator with equal probability. This is a stochastic characteristic of the theory that is not part of the dynamics, but is a classical type of stochasticity.

Once s_n is sampled, the active oscillators evolve for the time interval Δt_r, which is selected as a parameter representing the time of response computation. This evolution satisfies Kuramoto's differential equations

$$\frac{d\varphi_i}{dt} = \omega_i - \sum_{i \neq j} k_{ij} \sin\left(\varphi_i - \varphi_j + \delta_{ij}\right), \tag{4.18}$$

where k_{ij} is the coupling constant between oscillators i and j, and δ_{ij} is an anti-symmetric matrix representing phase differences, and i and j can be either O_{s_n}, O_{r_1}, or O_{r_2}. Equation (4.18) can be rewritten as

$$\frac{d\varphi_i}{dt} = \omega_i - \sum_{j} \left[k_{ij}^E \sin\left(\varphi_i - \varphi_j\right) + k_{ij}^I \cos\left(\varphi_i - \varphi_j\right)\right], \tag{4.19}$$

where $k_{ij}^E = k_{ij}\cos\left(\delta_{ij}\right)$ and $k_{ij}^I = k_{ij}\sin\left(\delta_{ij}\right)$, and this has an important physical interpretation: k_{ij}^E corresponds to excitatory couplings, and k_{ij}^I to inhibitory ones. In terms of those couplings, the evolution equation is

$$\frac{d\varphi_i}{dt} = \omega_i - \sum_{i \neq j} \left[k_{i,j}^E \sin\left(\varphi_i - \varphi_j\right) - k_{i,j}^I \cos\left(\varphi_i - \varphi_j\right)\right], \tag{4.20}$$

where ω_i is the oscillator's natural frequency. The solutions to (4.20) and the initial conditions randomly distributed at activation give us the phases at time $t_{r,n} = t_{s,n} + \Delta t_r$. The coupling strengths between oscillators determine their relative phase locking, which in turn corresponds to the computation of a given response, according to equation (4.13). The couplings are determined by reinforcement, but here we assume the values are given for each experimental condition (see [85] for details).

At this point the attentive reader may have guessed where quantum-like contextuality come from: the interference of two neural oscillators in (4.17). To see how interference renders quantum-like results, let us consider the following case discussed in details in [29]. Imagine that instead of a single stimulus, O_s, we have two stimuli, O_{s_1} and O_{s_2} which can be activated separately or simultaneously. The activation of stimulus O_{s_1} leads to a response (contrast) b_1 when

$$k^E_{s_1,r_1} = k^E_{r_1,s_1} = \alpha b_1 = -k^E_{s_1,r_2} = -k^E_{r_2,s_1}, \tag{4.21}$$

$$k^E_{r_1,r_2} = k^E_{r_2,r_1} = -\alpha, \tag{4.22}$$

and

$$k^I_{s_1,r_1} = k^I_{r_2,s_1} = \alpha\sqrt{1-b_1^2} = -k^I_{s_1,r_2} = -k^I_{r_1,s_1}, \tag{4.23}$$

$$k^I_{r_1,r_2} = k^I_{r_2,r_1} = 0, \tag{4.24}$$

where α is a convergence to synchronization parameter (the larger the α, the faster it converges). A similar set of couplings can be obtained for the other stimulus oscillator O_{s_2} if we require it to answer b_2.

Now, from (4.20) and couplings (4.21)–(4.24), the system is deterministic. However, the initial conditions are not the same at every trial, and if we assume a Gaussian distribution of initial phases at each trial, the responses given to the stimulus O_{s_1} will vary around the value b_1. To code a discrete response, such as a ± 1-valued random variable $\mathbf{A}$ we say the outcome of a random variable $\mathbf{A}$ is $+1$ if the response b_1 is greater or equal to 0.5, and -1 if the response is lesser than 0.5, and we interpret the value $+1$ as an action being preferred over no action. Then, if we carefully chose the parameters in (4.21)–(4.24) such that b_1 is slightly greater than 0.5, then $\mathbf{A}$ would be $+1$ with higher probability than -1. Thus we could say that an action is preferred given stimulus O_{s_1}. We could do the same type of setup for stimulus O_{s_2}, such that whenever this stimulus is presented an action is also preferred.

