Quantum Probability: An Introduction*

Guido Bacciagaluppi†

14 February 2014

The topic of probabilty in quantum mechanics is rather vast, and in this article, we shall choose to discuss it from the perspective of whether and in what sense quantum mechanics requires a generalisation of the usual (Kolmogorovian) concept of probability. We shall focus on the case of finite-dimensional quantum mechanics (which is analogous to that of discrete probability spaces), partly for simplicity and partly for ease of generalisation. While we shall largely focus on formal aspects of quantum probability (in particular the non-existence of joint distributions for incompatible observables), our discussion will relate also to notorious issues in the interpretation of quantum mechanics. Indeed, whether quantum probability can or cannot be ultimately reduced to classical probability connects rather nicely to the question of 'hidden variables' in quantum mechanics.

1 Quantum mechanics (once over gently)

If a spinning charged object flies through an appropriately inhomogeneous magnetic field, then according to the laws of classical physics it will experience a deflection, in the direction of the gradient of the field, proportional

*An abridged version of this essay will appear in A. Hájek and C. Hitchcock (eds.), *The Oxford Handbook of Probability and Philosophy*, Oxford University Press. The present version includes in particular many more footnotes and references and the Appendix with the proofs of the Lemma and Proposition of Section 5.

†Department of Philosophy, University of Aberdeen; Institut d'Histoire et de Philosophie des Sciences et des Techniques (CNRS, Paris 1, ENS); and Collegio Dottorale in Filosofia, Epistemologia e Storia della Cultura, Università di Cagliari. Address for correspondence: Department of Philosophy, University of Aberdeen, Old Brewery, High Street, Aberdeen AB24 3UB, Scotland, U.K. (e-mail: g.bacciagaluppi@abdn.ac.uk).

to its angular momentum in the same direction (i.e. proportional to how fast it is spinning along that axis). In quantum mechanics one observes a similar effect, except that the object is deflected only by discrete amounts, as if its classical angular momentum could take only certain values (varying by units of Planck's constant $\hbar$). Some particles even seem to possess an intrinsic such 'spin' (i.e. not derived from any rotational motion), e.g. so-called spin-1/2 systems, such as electrons, which get deflected by amounts corresponding to spin values $\pm\frac{\hbar}{2}$. Experiments of this kind are known as Stern–Gerlach experiments, and they can be used to illustrate some of the most important features of quantum mechanics.

Imagine a beam of electrons, say, moving along the y-axis and encountering a region with a magnetic field inhomogeneous along the x-axis. The beam will split in two, as can be ascertained by placing a screen on the other side of the experiment, and observing that particle detections are localised around two distinct spots on the screen (needless to say, real experiments are a little messier than this).[1]

The first thing to point out is that if we send identically prepared electrons one by one through such an apparatus, each of them will trigger only one detection, *either* in the upper half *or* the lower half of the screen, with probabilities depending on the initial preparation of the incoming electrons.

The second thing to point out is that the same is true whatever the direction in which the inhomogeneous magnetic field is laid, whether along the x-axis, the z-axis, or any other direction: the beam of electrons will always be split in two components, corresponding to a spin value $\pm\frac{\hbar}{2}$ along the direction of inhomogeneity of the field. If the incoming beam happens to be prepared by selecting one of the deflected beams in a previous Stern–Gerlach experiment (say, a beam of 'spin-x up' electrons), then the probabilities for detection in a further Stern–Gerlach experiment in a direction x' depend only on the angle ϑ between x and x' (and are given by $\cos^2(\vartheta/2)$ and $\sin^2(\vartheta/2)$). So, for example, the probability of measuring spin-z 'up' or 'down' in a beam of spin-x 'up' electrons is 1/2.

For each given preparation procedure (each prepared 'state') we thus have well-defined probability measures over the outcome spaces of various experiments. It is not obvious, however, what in general (if any) should

[1] Note also that reversing either the sign of the gradient or the polarity of the field will produce a deflection in the opposite direction (in both the classical and quantum case).

be the joint distribution for the outcomes of *different* experiments, because performing one kind of experiment (say, measuring spin-z) disturbs the probabilities relating to other subsequent experiments (say, spin-x). Indeed, if we imagine performing a spin-z followed by a spin-x measurement on electrons originally prepared in a spin-x up state, we shall get a 50–50 distribution for the results of the last measurement (whether we previosuly got spin-z up or down), although the original beam was 100% spin-x up. At least in this sense, different measurements in general are incompatible.

