THE LOGIC IN COMPUTER SCIENCE COLUMN

BY

YURI GUREVICH

Computer Science and Engineering
University of Michigan, Ann Arbor, MI 48109, USA
gurevich@umich.edu

Negative probabilities, II
What they are and what they are for

Andreas Blass and Yuri Gurevich

What are numbers and what are they for? [1]
— Richard Dedekind, 1888

Abstract

A signed probability distribution may extend a given traditional probability from observable events to all events. We formalize and illustrate this approach. We also illustrate its limitation. We argue that the right question is not what negative probabilities are but what they are for.

1 Introduction

The idea of negative probabilities arose in quantum mechanics [20, 13, 8, 10]. This is not surprising. The weirdness of quantum mechanics required bold fresh ideas. The physicists have been using signed probability distributions primarily in connection with phase spaces for quantum systems [23] as suggested by Wigner in his 1932 paper [20].

The present article was provoked by the paper "An operational interpretation of negative probabilities and no-signaling models" by Samson Abramsky and Adam Brandenburger [1]. It occurred to us that the right question may not be how to interpret negative probabilities but how to employ them.

Not everybody thinks that the idea of negative probabilities is a good one. Let's hear a critic. In 1935, Albert Einstein, Boris Podolsky and Nathan Rosen surmised that quantum mechanics is incomplete, i.e., that some hidden variables are missing [9]. John Bell proved that local hidden-variable theories contradict quantum-mechanical statistics [3]. It seems natural to try to save the hidden-variable approach by means of negative probabilities, and here is what our critic, Itamar Pitowsky, has to say about that [17, page 148].

---

[1] Was sind und was sollen die Zahlen?

> What makes the classical hidden variable theories "classical" is the identification of "mixtures of hidden variable states" as probability measures... The logical step to take ... is ... to use (cheap) tricks such as negative "probability", or even complex "probability" values. Formally we may be able to "solve" our problem, but then the term "probability" loses completely its meaning. ... It is absurd to talk about an urn containing -17 red balls or $3e^{i\pi/12}$ wooden balls.

We wrote about negative probabilities once [5]. Here we wish to defend the cheap trick of negative probabilities. Indeed, the standard frequential interpretation of probabilities does not apply to negative probabilities. But here is an appetizing analogy. For a long time, the standard interpretation of numbers was quantity. It is absurd to talk about the quantity of $3e^{i\pi/12}$. Formally the complex numbers allow us to "solve" equations like $x^2 = -1$ but, one might argue, the term "number" loses its meaning.

The mathematical trick of introducing complex numbers paid off richly. Some hard number-theoretic problems have been solved using the methods of complex analysis. Eventually it even became possible to give a physical meaning to complex numbers, e.g., a complex number $a + bi$ can be interpreted as impedance where the real part a is resistance and the imaginary part b is the reactance [21], and complex numbers of absolute value ≤ 1 can be interpreted as quantum amplitudes.

The original purpose of complex numbers was to solve certain algebraic equations with real coefficients. What is the corresponding purpose of negative probabilities? What plays the role of algebraic equations? These are the questions addressed in this paper. Our proposal is admittedly — and provably — limited. But it supports some of the usages of negative probabilities in the literature. One has to start somewhere.

For simplicity and to separate concerns, we restrict attention to finite spaces.

2   Signed probability spaces

We use, as a running example, a scenario due to Piponi [16], which "while artificial, is appealingly simple, and does convey some helpful intuitions" [1].

Example (Piponi's scenario). A machine produces boxes with pairs (l, r) of bits, each bit viewable through its own door. Somehow it is also possible to test whether the two bits are equal. The probability of each possible combination of two bits is

given by the following table:

00	01	10	11
$-1/2$	$1/2$	$1/2$	$1/2$

While the table is not available to the observer, the following three experiments are available.

1. Look through the left door. This allows you to find out eventually that $\mathcal{P}(l = 1) = 1$ and $\mathcal{P}(l = 0) = 0$.

2. Look through the right door. This allows you to find out eventually that $\mathcal{P}(r = 1) = 1$ and $\mathcal{P}(r = 0) = 0$.

3. Test whether the two bits are equal. This allows you to find out eventually that $\mathcal{P}(l \neq r) = 1$ and $\mathcal{P}(l = r) = 0$.

Notice that these six discovered probabilities are all nonnegative (as in traditional probability theory), that they match the probabilities computed from the table above, but that they are not mutually consistent in traditional probability theory. ◄

The following definition reflects our intent to work with finite spaces.

Definition 1. By a *signed probability space S* we mean a pair $(\Omega, \mathcal{P})$ where Ω is a nonempty set and $\mathcal{P}$ is a real-valued function on 2^Ω such that the following probability laws hold.

PL1. $\mathcal{P}(\Omega) = 1$.

PL2. If $e_1, e_2 \subseteq \Omega$ and $e_1 \cap e_2 = \emptyset$ then $\mathcal{P}(e_1 \cup e_2) = \mathcal{P}(e_1) + \mathcal{P}(e_2)$.

If, in addition, we have

PL3. $\mathcal{P}(e) \geq 0$ for all $e \subseteq \Omega$,

then $\mathcal{P}$ and S are *traditional*. ◄

Terminology. Here Ω is the *sample space*, and its elements are *sample points* or *outcomes*. Subsets of Ω are *events*. $\mathcal{P}$ is a *signed probability distribution*. $\mathcal{P}(e)$, even if it is negative, is called the *probability* of e. For brevity, when $\omega \in \Omega$, we write $\mathcal{P}(\omega)$ to mean $\mathcal{P}(\{\omega\})$.