The oscillator case above is equivalent to the example presented in Section 2 if we think of the two distinct stimuli O_{s_1} and O_{s_2} as corresponding to "won first bet" and "lost first bet," respectively, and $\mathbf{A} = 1$ as "accept second gamble" and $\mathbf{A} = -1$ as "reject second gamble." If O_{s_1} and O_{s_2} are inconsistent stimuli, the violation of the STP comes from the probabilities of response for such oscillator model when *both* oscillators are activated (in case of lack of knowledge), and the interference effects of the oscillations lead to the nonmonotonicity of probabilistic outcomes [29]. In other words, because of interference, neural oscillator models may exhibit contextual quantum-like features.

A natural question now arises from our oscillator model. Since QM brings so many features in addition to stochasticity and contextuality[18], it is worth investigating whether there are violations of CP from our neural model that cannot be described by QM. A natural starting point is the three random variable example, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, given in Section 3. It is straightforward to prove that a Hilbert space description of three observables, represented by the Hermitian operators X, Y, and Z, where we can observe them in a pairwise fashion implies that we can observe all three simultaneous. In other words, if $[X,Y] = [X,Z] = [Y,Z] = 0$, then there exists a basis where X, Y, and Z are simultaneously diagonal. This means that if we can concoct an experiment to measure X and Y together, another to measure Y and Z, and yet another to measure X and Z, quantum mechanics predicts it to be possible to create an experiment where all three observables, X, Y, and Z are measured simultaneously. Since a simultaneous measure of three observables is

[18]Non-locality, as we talked about in Section 3, is one prominent case, but there are many non-trivial results in QM that bear no clear connection to social systems, such as the no-cloning theorem [37], or the monogamy of entanglement [92], to mention a few.

a guarantee of the existence of a joint probability distribution (by simply counting how many times each elementary event shows up), the three random-variable example provided in Section 3 cannot be described by QM[19].

However, as we showed in [28, 31], in a more complicated three stimulus and six response oscillators, there are couplings between oscillators that give a higher probability of anti-correlation between pairwise activations of stimuli. For strong enough anti-correlations, there are no joint probability distribution, as Suppes and Zanotti [87] proved that a joint probability exist iff

$$\begin{aligned} -1 &\leq E(\mathbf{XY}) + E(\mathbf{XZ}) + E(\mathbf{YZ}) \\ &\leq 1 + 2\min\{E(\mathbf{XY}), E(\mathbf{XZ}), E(\mathbf{YZ})\}. \end{aligned} \tag{4.25}$$

Thus, there are neural oscillator models that exhibit a type of contextuality that cannot be modeled by QM.

In this Section we presented a neural oscillator model that reproduces not only SR theory, but also displays the nonmonotonicity associated with contextual quantum-like behavior. We also showed that such model poses difficulties for quantum descriptions, as it implies the theoretical existence of systems that would not be describable by QM. In the next Section we introduce an alternative stochastic model that we believe could be a natural replacement for the quantum formalism in such cases where QM is not applicable.

5. Negative Probabilities

In this Section we introduce the idea of negative probabilities as a way to describe certain contextual stochastic processes. Historically, negative probabilities (NP) were first encountered in QM, when Wigner attempted to produce a joint probability distribution for momentum and position that would give the same outcomes as quantum statistical mechanics (for a somewhat old review, see [75]). Wigner dismissed NP as meaningless, and called them quasi-probability distributions[20]. Later, Dirac used NP to approach problems in quantum electrodynamics [38], and Feynman use them to describe the two-slit and spin [46]. Dirac and Feynman's views were similar to Wigner, but they thought of NP as a nice accounting tool that could perhaps be as useful as negative numbers in mathematics. However, not all quantum mechanical setups allow for negative probabilities (e.g. the two-slit experiment can be shown to allow for negative probabilities only under certain counterfactual reasoning [34, 35]). We emphasize that here NP always mean that a joint probability distribution takes negative values for non-observable events (such as joint values of position and momentum), but is always non-negative for observable events.

[19]This is a point mentioned by Kochen in [69], but in [30] we showed that by increasing the Hilbert space and adding a fourth variable corresponding to context, we can artificially reproduce the correlations that violate a joint probability distribution.

[20]Perhaps very much in the same way that mathematicians had problems with negative numbers. For instance, as late as the 1800's, the famous mathematician Augustus De Morgan, stated the following [74, pg. 72]. "Above all, he [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."