What is truly remarkable, however, and makes a straightforward hypothesis of disturbance untenable, is that such a putative disturbance can be *undone* if the spin-z up and spin-z down beam are brought together again before the spin-x measurement. In this case one obtains again spin-x up with probability 1, and this *even* if the whole experiment is performed on one electron at a time. Thus this case cannot be explained by interaction between different electrons when the two beams are brought together again. It reminds one rather of typically wave-like phenomena: the up components of the two beams appear to have interfered constructively, and the down components of the two beams appear to have interfered destructively, although it is each individual electron that displays this wave-like interference behaviour.

There is thus a genuine puzzle (one of many!) about whether and how the probability measures defined over the outcome spaces of the different experiments can be combined. This will be one of the main questions discussed in this article.

Let us very briefly review some standard bits of formalism (mainly to ease the transition into the more abstract setting of Section 3).[2] The natural mathematical framework for describing interference phenomena is a vector space, where any two elements, call them 'states', $|\psi\rangle$ and $|\varphi\rangle$ can be linearly superposed,

$$\alpha|\psi\rangle + \beta|\varphi\rangle \ . \tag{1}$$

The spin degree of freedom of an electron is described by a two-dimensional

[2] For more comprehensive but still relatively gentle introductions to quantum mechanics (and its philosophy), see e. g. Albert (1992), Ghirardi (1997), Wallace (2008), Bacciagaluppi (2013), and the relevant articles in the *Stanford Encyclopedia of Philosophy* (`http://plato.stanford.edu/`). For treatments emphasising the modern notion of effect-valued observables, used below in Section 3, see Busch, Lahti and Mittelstaedt (1991) and Busch, Grabowski and Lahti (1995).

complex vector space (with the usual scalar product, which we denote $\langle\varphi|\psi\rangle$). Vectors are usually normalised to unit length, since any two vectors $|\psi\rangle$ and $\alpha|\psi\rangle$ that are multiples of each other are considered physically equivalent. Each pair of up and down states is taken to correspond to an orthonormal basis, e.g. the spin-x and spin-z states, related by

$$|+_z\rangle = \frac{1}{\sqrt{2}}(|+_x\rangle + |-_x\rangle) \ , \qquad |-_z\rangle = \frac{1}{\sqrt{2}}(|+_x\rangle - |-_x\rangle). \tag{2}$$

Thus, while each spin-z state can be split into both up and down components in the spin-x basis, the down components, say, can cancel out again if the two spin-z states are appropriately combined:

$$\frac{1}{\sqrt{2}}(|+_z\rangle + |-_z\rangle) = \frac{1}{\sqrt{2}}\Big(\frac{1}{\sqrt{2}}(|+_x\rangle + |-_x\rangle) + \frac{1}{\sqrt{2}}(|+_x\rangle - |-_x\rangle)\Big) = |+_x\rangle \ . \tag{3}$$

Temporal evolution is given by the action of an appropriate group of linear operators (i. e. of linear mappings, which map superpositions into the corresponding superposition) on the states:

$$|\psi\rangle \mapsto U_{t_2 t_1}|\psi\rangle \ . \tag{4}$$

These operators are in fact unitary, i.e. preserve the length of vectors (and more generally scalar products between them). In a Stern–Gerlach measurement, the relevant unitary evolution is generated by an operator containing a term proportional to a 'spin operator', e.g. the z-spin operator S_z, written

$$S_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{5}$$

in the spin-z basis, which thus simply multiplies by a scalar the spin-z vectors:

$$S_z|\pm_z\rangle = \pm\frac{\hbar}{2}|\pm_z\rangle \tag{6}$$

(the spin-z states are *eigenstates* of the operator with corresponding *eigenvalues* $\pm\frac{\hbar}{2}$). During the measurement, this term couples the spin-z eigenstates (which it leaves invariant) to the spatial degrees of freedom of the electron, thus deflecting the motion of the electron:

$$|\pm_z\rangle|\psi\rangle \mapsto U|\pm_z\rangle|\psi\rangle = |\pm_z\rangle|\psi_\pm\rangle \tag{7}$$

(where we assume that $|\psi_\pm\rangle$ are states of the spatial degrees of freedom of the electron in which the electron is localised in two non-overlapping regions).