Notation. The complement $\Omega - e$ of an event e will be denoted $\bar{e}$. The collection of all subsets of a set s will be denoted 2^s. Disjoint union of sets s_1, s_2 will be denoted $s_1 + s_2$. If S is a set of sets then $\bigcup S = \bigcup_{s \in S} s$.

In the example above, the sample space Ω consists of the four binary strings $00, 01, 10, 11$, and the probability distribution $\mathcal{P}$ is given by the table.

Definition 2. A *test* for a signed probability space $S = (\Omega, \mathcal{P})$ is given by (and, mathematically speaking, can be identified with) a partition of Ω into parts of nonnegative probability.

An *execution* of a test picks out one of its parts. The example above explicitly exhibits three tests.

> **Q:** Normally, in traditional probability theory, an execution of a probability trial picks out an outcome ω with probability $\mathcal{P}(\omega)$. Why don't you do that in general? Require that only outcomes of nonnegative probability are picked out.

> **A:** The proposed test is impossible in the case of nontraditional probability distribution. Notice that, in the example, the probability of the event $e^+ = \{\omega \in \Omega : \mathcal{P}(\omega) \geq 0\}$ is more than 1. What would it mean to pick an outcome from e^+ according to a distribution with total probability > 1? Our definition of test intends to reflect measurements in quantum mechanics.

3 Observation frames

Our goal in this section is to formalize the notions of an observable event and a coobservable set of events. Intuitively, a set E of events is coobservable if there is a test τ_E that allows us to observe, for all $e \in E$, whether e occurred or not. Further, e is observable if the singleton set $\{e\}$ is coobservable.

Definition 3. An *observation frame* is a pair (Ω, CO), such that Ω is a nonempty set, CO is a collection of subsets of Ω, and the following axioms hold:

CO1. If $X \subseteq Y \in \mathrm{CO}$ then $X \in \mathrm{CO}$.

CO2. If $e_1, e_2 \in X \in \mathrm{CO}$ then $X \cup \{\overline{e_1}\} \in \mathrm{CO}$ and $X \cup \{e_1 \cup e_2\} \in \mathrm{CO}$.   ◄

Terminology. Event sets in CO are *coobservable*. An event e is *observable* if the set $\{e\}$ is coobservable. For brevity, maximal coobservable sets, maximal in the inclusion order, will be called *ensembles*.   ◄

Notation. The set $\bigcup \mathrm{CO}$ of the observable events will be denoted Ob.

It follows from the definition that every ensemble is a Boolean algebra of subsets of Ω.

`Q:` How do you justify CO2.

`A:` Let $e_1, e_2 \in X \in \mathrm{CO}$. Since X is coobservable, there exists a test τ_X that allows us to observe, for all $e \in X$, whether e occurred or not. Therefore τ_X also allows us to observe whether $\overline{e_1}$ occurred or not: it occurred if and only if e_1 didn't occur. And τ_X allows us to observe whether $e_1 \cup e_2$ occurred or not: it occurred if and only if e_1 occurred or e_2 occurred.

Proviso 1 (Finiteness)**.** By default, observation frames are finite, i.e., their sample spaces are finite.

Remark. The definition of observation frames should be more general by excluding Ω and dealing only with coobservation. But, at this initial point of our investigation, we are willing to sacrifice the generality. ◄

Since every coobservable set is a subset of an ensemble and every subset of an ensemble is coobservable, the whole collection CO of coobservable event sets is determined by the ensembles.

Further, due to the finiteness proviso, the Boolean algebra of any ensemble E is atomic. The atoms partition the sample space; let us call that partition Π_E. The partition Π_E uniquely determines the ensemble E. Thus the collection CO can be given by the table of ensemble-induced partitions Π_E.

Example (The observation frame of Piponi's scenario)**.** Piponi's scenario gives rise to the following observation frame. The sample space Ω consists of the four binary strings $00, 01, 10, 11$, and there are three ensembles giving rise to the following partitions:

$$\{\{00, 01\}, \{10, 11\}\}$$
$$\{\{00, 10\}, \{01, 11\}\}$$
$$\{\{00, 11\}, \{01, 10\}\}$$

Furthermore, there is the least common refinement Π of all the ensemble-induced partitions; it has the smallest number of parts. Notice that, for any part $e \in \Pi$, different outcomes in e cannot be distinguished. For all practical purposes, members of Π can be viewed as singletons.

Proviso 2 (Fat outcomes)**.** By default, each part of the common refinement contains a single outcome.

4 Observation spaces

Definition 4. An *observation space* is a triple $(\Omega, \mathrm{CO}, \mathcal{P})$, such that

- (Ω, CO) is an observation frame,

- $\mathcal{P}$ is a real-valued function on $\mathrm{Ob} = \bigcup \mathrm{CO}$ satisfying the following versions of the probability laws PL1–PL3.

  - $\mathcal{P}(\Omega) = 1$.
  - If $(e_1, e_2) \in \mathrm{CO}$ and $e_1 \cap e_2 = \emptyset$ then $\mathcal{P}(e_1 + e_2) = \mathcal{P}(e_1) + \mathcal{P}(e_2)$.
  - $\mathcal{P}(e) \geq 0$ for all $e \in \mathrm{Ob}$. ◁

Since each observable event belongs to some ensemble, it suffices to define $\mathcal{P}$ on every ensemble (and ensure that every event gets the same probability in every ensemble that contains it). Since any ensemble E is a Boolean algebra of sets, the probability distribution on E is determined by the probabilities assigned to the parts of the ensemble partition Π_E.