Before we delve further into our discussion of NP, we formally define it (from now on we follow [35]). We start with a preliminary definition related to marginal expectations that are observable.

Definition 3. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, and let $(\Omega_i, \mathcal{F}_i, p_i)$, $i = 1, \ldots, n$, a set of n probability spaces, $\mathcal{F}_i \subseteq \mathcal{F}$ and $\Omega_i \subseteq \Omega$. Then $(\Omega, \mathcal{F}, p)$,where p is a real-valued function, $p : \mathcal{F} \to [0,1]$, $p(\Omega) = 1$, is *compatible* with the probabilities p_i's iff

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

Furthermore, the marginals p_i are *viable* iff p is a probability measure.

Intuitively, we can think of the p_i's as observable marginal probabilities on subspaces of a larger sample space Ω. Then such marginals are *viable*[21] if it is possible to "sew" them together to produce a larger probability function over the whole Ω [42, 39, 32, 35].

As mentioned, in QM the marginals are not always viable, but are compatible with a real-valued function p that has the characteristic of being somewhere negative. This motivates the following definition.

Definition 4. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, P and P' real-valued functions, $P : \mathcal{F} \to \mathbb{R}$, $P' : \mathcal{F} \to \mathbb{R}$, and let $(\Omega_i, \mathcal{F}_i, p_i)$, $i = 1, \ldots, n$, a set of n probability spaces, $\mathcal{F}_i \subset \mathcal{F}$ and $\Omega_i \subseteq \Omega$. Then $(\Omega, \mathcal{F}, P)$ is a negative probability space, and P a negative probability, if and only if $(\Omega, \mathcal{F}, P)$ is compatible with the probabilities p_i's and

$$\begin{array}{ll} \text{N1.} & \forall \left(P'\right) \left(\sum_{\omega_i \in \Omega} \left|P\left(\{\omega_i\}\right)\right| \leq \sum_{\omega_i \in \Omega} \left|P'\left(\{\omega_i\}\right)\right| \right) \\ \text{N2.} & \sum_{\omega_i \in \Omega} P\left(\{\omega_i\}\right) = 1 \\ \text{N3.} & P\left(\{\omega_i, \omega_j\}\right) = P\left(\{\omega_i\}\right) + P\left(\{\omega_j\}\right), \quad i \neq j. \end{array}$$

In this definition we replaced Kolmogorov's nonnegativity axiom with a minimization of the L1 norm of P. There is an intuitive reason to do so: we seek a quasi-probability distribution that is as close to a proper distribution as possible. This departure from a proper norm is the motivation for the following definition.

Definition 5. Let $(\Omega, \mathcal{F}, P)$ be a negative probability space. Then, the *minimum L1 probability norm*, denoted M^*, or simply *minimum probability norm*, is given by $M^* = \sum_{\omega_i \in \Omega} \left|P\left(\{\omega_i\}\right)\right|$.

In [35] we proved that P is a probability (and therefore $(\Omega, \mathcal{F}, P)$ is a probability space) if and only if $M^* = 1$. Since M^* can be greater than one for systems with negative probability, and since negative probabilities come from the impossibility of defining a proper probability distribution that can put together the different marginals, we interpret M^* as a measure of contextuality. In other words, not only does the existence of NP lead to contextuality, but the more they depart from a proper distribution the more there contextuality there is [32].

An important result for negative probabilities relies on the following definition [35].

---

[21] A term coined by [53].

Definition 6. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, and let $(\Omega_i, \mathcal{F}_i, p_i)$, $i = 1, \ldots, n$, a collection of n probability spaces, $\mathcal{F}_i \subseteq \mathcal{F}$ and $\Omega_i \subseteq \Omega$. Then the probabilities p_i are *contextually biased*[22] if there exists an a in $\mathcal{F}_i$ and in $\mathcal{F}_j$, $i \neq j$, $b \neq a \neq b'$, $\sum_{\forall b \in \mathcal{F}_j} p\,(a \cap b) \neq \sum_{\forall b' \in \mathcal{F}_i} p\,(a \cap b')$.

In references [1, 6, 76, 73] it was independently proven that NP (in the sense we use above) exist if and only if the marginals p_i are not contextually biased. Thus, it follows that for many systems where proper joint probability distributions cannot be defined, we can still define NP if such systems are not contextually biased. Another way is to say that a collection of probabilities p_i are compatible if and only if they are not contextually biased.