Measurable quantities in quantum mechanics (usually called 'observables') are traditionally associated with such operators, the corresponding eigenvalues being the values the observable can take.[3] Two observables thus understood will be compatible if they have all eigenvectors in common, or equivalently if the associated operators commute, i.e. $AB|\psi\rangle = BA|\psi\rangle$ for all states $|\psi\rangle$. Incompatibility of quantum mechanical observables is intuitively related to the idea that measurements of non-commuting observables generally require *mutually exclusive* experimental arrangements (implemented through appropriate unitary operators).

Note that if the initial state of the electron is a superposition of spin-z states, e.g. the spin-x up state (3), the linearity of the evolution will preserve the superposition:

$$\frac{1}{\sqrt{2}}(|+_z\rangle + |-_z\rangle)|\psi\rangle \mapsto \frac{1}{\sqrt{2}}(|+_z\rangle|\psi_+\rangle + |-_z\rangle|\psi_-\rangle) \ . \tag{8}$$

The state (8) has no longer product from, unlike the states (7). Indeed, in quantum mechanics the composition of degrees of freedom (or of different systems) proceeds by taking the *tensor product* of the vector spaces describing the different degrees of freedom (or systems), which is the space of all linear superpositions of product states. Non-product states are called *entangled*, and are the source of some of the most peculiar features of quantum mechanics. We can ignore them further, however, until we make contact with the discussion of the Bell inequalities in Section 5.

The last thing we need to recall from elementary treatments of quantum mechanics is what happens upon measurement. Namely, when a certain observable is measured, say S_z, the state of the system undergoes a stochastic transformation (a 'collapse', or 'reduction', or 'projection'), given mathematically by the projection onto one of the eigenstates of the measured observable, i.e. by the application of the projection operator $P^z_+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ (or $P^z_- \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$) — the operator with eigenvectors $|\pm_z\rangle$ and corresponding

[3]In order for such an operator to generate a unitary group, it has to be self-adjoint, which in finite dimensions simply means that the corresponding matrix is conjugate symmetric, e.g. $\begin{pmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$. This implies further that all its eigenvalues are real and that the vector space has an orthonormal basis composed of eigenvectors of the operator.

eigenvalues 1 and 0 (or 0 and 1). Thus, for instance,

$$|+_x\rangle = \frac{1}{\sqrt{2}}(|+_z\rangle + |-_z\rangle) \mapsto P^z_+ \frac{1}{\sqrt{2}}(|+_z\rangle + |-_z\rangle) = \frac{1}{\sqrt{2}}|+_z\rangle \tag{9}$$

or

$$|+_x\rangle \mapsto P^z_- \frac{1}{\sqrt{2}}(|+_z\rangle + |-_z\rangle) = \frac{1}{\sqrt{2}}|-_z\rangle \ . \tag{10}$$

(The final state is then thought of as renormalised, i. e. rescaled to the unit vector $|+_z\rangle$ or $|-_z\rangle$, respectively.)

The probability for such a transformation is given by the so-called 'Born rule': it is the modulus squared of the coefficient of the corresponding component in the initial state, or equivalently the squared norm of the (unnormalised) collapsed state. In the case of both (9) and (10) this equals 1/2. (Note that in a Stern–Gerlach experiment what we measure is in fact whether the electron impinges on the upper or lower half of the screen, thus collapsing the state (8) to one of $\frac{1}{\sqrt{2}}|\pm_z\rangle|\psi_\pm\rangle$, each with probability 1/2.)

2 Classical probability (with an eye to quantum mechanics)

We shall now look at the usual Kolmogorovian notion of probability, formulated with an emphasis on aspects relevant for the analogy with quantum probability in the next section.