Problem 1 (Extension)**.** Given an observation space $(\Omega, \mathrm{CO}, \mathcal{P})$, do the following.

1. Decide whether there is a traditional probability distribution that extends $\mathcal{P}$ from observable events to all events.

2. If such a traditional extension exists then find one.

3. Otherwise decide whether there is a signed probability distribution that extends $\mathcal{P}$ from observable events to all events.

4. If such a signed extension exists then find one.

The Fat-outcomes proviso of the previous section makes the Extension Problem trivial in the case of a single ensemble: the given $\mathcal{P}$ is already defined on all events. If there are exactly two ensembles A and B, imposing partitions $\Pi_A = \{A_1, \ldots, A_m\}$ and $\Pi_B = \{B_1, \ldots, B_n\}$, and if every intersection $A_i \cap B_j \neq \emptyset$, then there is a traditional solution for the Extension Problem: set $\mathcal{P}(A_i \cap B_j) = \mathcal{P}(A_i) \cdot \mathcal{P}(B_j)$.

Example (The observation space of Piponi's scenario)**.** The observation frame of the scenario is described in the previous Example. It has exactly three ensembles. It remains only to specify the probability distribution $\mathcal{P}$. It is given by the following table.

$$\mathcal{P}\{00, 01\} = 0 \quad \mathcal{P}\{10, 11\} = 1$$
$$\mathcal{P}\{00, 10\} = 0 \quad \mathcal{P}\{01, 11\} = 1$$
$$\mathcal{P}\{00, 11\} = 0 \quad \mathcal{P}\{01, 10\} = 1$$

The following simple theorem will turn out to be useful.

Theorem 5 (Symmetry). *Let G be a group of automorphisms of an observation space $O = (\Omega, \mathrm{CO}, \mathcal{P})$. If Q is a possibly-signed extension of $\mathcal{P}$ to all events, then so is the average*

$$R(e) = \frac{1}{|G|} \sum_{g \in G} Q(ge).$$

`Proof.` If event e is observable then $Q(ge) = Q(e) = \mathcal{P}(e)$ for any $g \in G$; hence $R(e) = \mathcal{P}(e)$. It remains to check that R is a possibly-signed probability distribution. Since Ω is observable, $R(\Omega) = \frac{1}{|G|} \sum_{g \in G} \mathcal{P}(\Omega) = \mathcal{P}(\Omega) = 1$. Further, if events e_1, e_2 are disjoint and coobservable then

$$R(e_1 \cup e_2) = \frac{1}{|G|} \sum_{g \in G} Q(g(e_1 \cup e_2)) = \frac{1}{|G|} \sum_{g \in G} [Q(ge_1) + Q(ge_2)] = R(e_1) + R(e_2). \qquad \text{QED}$$

5   Bell's theorem and negative probabilities

Quantum theory is contextual in the sense that the value of an observable O, measured as a part of one context, may differ from the value of O measured as a part of another context. Attempts to avoid contextuality may lead to negative probabilities. This will be illustrated in the present section. We start with two Gedankenexperiments exhibiting the contextuality of quantum mechanics.

Prepare the state $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, known as the spin singlet state, of a pair of spin 1/2 particles, e.g., electrons. Here $|0\rangle$ represents spin up in the z direction, and $|1\rangle$ represents spin down. Choose an arbitrary direction **a** and measure spin in direction **a** on both particles, getting $+\frac{1}{2}$ if the spin is up or $-\frac{1}{2}$ if the spin is down. According to quantum mechanics, the results of the two measurements are opposite to one another: one measurement yields $+\frac{1}{2}$ and the other $-\frac{1}{2}$ [14, Box 2.7]. This is true even if each measurement is performed outside of the lightcone of the other and thus cannot possibly affect the other measurement.

Alternatively, one can work with photons, which are spin 1 particles. In this connection, see Figure 1 which, together with the caption, is borrowed from [2].
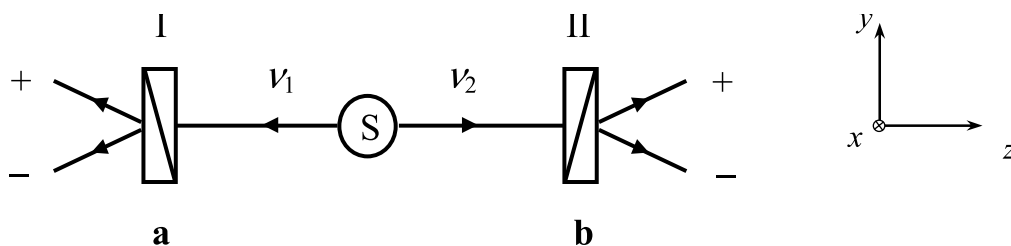
Figure 1. Einstein-Podolsky-Rosen-Bohm Gedankenexperiment with photons. The two photons ν_1 and ν_2 , emitted in the state $\left|\Psi(1,2)\right\rangle$ of Equation (1), are analyzed by linear polarizers in orientations **a** *and* **b**. *One can measure the probabilities of single or joint detections in the output channels of the polarizers.*

The two photons are moving along the z axis. Formula $|\Psi(1,2)\rangle$ in the figure is $\frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$ where $|0\rangle$ and $|1\rangle$ are unit vectors in the Hilbert space for the quantum system of one photon. $|0\rangle$ and $|1\rangle$ correspond to polarization in the directions of the x and y axes respectively. If the orientations $\mathbf{a}, \mathbf{b}$ of the analyzers in the figure coincide then the two measurement outcomes are guaranteed to coincide: both are $+1$ or both are -1.