We now present an example of a nontrivial application of negative probabilities to decision making [30]. In this example, Deana is a decision maker who wants to bet on the stock market (well, some "simple" version of it). She she wants to invest in three companies, creatively named here X, Y, and Z. Since she knows nothing about X, Y, and Z, she contacts three "experts," Alice, Bob, and Carlos, who provide her with expected outcomes of X, Y, and Z. However, each expert is specialized only on two of the companies, but not all. Furthermore, perhaps because of a bias, experts may give information that is inconsistent. For example, say we create the following ± 1-valued random variables, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, corresponding to their beliefs of a stock value going up if $+1$ and down if -1. Our experts all agree that the probabilities of stocks of X, Y, and Z going up are the same as going down, and therefore we can say that

$$E\,(\mathbf{X}) = E\,(\mathbf{Y}) = E\,(\mathbf{Z}) = 0. \tag{5.1}$$

But since Alice only knows about X and Y, she can only tell us that her belief is

$$E_A\,(\mathbf{XY}) = -1, \tag{5.2}$$

where we put a subscript on the expectation to emphasize that it is Alice's subjective belief[23]. Equation (5.2) has the simple interpretation: Alice believes that if X's stocks go up/down then Y's will go down/up with certainty. Bob's and Carlos's beliefs are that

$$E_B\,(\mathbf{XZ}) = -\frac{1}{2}, \tag{5.3}$$

and

$$E_C\,(\mathbf{YZ}) = 0. \tag{5.4}$$

It is easy to see from (4.25) that (5.2)–(5.4) are not viable, but (5.1) imply the probabilities that lead to Alice, Bob, and Carlos's expectations are compatible. Therefore, there exists a negative probability distribution consistent with (5.1)–(5.4).

What is Deana to do with the inconsistent information she got from Alice, Bob, and Carlos? A standard approach is to start with a prior distribution and use their information to update the posterior using Bayes's theorem. However, as demonstrated in [30], such approach has a shortcoming: it does not tell us anything new

[22]Here we adopt and adapt the terminology of [39].

[23]An example of inconsistent information like the one we present here is not easily translatable into objective interpretations of probabilities.

about the triple moment $E(\mathbf{XYZ})$. The Bayesian approach does not update the triple moment, and its value comes purely from Deana's prior.

The lack of update for the triple moment presents a difficulty. To show it, take the deterministic (and consistent with a proper joint probability) case where Deana was given $E_A(\mathbf{XY}) = E_B(\mathbf{XZ}) = E_C(\mathbf{YZ}) = 1$. It is immediately clear from these correlations that $E(\mathbf{XYZ}) = 1$. So, since the experts are *not* disagreeing, and since their judgment leads to a specific value for the triple moment, why should Deana not take this into account? Why should her bet on the triple moment be related simply to her prior on it? This seems to be a failure of the Bayesian approach. The situation is different for NP. Because we assume that a joint quasi-probability distribution exists (albeit negative) and that the best joints (as they are not unique) minimize the L1 norm, as they are the closest to a "rational" and consistent joint, then we are constrained to only the best joints. In the deterministic case of 1 correlations, this leads to the correct prediction that the triple moment is 1.

The minimization of the L1 norm also has a consequence for the inconsistent pairwise expectations (5.2)–(5.4). It restricts the possible values for the triple moments to the range [30]

$$-\frac{1}{4} \leq E(\mathbf{XYZ}) \leq \frac{1}{2}.$$

Given that the Bayesian approach provides no information to Deana, it should possible to devise a Dutch book between NP and Bayesian approaches for certain situations[24]. NP provides normative information that goes beyond the Bayesian approach.

The situation is a little better between NP and QC. As we mentioned, the $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ example is only describable via QM with supplementary assumptions, as done in [30], where an extra dimension to the Hilbert space was added corresponding to the internal states of belief of Alice, Bob, and Carlos. However, the triple moment correlation needs to be explicitly given in the state vector, and there are no arguments to limit its values. So, QC is in better shape than the Bayesian approach because even though it does not provide an advantage over the other approaches, it at least makes it explicit that the triple moment are included *ad hoc*.