Einstein, Podolsky and Rosen (EPR) saw contextuality as an indication that quantum mechanics is incomplete, i.e., that some hidden variables are missing [9]. Bell famously proved that local hidden-variable theories contradict quantum-mechanical statistics [3]. Subsequent experiments supported the latter.

In general, the orientations $\mathbf{a}$, $\mathbf{b}$ of the two analyzers on Figure 1 may be different. If θ is the angle $\angle(\mathbf{a}, \mathbf{b})$ between vectors $\mathbf{a}$ and $\mathbf{b}$ then the two outcomes $(+1, +1)$ and $(-1, -1)$ in which the two measurements give us the same result, have probability $\frac{1}{2}\cos^2\theta$ each, and the other two outcomes $(+1, -1)$ and $(-1, +1)$ have probability $\frac{1}{2}\sin^2\theta$ each [15, §6-2]. In particular, if $\mathbf{a} = \mathbf{b}$ then the probability of getting the same result is 1.

To illustrate how Bell's Theorem leads to negative probabilities, David Schneider played with three orientations $\vec{A}, \vec{B}, \vec{C}$ in his blog post [18]. The angle $\angle(\vec{A}, \vec{C}) = 3\pi/8$, and $\vec{B}$ is in between so that the angle $\angle(\vec{A}, \vec{B}) = \pi/4$ and $\angle(\vec{B}, \vec{C}) = \pi/8$. (We swapped Schneider's $\vec{B}$ and $\vec{C}$ so that our $\vec{B}$ is between $\vec{A}$ and $\vec{C}$.) He arrived at negative probabilities, implicitly assuming that (in terms of §4) there is some solution of the appropriate Extension Problem. In the rest of this section, we explain Schneider's derivation and then address the implicit assumption.

Imagine that we work with a noncontextual hidden-variable theory where the measurements are determined locally, "so the first thing we need to do is momentarily forget all our knowledge of quantum mechanics," [14, page 148]. But we do use the data obtained in three optical experiments of the kind depicted in Figure 1.

1. Experiment AB involves orientations $\vec{A}$ and $\vec{B}$,

2. Experiment BC involves orientations $\vec{B}$ and $\vec{C}$,

3. Experiment AC involves orientations $\vec{A}$ and $\vec{C}$.

This gives rise to the following observation space $T = (\Omega, \mathrm{CO}, \mathcal{P})$.

Sample space. The sample space Ω consists of eight sample points

$1, a.k.a. \quad + ++, \quad 2, a.k.a. \quad + +-, \quad 3, a.k.a. \quad + -+, \quad 4, \quad a.k.a. + --,$
$5, a.k.a. \quad - ++, \quad 6, a.k.a. \quad - +-, \quad 7, a.k.a. \quad - -+, \quad 8, \quad a.k.a. - - - .$

In the three-letter words abc in alphabet $\{+, -\}$, the first letter a is the result of measuring the spin in the $\vec{A}$ direction. Similarly for the second letter b and the third letter c using the $\vec{B}$ and $\vec{C}$ directions.

Coobservation. The first two letters a, b of outcomes abc give rise to an equivalence relation $abc \equiv a'b'c' \iff (a = a' \wedge b = b')$ whose four equivalence classes $\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8\}$ form a partition Π_{12} of Ω. Thanks to experiment AB, the four parts of Π_{12} are coobservable. Define partitions Π_{23} and Π_{31} similarly. The four parts of Π_{23} are coobservable thanks to the experiment BC, and the four parts of Π_{13} are coobservable thanks to the experiment AC.

The four parts of any partition Π_{ij} generate a Boolean algebra E_{ij} of subsets of Ω. Define

$$\mathrm{CO} = \{E : E \subseteq E_{ij} \text{ for some } i, j\},$$

so that each E_{ij} is an ensemble.

Probability distribution. Since the experimental results support quantum mechanics, the experiments AB, BC and AC produce results approximating the following three tables.

$$\mathcal{P}(+ + \pm) = \mathcal{P}\{1, 2\} = \frac{1}{2}\cos^2(\pi/4) = 1/4,$$

$$\mathcal{P}(- - \pm) = \mathcal{P}\{7, 8\} = \frac{1}{2}\cos^2(\pi/4) = 1/4, \qquad \text{(AB)}$$

$$\mathcal{P}(+ - \pm) = \mathcal{P}\{3, 4\} = \frac{1}{2}\sin^2(\pi/4) = 1/4,$$

$$\mathcal{P}(- + \pm) = \mathcal{P}\{5, 6\} = \frac{1}{2}\sin^2(\pi/4) = 1/4.$$

$$\mathcal{P}(\pm + +) = \mathcal{P}\{1, 5\} = \frac{1}{2}\cos^2(\frac{\pi}{8}) = \frac{1}{8}(2 + \sqrt{2}),$$

$$\mathcal{P}(\pm - -) = \mathcal{P}\{4, 8\} = \frac{1}{2}\cos^2(\frac{\pi}{8}) = \frac{1}{8}(2 + \sqrt{2}), \qquad \text{(BC)}$$

$$\mathcal{P}(\pm + -) = \mathcal{P}\{2, 6\} = \frac{1}{2}\sin^2(\frac{\pi}{8}) = \frac{1}{8}(2 - \sqrt{2}),$$

$$\mathcal{P}(\pm - +) = \mathcal{P}\{3, 7\} = \frac{1}{2}\sin^2(\frac{\pi}{8}) = \frac{1}{8}(2 - \sqrt{2}).$$

$$\mathcal{P}(+ \pm +) = \mathcal{P}\{1, 3\} = \frac{1}{2}\cos^2(3\pi/8) = \frac{1}{8}(2 - \sqrt{2}),$$

$$\mathcal{P}(- \pm -) = \mathcal{P}\{6, 8\} = \frac{1}{2}\cos^2(3\pi/8) = \frac{1}{8}(2 - \sqrt{2}), \qquad \text{(AC)}$$

$$\mathcal{P}(+ \pm -) = \mathcal{P}\{2, 4\} = \frac{1}{2}\sin^2(3\pi/8) = \frac{1}{8}(2 + \sqrt{2})$$

$$\mathcal{P}(- \pm +) = \mathcal{P}\{5, 7\} = \frac{1}{2}\sin^2(3\pi/8) = \frac{1}{8}(2 + \sqrt{2}).$$

Let $[a = b]$ be the event that the measurements for orientations $\vec{A}, \vec{B}$ coincide, and $[a \neq b]$ be the complementary event, that these measurements are distinct. Define events $[b = c], [b \neq c], [a = c]$ and $[a \neq c]$ similarly. We have

$$[a = b] = (+ + \pm) \cup (- - \pm) = \{1, 2, 7, 8\}, \quad \mathcal{P}[a = b] = 1/2,$$

$$[a \neq b] = (+ - \pm) \cup (- + \pm) = \{3, 4, 5, 6\}, \quad \mathcal{P}[a \neq b] = 1/2,$$

$$[b = c] = (\pm + +) \cup (\pm - -) = \{1, 4, 5, 8\}, \quad \mathcal{P}[b = c] = \cos^2(\pi/8) = \frac{1}{4}(2 + \sqrt{2}),$$

$$[a = c] = (+ \pm +) \cup (- \pm -) = \{1, 3, 6, 8\}, \quad \mathcal{P}[a = c] = \sin^2(\pi/8) = \frac{1}{4}(2 - \sqrt{2}).$$

Suppose that a possibly-signed probability distribution Q extends $\mathcal{P}$ to all events. Let U be the nonobservable event $\{3, 6\}$. The following computation shows that Q cannot be traditional.

$$2Q(U) = [Q(U) + Q([a = c] - U)] + [Q(U) + Q([a \neq b] - U)]$$
$$- [Q([a = c] - U) + Q([a \neq b] - U)]$$

$$= \mathcal{P}[a = c] + \mathcal{P}[a \neq b] - \mathcal{P}[b = c] = \frac{1}{4}(2 - \sqrt{2}) + \frac{1}{2} - \frac{1}{4}(2 + \sqrt{2})$$

$$= \frac{1}{2}(1 - \sqrt{2})$$

$$Q(U) = \frac{1}{4}(1 - \sqrt{2}) < 0.$$

Now let's address the question whether there is any solution of the Extension Problem in our case.

Consider the transformation g of Ω that, for any outcome abc, replaces every letter by its opposite. For example, $g(+ - +) = - + -$. It is easy to see that g is an automorphism of the observation space T. By Theorem 5, the average $R(e) = \frac{1}{2}(Q(ge) + Q(e))$ is a signed probability distribution that extends $\mathcal{P}$ to all events.

Since $Q\{3, 6\} = \frac{1}{4}(1 - \sqrt{2})$, we have:

$$R(3) = R(6) = \frac{1}{2}(Q(3) + Q(6)) = \frac{1}{2}Q\{3, 6\} = \frac{1}{8}(1 - \sqrt{2})$$

$$R(1) = R(8) = \frac{1}{8}(2 - \sqrt{2}) - \frac{1}{8}(1 - \sqrt{2}) = \frac{1}{8} \qquad \text{by (AC)}$$

$$R(2) = R(7) = \frac{1}{4} - \frac{1}{8} = \frac{1}{8} \qquad \text{by (AB)}$$

$$R(4) = R(5) = \frac{1}{4} - \frac{1}{8}(1 - \sqrt{2}) = \frac{1}{8}(1 + \sqrt{2}) \qquad \text{by (AB)}$$

To prove that R is consistent with $\mathcal{P}$, it suffices to check that these probabilities satisfy the constraints (AB), (BC) and (AC) where $\mathcal{P}\{k, l\}$ is replaced with $R(k) + R(l)$.