We end this Section with some comments about the meaning of NP. In this paper we take Feynman and Dirac's views: NP are a useful accounting tool. However, there are ways to interpret them. For example, Khrennikov showed that in the frequentist interpretation of von Mises, negative probabilities appear when we have sequences in the usual Archimedian metric that violate the principle of stabilization, and therefore do not converge to a specific probability value[25]. In those cases, a p-adic metric makes such sequences convergent, and negative probabilities appear as the p-adic limiting case [61, 62, 63, 64, 65, 66]. Abramsky and Brandenburger [1, 2] interpret NP in the context of sheaf theory. For them NP comes from two independent types of events belonging to different types. One type of event erases recordings of the other type, and this allows for the observed correlations. Finally, closely related to Abramsky and Brandenburger's, is Szekely's interpretation, who thinks of negative probabilities P as related to a proper probability p via a convolution equation $P * f = p$, which always exists [80, 88]. This convolution means

[24] A Dutch Book is the name given to a strategy that would allow one of the gamblers to win for sure over the other gamblers in a game [7].

[25] Pseudo-random sequences may have this property.

that for a random variable $\mathbf{X}$ whose probability distribution is P, there exists two other random variables, $\mathbf{X}_+$ and $\mathbf{X}_-$ with proper probability distributions (p and f, respectively) and such that $\mathbf{X} = \mathbf{X}_+ - \mathbf{X}_-$. Our interpretation can be though as subjective: NP are an accounting tools, but provide us the best subjective information about systems who do not have an objective probability distribution, as it is the closest distribution to a proper one (via normalization of the L1 norm).

6. Final Remarks

In this paper we discussed how some of the well-known examples used in QC are connected to the contextuality of the two-slit experiment. A neural oscillator model was introduced based on reasonable neurophysiological assumptions that reproduce behavioral SR theory and the nonmonotonic character of QC. Such a neural model produces outcomes for certain situations that are not naturally modeled by QM, as in the case with six-response oscillators. However, such examples could be modeled by NP. More importantly, NP did not only provide a way to describe such systems, but was also normative.

QC comes from the idea that human decision making is better describable by the mathematics of quantum mechanics, with its probability associated to density operators in a Hilbert space. However, there are possible situations where QC is not appropriate, such as the $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ example. Furthermore, we saw that the $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ example shows up in neural oscillator models that reproduce standard SR theory, but also in decision-making situations. Therefore, as an extended probability theory, QC is too restrictive, leaving out perhaps important situations. Furthermore, QC is mostly descriptive, not offering, as far as we are aware, any normative power. We contrast this with NP, which describes many of the QC systems (those with compatible probabilities), but also those created by inconsistent oscillators or inconsistent information. Given how NP offers normative information via the minimization of the L1 norm, which is computationally simple for biological (as well as computer) systems, perhaps it is not unreasonable to hypothesize that such processes actually happen in our brain. This, we believe is an exciting perspective, and we hope to further investigate it in the future.

Bayesians have problems with not updating their triple moment, even faced with indirect information about them. This suggests the existence of a Dutch Book. An interesting question is how such a Dutch Book could be constructed. For example, if Alice, Bob, and Carlos are subject to confirmation biases, could we model it (similar to Fine's prism model [49, 50]) and show that NP outperforms Bayesianism? Furthermore, if our L1 norm hypothesis for the brain is correct, wouldn't human decision makers unconsciously follow a strategy that would win bets with "rational" Bayesians? These questions also present a research program that we believe will be fruitful.