The (AB) constraints and the first two of the (AC) constraints are satisfied in a trivial way (because of the way they have been used to compute the outcome probabilities). $R(2) + R(4) = R(7) + R(5) = \frac{1}{8} + \frac{1}{8}(1 + \sqrt{2}) = \frac{1}{8}(2 + \sqrt{2}) = \mathcal{P}\{2, 4\} = \mathcal{P}\{5, 7\}$, and so the remaining two (AC) constraints are satisfied. We check the (BC) constraints.

$$R(1) + R(5) = R(8) + R(4) = \frac{1}{8} + \frac{1}{8}(1 + \sqrt{2}) = \frac{1}{8}(2 + \sqrt{2}) = \mathcal{P}\{1, 5\} = \mathcal{P}\{4, 8\}$$

$$R(2) + R(6) = R(3) + R(7) = \frac{1}{8} + \frac{1}{8}(1 - \sqrt{2}) = \frac{1}{8}(2 - \sqrt{2})$$

6 Hardy's Gedankenexperiment: Contextuality without negativity

The previous section may give one the idea that contextuality always leads to negative probabilities. In this section, building on Lucien Hardy's article [11] and also influenced by David Mermin's article [12], we illustrate that this is not so.

Experiment. Two one-qubit particles emerge from a common source heading for two far apart detectors. Aside from the passage of the particles from the source to the detectors, there are no connections between the source and either detector

or between the two detectors. The following picture is borrowed from [12] (and slightly modified).
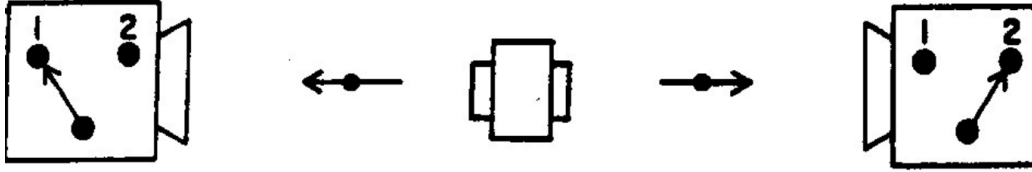
Each of the detectors is randomly set, ahead of time, to one of two modes, indicated by "1" and "2" in the picture. Four cases arise, two possible settings on each of the two detectors.

When a particle arrives at a detector, that detector performs a measurement and exhibits the result. In mode 1, observable Z is measured. Its value is $+1$ in state $|0\rangle$ and -1 in state $|1\rangle$. In mode 2, observable X is measured. Its value is $+1$ in state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and -1 in state $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. Initially, the two particles are in state

$$|\psi\rangle = \frac{1}{\sqrt{3}}\big(|01\rangle + |10\rangle - |00\rangle\big). \tag{1}$$

That completes the description of the experiment. ◄

In the rest of the section, we analyse the experiment. The initial state is given to us in the basis $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$. It will be convenient to express it in three additional bases. In the basis $|0+\rangle$, $|0-\rangle$, $|1+\rangle$, $|1-\rangle$, we have

$$\sqrt{3}|\psi\rangle = -|0\rangle(|0\rangle - |1\rangle) + |1\rangle(|+\rangle + |-\rangle) = -\sqrt{2}|0-\rangle + \frac{1}{\sqrt{2}}|1+\rangle + \frac{1}{\sqrt{2}}|1-\rangle. \tag{2}$$

In the basis $|+0\rangle$, $|+1\rangle$, $|-0\rangle$, $|-1\rangle$, we have

$$\sqrt{3}|\psi\rangle = |01\rangle - (|0\rangle - |1\rangle)|0\rangle = \frac{(|+\rangle + |-\rangle)|1\rangle}{\sqrt{2}} - \sqrt{2}|-0\rangle = \frac{1}{\sqrt{2}}|+1\rangle + \frac{1}{\sqrt{2}}|-1\rangle - \sqrt{2}|-0\rangle. \tag{3}$$

In the basis $|++\rangle$, $|+-\rangle$, $|-+\rangle$, $|--\rangle$, we have

$$\sqrt{3}|\psi\rangle = \frac{1}{2}|++\rangle - \frac{1}{2}|+-\rangle - \frac{1}{2}|-+\rangle - \frac{3}{2}|--\rangle. \tag{4}$$

`Contextuality.` Again, suppose that we work with a noncontextual hidden-variable theory where the measurements are determined locally. Let $f_l(Z)$ and $f_l(X)$ be the sets of values that may occur as the result of the Z measurement and X measurement respectively on the left. Define $f_r(Z)$ and $f_r(X)$ similarly. By the noncontextuality assumption, these sets depend only on what happens on their side of the common source. But this leads to a contradiction.

Indeed, consider the case ZZ, where both modes are 1 and thus observable Z is measured on the left and the right. By (1), the conditional probability $P[(+1, +1) \mid ZZ]$ that we have $+1$ on the left and the right is $1/3$. Hence $+1 \in f_l(Z) \cap f_r(Z)$. In particular, $+1 \in f_l(Z)$.

By (4), in the case XX, where both modes are 2 and thus observable X is measured on the left and the right, the conditional probability $\mathcal{P}[(+1, +1)|XX]$ that $+1$ is produced on the left and the right is positive (namely $1/12$) and therefore $+1 \in f_r(X)$.

Now let's consider the case ZX, where the left mode is 1 and so Z is measured on the left and where right mode is 2 and so X is measured on the right. By noncontextuality, it must be possible to have $+1$ on the left and on the right in the same trial. But this does not happen. For, by (2), the conditional probabilities in the ZX case are as follows.

$$\mathcal{P}[(+1, -1) \mid ZX] = 2/3, \quad \mathcal{P}[(-1, +1) \mid ZX] = 1/6, \quad \mathcal{P}[(-1, -1) \mid ZX] = 1/6,$$

so that $\mathcal{P}[(+1, +1) \mid ZX] = 0$ which gives the desired contradiction.