References

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

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

[3] Diederik Aerts. Quantum structure in cognition. *Journal of Mathematical Psychology*, 53(5):314–348, October 2009.
[4] Diederik Aerts, Jan Broekaert, Liane Gabora, and Tomas Veloz. The guppy effect as interference. In Jerome R. Busemeyer, François Dubois, Ariane Lambert-Mogiliansky, and Massimo Melucci, editors, *Quantum Interaction*, number 7620 in Lecture Notes in Computer Science, pages 36–47. Springer Berlin Heidelberg, 2012.
[5] Diederik Aerts and Sandro Sozzo. Quantum entanglement in concept combinations. arXiv e-print 1302.3831, February 2013.
[6] S.W. Al-Safi and A.J. Short. Simulating all nonsignaling correlations via classical or quantum theory with negative probabilities. *Physical Review Letters*, 111(17):170403, October 2013.
[7] Paul Anand, Prasanta Pattanaik, and Clemens Puppe, editors. *The Handbook of Rational and Social Choice: an overview of new foundations and applications*. Oxford University Press, Oxford, England, 2009.
[8] Mehrdad Ashtiani and Mohammad Abdollahi Azgomi. A survey of quantum-like approaches to decision making and cognition. *Mathematical Social Sciences*, 2015.
[9] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of bell's inequalities using time- varying analyzers. *Physical Review Letters*, 49(25):1804–1807, December 1982.
[10] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental tests of realistic local theories via bell's theorem. *Physical Review Letters*, 47(7):460–463, August 1981.
[11] H. Atmanspacher and H. Römer. Order effects in sequential measurements of non-commuting psychological observables. *Journal of Mathematical Psychology*, 56(4):274–280, August 2012.
[12] J.S. Bell. On the einstein-podolsky-rosen paradox. *Physics*, 1(3):195–200, 1964.
[13] J.S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38(3):447–452, 1966.
[14] J.S. Bell. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, June 2004.
[15] George Boole. *An Investigation of the Laws of Thought: On which are Founded the Mathematical Theories of Logic and Probabilities*. Dover Publications, 1854.
[16] Rachael Briggs. Normative theories of rational choice: Expected utility. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition, 2014.
[17] Jerome R. Busemeyer and Adele Diederich. *Cognitive modeling*. SAGE, 2010.
[18] Jerome R. Busemeyer, Emmanuel M. Pothos, Riccardo Franco, and Jennifer S. Trueblood. A quantum theoretical explanation for probability judgment errors. *Psychological Review*, 118(2):193–218, 2011.
[19] Jerome R. Busemeyer, Zheng Wang, Andrei Khrennikov, and Irina Basieva. Applying quantum principles to psychology. *Physica Scripta*, 2014(T163):014007, December 2014.
[20] Jerome R. Busemeyer, Zheng Wang, and James T. Townsend. Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*, 50(3):220–241, June 2006.
[21] J.R. Busemeyer and P.D. Bruza. *Quantum models of cognition and decision*. Cambridge Univ. Press, Cambridge, 2012.
[22] A. Cabello. Quantum physics: Correlations without parts. *Nature*, 474(7352):456–458, June 2011.
[23] A. Cabello. Simple explanation of the quantum violation of a fundamental inequality. *Physical Review Letters*, 110(6):060402, February 2013.
[24] A. Cabello. The exclusivity principle singles out the quantum violation of the bell inequality. *arXiv:1406.5656 [quant-ph]*, June 2014. arXiv: 1406.5656.
[25] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, October 1969.
[26] Allan M. Collins and Elizabeth F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975.
[27] Richard T. Cox. *The Algebra of Probable Inference*. The John Hopkins Press, Baltimore, 1961.
[28] J. Acacio de Barros. Joint probabilities and quantum cognition. In A. Khrennikov, A. L Migdall, S. Polyakov, and H. Atmanspacher, editors, *AIP Conference Proceedings*, volume 1508, pages 98–107, Vaxjo, Sweden, December 2012. American Institute of Physics.
[29] J. Acacio de Barros. Quantum-like model of behavioral response computation using neural oscillators. *Biosystems*, 110(3):171–182, December 2012.

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

[31] J. Acacio de Barros. Beyond the quantum formalism: Consequences of a neural-oscillator model to quantum cognition. In Hans Liljenström, editor, *Advances in Cognitive Neurodynamics (IV)*, Advances in Cognitive Neurodynamics, pages 401–404. Springer Netherlands, 2015.

[32] J. Acacio de Barros, E.N. Dzhafarov, J.V. Kujala, and G. Oas. Unifying two methods of measuring quantum contextuality. *arXiv:1406.3088 [quant-ph]*, June 2014. arXiv: 1406.3088.

[33] J. Acacio de Barros and G. Oas. Response selection using neural phase oscillators. In Colleen E. Crangle, Adolfo Garcia de la Sienra, and Helen E. Logino, editors, *Foundations and Methods from Mathematics to Neuroscience: Essays Inspired by Patrick Suppes*. CSLI Publications, Stanford University, Stanford, California, 2014.

[34] J. Acacio de Barros and Gary Oas. Negative probabilities and counter-factual reasoning in quantum cognition. *Physica Scripta*, T163:014008, 2014.

[35] J. Acacio de Barros, Gary Oas, and Patrick Suppes. Negative probabilities and counterfactual reasoning on the double-slit experiment. In J.-Y. Beziau, D. Krause, and J.B. Arenhart, editors, *Conceptual Clarification: Tributes to Patrick Suppes (1992-2014)*. College Publications, London, 2015.

[36] J. Acacio de Barros and Patrick Suppes. Quantum mechanics, interference, and the brain. *Journal of Mathematical Psychology*, 53(5):306–313, October 2009.

[37] D. Dieks. Communication by EPR devices. *Physics Letters A*, 92(6):271–272, November 1982.

[38] P.A.M. Dirac. {B}akerian lecture. {T}he physical interpretation of quantum mechanics. *Proceedings of the Royal Society of London B*, A180:1–40, 1942.

[39] Ehtibar N. Dzhafarov and Janne V. Kujala. Contextuality in generalized klyachko-type, bell-type, and leggett-garg-type systems. *arXiv:1411.2244 [physics, physics:quant-ph]*, November 2014. arXiv: 1411.2244.

[40] Ehtibar N. Dzhafarov, Janne V. Kujala, and Jan-Åke Larsson. Contextuality in three types of quantum-mechanical systems. *arXiv:1411.2244 [physics, physics:quant-ph]*, November 2014. arXiv: 1411.2244.

[41] E.N. Dzhafarov and J.N. Kujala. A qualified kolmogorovian account of probabilistic contextuality. *Lecture Notes in Computer Science*, 8369:201–212, 2014.

[42] E.N. Dzhafarov and J.V. Kujala. All-possible-couplings approach to measuring probabilistic context. *PLoS ONE*, 8(5):e61712, May 2013.

[43] E.N. Dzhafarov and J.V. Kujala. Random variables recorded under mutually exclusive conditions: Contextuality-by-default. *arXiv:1309.0962 [quant-ph]*, September 2013.

[44] E.N. Dzhafarov and J.V. Kujala. Contextuality is about identity of random variables. *Physica Scripta*, 2014. In Press. arXiv:1405.2116 [quant-ph].

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

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

[47] R.P. Feynman, R.B. Leighton, and M.L. Sands. *The Feynman lectures on physics: Mainly mechanics, radiation, and heat*, volume 3. Basic Books, New York, NY, 2011.

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

[49] A. Fine. Some local models for correlation experiments. *Synthese*, 50(2):279–294, 1982.

[50] Arthur Fine. *The Shaky Game*. University of Chicago Press, February 2009.

[51] M. C Galavotti. *Philosophical introduction to probability*, volume 167 of *CSLI Lecture Notes*. CSLI Publications, Stanford, CA, 2005.

[52] O. Gühne, C. Budroni, A. Cabello, M. Kleinmann, and J.-Å. Larsson. Bounding the quantum dimension with contextuality. *Physical Review A*, 89(6):062107, June 2014.

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

[54] E. Haven and A. Khrennikov. *Quantum Social Science*. Cambridge Univ. Press, Cambridge, 2013.