Observation space. The sample space Ω consists of 16 outcomes for the combination of the random settings of modes and the observations of the results: $(ZZ, \pm1, \pm1)$, $(ZX, \pm1, \pm1)$, $(XZ, \pm1, \pm1)$ and $(XX, \pm1, \pm1)$.

In the sense of the observation space, all 16 outcomes are observable. For example, consider the case ZZ where both modes are 1. By (1), the conditional probabilities $\mathcal{P}[(+1, +1)|ZZ], \mathcal{P}[(+1, -1)|ZZ], \mathcal{P}[(-1, +1)|ZZ]$ are $1/3$. Accordingly

$$\mathcal{P}[\{(+1, +1), (+1, -1), (-1, +)\} \mid ZZ] = 1$$

and so $\mathcal{P}[(-1, -1) \mid ZZ] = 0$. Thus, in the case ZZ, it is impossible to have -1 on the left and on the right in the same trial. But, like event $\emptyset$, the event $\{(ZZ, -1, -1)\}$ is observable in the sense of §3.

Thus $\mathcal{P}$ is defined on all events. The extension problem is trivial in Hardy's case, and negative probabilities do not arise.

Remark. The presence of contextuality and the absence of negativity seem to contradict Robert Spekkens's claim that negativity and contextuality are equivalent forms of nonclassicality [19]. Earlier, in a lengthy footnote in Section 6 of [6], we showed that the equivalence claim is unsubstantiated.

7 Limitation

The notion of observation spaces was motivated by quantum mechanics with its observables, i.e., Hermitian operators, which may or may not be coobservable, i.e.,

commeasurable. Unfortunately, as this section shows, this notion is too simplistic to faithfuly model more complicated sets of quantum mechanical observables.

Let $\mathcal{S}$ be a set of Hermitian operators on a finite-dimensional Hilbert space $\mathcal{H}$, and let (Ω, CO) be an observation frame. By a *model* of $\mathcal{S}$ in (Ω, CO) we mean a partial function μ from closed subspaces of $\mathcal{H}$ to subsets of Ω such that:

M1 For each operator $A \in \mathcal{S}$ and each sum E of eigenspaces of A, $\mu(E)$ is defined and is an observable event in (Ω, CO).

M2 For each operator $A \in \mathcal{S}$, the collection $\mathcal{E}_A$ of all sums of eigenspaces of A has $\{\mu(E) : E \in \mathcal{E}_A\}$ coobservable in (Ω, CO).

M3 If E is a sum of eigenspaces of some $A \in \mathcal{S}$, then the complementary sum $E^\perp$ satisfies $\mu(E^\perp) = \Omega - \mu(E)$.

M4 If E_1 and E_2 are sums of eigenspaces of a single $A \in \mathcal{S}$, then $\mu(E_1 + E_2) = \mu(E_1) \cup \mu(E_2)$.

Q: Explain M1. Why are we talking about sums of eigenspaces?

A: Given a set V of eigenvalues of a Hermitian operator $A \in \mathcal{S}$, let E be the sum of the corresponding eigenspaces of A. As a closed subspace of the Hilbert space, E is an event. When we measure the Hermitian operator A, the result v is one of its eigenvalues. By observing whether v is in V we know whether E occurred. So E should be observable in (Ω, CO).

The same justification applies to M2.

Q: Actually M2 implies M1.

A: Almost. M1 contains the requirement that $\mu(E)$ is defined which is used implicitly in M2.

Theorem 6. *There is a finite set $\mathcal{S}$ of Hermitian operators on $\mathbb{C}^4$ that admits no model in any observation frame.*

Proof. Consider the 18-vector, 9-basis example of Cabello et al. [7] given in color on the Wikipedia page for "Kochen-Specker Theorem" [22]. For each of the 9 bases B there, invent an operator S_B (Hermitian, with 4 distinct eigenvalues) whose eigenspaces are exactly the 4 vectors in B (and their scalar multiples). Let $\mathcal{S}$ be the set of these 9 S_B's, and suppose, toward a contradiction, that μ were a model of $\mathcal{S}$ in (Ω, CO).

Let B be any one of the nine bases, and consider the sums E of eigenspaces of S_B, i.e., the "coordinate subspaces" of $\mathbb{C}^4$ with respect to the basis B. They

constitute a Boolean algebra BA(B) and, by the requirements for a model, their μ-images are defined and constitute a Boolean subalgebra of 2^{Ω}; furthermore, when restricted to BA(B), μ is a Boolean homomorphism. The atoms of BA(B) are the eigenspaces themselves. The μ-image of BA(B) is a finite Boolean algebra of events in (Ω, CO). Its atoms are the μ-images of the four eigenspaces. So each $\omega \in \Omega$ is in exactly one of those four μ-images.

Fix some point $\omega \in \Omega$. It selects, for each of the 9 bases B, exactly one of the 4 vectors in B. Furthermore, when the same vector occurs in two bases, then it is selected from one if and only if it is selected from the other. This is because for $\vec{v}$ to be selected means that (writing $\langle \vec{v} \rangle$ for the subspace generated by $\vec{v}$) $\omega \in \mu\langle \vec{v} \rangle$, and this doesn't depend on the rest of a basis.

But the point of the example of Cabello et al. is precisely that such a selection is impossible. Indeed, we would have 9 selections, one from each basis, but every selection would occur twice because, as indicated by the colors in [22], every vector occurs in exactly two bases. So 9 would have to be even. $\qquad\square$

The $\mathbb{C}^4$ in the theorem can be improved to $\mathbb{C}^3$ at the cost of using a more complicated example, as in the original Kochen-Specker proof.

Q: This is too bad that there is such a brutal limitation on observation spaces. And, since the theorem speaks about modeling in an observation frame, rather than observation space, negative probabilities are irrelevant.

A: It is not completely obvious to us at this point that negative probabilities are irrelevant. We can weaken conditions M3 and M4 by requiring that the equalities hold only up to an error of probability zero. The resulting weaker models would be defined not in observation frames but in observation spaces because the weaker notion involves probabilities.

If we stick to nonnegative probabilities, this doesn't buy us anything. But once negative probabilities enter the picture, errors of probability zero become more complicated. They may involve cancellation between outcomes of positive probability and outcomes of negative probability.

Q: I have another question. It seems that the proof of the limitative Theorem 6 crucially uses the existence of points in observation frames. Did you consider working more abstractly with just observable events and coobservable sets of events?

`A:` Yes, a little. We found that a "pointless" abstraction of observation frames seems to lead naturally to orthomodular lattices, extensively studied (though not by us) in quantum logic [4].

Future work

We introduced observation spaces and the extension problem. Observation spaces allow us to model some interesting quantum mechanical situations. And the extension problem amounts to asking when does one need negative probabilities and what can one accomplish with negative probabilities. This is a small step toward understanding what negative probabilites can be used for. Observation spaces do not allow us to model straightforwardly the more complicated situations involved in proofs of the Kochen-Specker theorem. We'd like to understand what the next step should be.

References

[1] Samson Abramsky and Adam Brandenburger, "An operational interpretation of negative probabilities and no-signaling models," arXiv:1401.2561v2, 2014, `https://arxiv.org/abs/1401.2561v2`.

[2] Alain Aspect, "Bell's theorem: The naive view of an experimentalist," in *Quantum [un]speakables*, Reinhold A. Bertlmann and Anton Zeilinger (eds.), Springer 2002, pages 119–153.

[3] John Bell, "On the Einstein-Podolsky-Rosen paradox," Physics 1 (1964), 195–200.

[4] Garett Birkhoff and John von Neumann, "The Logic of Quantum Mechanics," *Annals of Mathematics* 37 (1936) 823–843.

[5] Andreas Blass and Yuri Gurevich, "Negative probabilities," Bulletin of EATCS 115 2004.

[6] Andreas Blass and Yuri Gurevich, "Optimal no-go theorem on hidden-variable predictions of effect expectations," Physical Review A 97:3 2018.

[7] Adán Cabello, Jose M. Estebaranz, Guillermo Garcia Alcaine, "A proof with 18 vectors of the Bell-Kochen-Specker theorem," `https://arxiv.org/abs/quant-ph/9706009`

[8] Paul A.M. Dirac, "The physical interpretation of quantum mechanics," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 180:980 (1942) 1–40.

[9] Albert Einstein, Boris Podolsky and Nathan Rosen, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 47 (1935) 777–780.

[10] Richard P. Feynman, "Negative probabilities," in *Quantum Implications: Essays in Honor of David Bohm*, ed. F. D. Peat and B. Hiley, Routledge & Kegan Paul (1987) 235–248.

[11] Lucien Hardy, "Non-locality for two particles without inequalities for almost all entangled states," *Physical Review Letters* 71 (1993) 1665–1668.

[12] N. David Mermin, "Quantum mysteries refined," *American Journal of Physics* 62 (1994) 880–887.

[13] José E. Moyal, "Quantum mechanics as a statistical theory," Proc. Cambridge Phil. Soc. 45 (1949) 99–124.

[14] Michael A. Nielsen and Isaac L. Chuang, "Quantum computation and quantum information," 10th Anniversary edition, Cambridge University Press, 2010.

[15] Asher Peres, "Quantum theory: concepts and methods," Kluwer 2002.

[16] Dan Piponi, "Negative probabilities," A blog post, April 12, 2008, `http://blog.sigfpe.com/2008/04/negative-probabilities.html`,

[17] Itamar Pitowsky, "Quantum probability, quantum logic," Springer Lecture Notes in Physics 321, 1989.

[18] David R. Schneider, "Bell's Theorem and Negative Probabilities," blog post, 2005–2007, `http://www.drchinese.com/David/Bell_Theorem_Negative_Probabilities.htm`

[19] Robert W. Spekkens, "Negativity and contextuality are equivalent notions of non-classicality," *Physical Review Letters* 101(2) (2008) 020401.

[20] Eugene P. Wigner, "On the quantum correction for thermodynamic equilibrium," Physical Review 40 (1932) 749–759.

[21] Wikipedia contributors, "Electric Impedance," Wikipedia, 7 April 2018, `https://en.wikipedia.org/w/index.php?title=Electrical_impedance&oldid=836085652`.

[22] Wikipedia contributors, "Kochen-Specker theorem," Wikipedia, 4 April 2018, `https://en.wikipedia.org/w/index.php?title=Kochen-Specker_theorem&oldid=828375518`.

[23] Cosmas K. Zachos, David B. Fairlie, and Thomas L. Curtright (eds.), *Quantum Mechanics in Phase Space: An Overview with Selected Papers,* World Scientific Series in 20th Century Physics, vol. 34 (2005).