[55] Emmanuel Haven and Andrei Khrennikov. A brief introduction to quantum formalism. In Emmanuel Haven and Andrei Khrennikov, editors, *The Palgrave Handbook of quantum models in social science: applications and grand challenges*. Palgrave MacMillan, 2015.
[56] F. Holik, M. Saenz, and A. Plastino. A discussion on the origin of quantum probabilities. *Annals of Physics*, 340(1):293–310, January 2014.
[57] Eugene M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Cambridge, Massachusetts, 2007.
[58] E. T Jaynes. *Probability theory: the logic of science*. Cambridge Univ Press, Cambridge, Great Britain, 2003.
[59] Richard Jeffrey. *Probability and the Art of Judgment*. Cambridge University Press, Cambridge, UK, 1992.
[60] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.
[61] A. Khrennikov. p-adic probability theory and its applications. {T}he principle of statistical stabilization of frequencies. *Theoretical and Mathematical Physics*, 97(3):1340–1348, December 1993.
[62] A. Khrennikov. p-adic statistical models. *Doklady Akademii Nauk*, 330(3):300–304, 1993.
[63] A. Khrennikov. Statistical interpretation of p-adic valued quantum field theory. 1993.
[64] A. Khrennikov. Discrete {Q}p-valued probabilities. *Russian Academy of Sciences-Doklady Mathematics-AMS Translation*, 48(3):485–490, 1994.
[65] A. Khrennikov. *p-Adic Valued Distributions in Mathematical Physics*, volume 309 of *Mathematics and Its Applications*. Springer Science+Business Media, Dordrecht, Holland, 1994.
[66] A. Khrennikov. *Interpretations of probability*. Walter de Gruyter, Berlin, 2009.
[67] A. Khrennikov. *Ubiquitous Quantum Structure*. Springer Verlag, Heidelberg, 2010.
[68] Andrei Khrennikov, Irina Basieva, Ehtibar N. Dzhafarov, and Jerome R. Busemeyer. Quantum models for psychological measurements: An unsolved problem. *PLoS ONE*, 9(10):e110909, October 2014.
[69] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. In C.F. Hooker, editor, *The Logico-Algebraic Approach to Quantum Mechanics*, pages 293–328. D. Reidel Publishing Co., Dordrecht, Holland, 1975.
[70] Simon Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87, 1967.
[71] A.N. Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Co., Oxford, England, 2nd edition, 1956.
[72] Yoshiki Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Dover Publications, Inc., Mineola, New York, 1984.
[73] Elena R. Loubenets. Context-invariant quasi hidden variable (qHV) modelling of all joint von neumann measurements for an arbitrary hilbert space. *Journal of Mathematical Physics*, 56(3):032201, March 2015.
[74] Augustus De Morgan. *On the Study and Difficulties of Mathematics*. Open Court Publishing Company, 1910.
[75] G. Mückenheim. A review of extended probabilities. *Physics Reports*, 133(6):337–401, 1986.
[76] G. Oas, J. Acacio de Barros, and C. Carvalhaes. Exploring non-signalling polytopes with negative probability. *Physica Scripta*, T163:014034, 2014.
[77] A. Pais. *Inward bound: of matter and forces in the physical world*. Oxford University Press, Oxford, UK, 1986.
[78] A. Peres. *Quantum Theory: Concepts and Methods*. Springer Science & Business Media, September 1995.
[79] E. M Pothos and J. R Busemeyer. A quantum probability explanation for violations of 'rational' decision theory. *Proceedings of the Royal Society B: Biological Sciences*, 276(1665):2171–2178, 2009.
[80] Imre Z. Ruzsa and Gábor J. Székely. Convolution quotients of nonnegative functions. *Monatshefte für Mathematik*, 95(3):235–239, September 1983.
[81] L. J Savage. *The foundations of statistics*. Dover Publications Inc., Mineola, New York, 2nd edition, 1972.
[82] E. Shafir and A. Tversky. Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, 24(4):449–474, October 1992.

[83] P. Suppes and R. C Atkinson. *Markov learning models for multiperson interactions.* Stanford University Press, Stanford, California, 1960.

[84] P. Suppes and J. Acacio de Barros. Photons, billiards and chaos. In Paul Weingartner and Gerhard Schurz, editors, *Law and Prediction in the Light of Chaos Research*, volume 473, pages 189–201. Springer Berlin Heidelberg, 1996.

[85] P. Suppes, J. Acacio de Barros, and G. Oas. Phase-oscillator computations as neural models of stimulus–response conditioning and response selection. *Journal of Mathematical Psychology*, 56(2):95–117, April 2012.

[86] Patrick Suppes. *Representation and Invariance of Scientific Structures*. CSLI Publications, Stanford, California, 2002.

[87] Patrick Suppes and Mario Zanotti. When are probabilistic explanations possible? *Synthese*, 48(2):191–199, 1981.

[88] Gábor J. Székely. Half of a coin: negative probabilities. *Wilmott Magazine*, 50:66–68, 2005.

[89] Jennifer S. Trueblood and Jerome R. Busemeyer. A quantum probability account of order effects in inference. *Cognitive Science*, 35(8):1518–1552, November 2011.

[90] A. Tversky and E. Shafir. The disjunction effect in choice under uncertainty. *Psychological Science*, 3(5):305–309, September 1992.

[91] Zheng Wang and Jerome R. Busemeyer. A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, 5(4):689–710, October 2013.

[92] Dong Yang. A simple proof of monogamy of entanglement. *Physics Letters A*, 360(2):249–250, December 2006.

*School of Humanities and Liberal Studies, San Francisco State University, San Francisco, CA; †Stanford Pre-Collegiate Studies, Stanford University, Stanford, CA