In the standard formulation of Kolmogorov's axioms, a probability space is a triple $(\Omega, \mathcal{B}, p)$, where Ω is a set, the event space $\mathcal{B}$ is a (σ-)field of subsets of Ω (i. e. closed under complements, (denumerable) intersections and unions, and containing Ω) and p is a normalised (σ-)additive real measure on $\mathcal{B}$, in the sense that

(C1) For all $b \in \mathcal{B}$, $p(b) \in [0,1]$,

(C2) $p(\Omega) = 1$,

(C3) For all finite (or denumerable) families $\{b_i\}$ of mutually disjoint sets in $\mathcal{B}$,

$$p(\bigcup_i b_i) = \sum_i p(b_i) \ .$$

Equivalently, since every (σ-)field of sets is a Boolean (σ-)algebra, and by Stone's theorem every Boolean (σ-)algebra is representable as a (σ-)field of sets, one can take the event space $\mathcal{B}$ to be a Boolean (σ-)algebra, and re-express (C1)–(C3) accordingly:

(C1′) For all $b \in \mathcal{B}$, $p(b) \in [0,1]$,

(C2′) $p(1) = 1$,

(C3′) ll For all finite (or denumerable) families $\{b_i\}$ of mutually disjoint events (i. e. with $b_i \wedge b_j = 0$ for $i \neq j$),

$$p(\bigvee_i b_i) = \sum_i p(b_i) \ ,$$

where we use 0 and 1 to denote also the zero and unit elements of the algebra. In the following, we shall take the set Ω to be discrete (finite or denumerable), and we shall take for simplicity all singletons $\{\omega\} \subset \Omega$ to be measurable.

We shall now introduce a (suitably general) notion of *observable*. Consider first the random variables with values in the unit interval $e : \Omega \to [0,1]$ (so-called *response functions* or *effects*), and as a special case the response functions $\chi : \Omega \to \{0,1\}$ (these are identical with the *characteristic functions* of measurable subsets $\Sigma \subset \Omega$, i. e. functions that take the value 1 on Σ and 0 on $\Omega \setminus \Sigma$). We now define an *observable* as a (finite or denumerable) family of response functions $\{e_i\}_{i \in I}$ such that

$$\sum_{i \in I} e_i = 1 \tag{11}$$

(where 1 is the random variable that is identically 1).[4]

For each such observable, the probability measure p induces a probability measure on (the Boolean algebra generated by) the family of functions $\{e_i\}$ (or on the index set I), which we also denote by p:

$$p(e_i) := \sum_{\omega \in \Omega} e_i(\omega) p(\{\omega\}) \ , \tag{12}$$

[4]One could consider also continuous or partially continuous families of response functions as defining observables (even though the probability space itself is discrete), but we shall ignore them for simplicity, and keep everything discrete.

and for any subset J of I:

$$p(\sum_{i\in J} e_i) := \sum_{i\in J}\sum_{\omega\in\Omega} e_i(\omega)p(\{\omega\}) \ . \tag{13}$$

(This is correctly normalised because of (11) and (C3$'$).) In the special case in which all e_i are 'sharp' ($e_i(1-e_i)=0$, where 0 is the random variable that is identically 0) — i. e. characteristic functions —, we see that the probabilities are just the measures of the sets defined by the characteristic functions $\sum_{i\in J} e_i$, so that the 'sharp observables' are in bijective correspondence with the (finite or denumerable) partitions of Ω, and 'measuring' a sharp observable is simply a procedure for distinguishing between the events forming such a partition.

General observables (at least in the classical case) can be interpreted as noisy or fuzzy or unsharp versions of sharp observables. Indeed, take the following observable, given by the resolution of the identity

$$\sum_{\omega\in\Omega} \chi_{\{\omega\}} = 1 \tag{14}$$

(we can call this the *finest sharp observable*). Now, every effect e can be written as

$$\sum_{\omega\in\Omega} e(\omega)\chi_{\{\omega\}} \ . \tag{15}$$

Since for any observable $\{e_i\}$, in particular also for (14), the probability of each e_i is given by (12), we have that

$$p(\chi_{\{\omega\}}) = \sum_{\omega'\in\Omega} \chi_{\{\omega\}} p(\{\omega'\}) = \chi_{\{\omega\}} p(\{\omega\}) = p(\{\omega\}) \ , \tag{16}$$

and thus

$$p(e_i) = \sum_{\omega\in\Omega} e_i(\omega)p(\chi_{\{\omega\}}) \ . \tag{17}$$

And we see that $e_i(\omega)$ can be interpreted as the *conditional probability* for the response e_i in the experiment $\{e_i\}$, given that a (counterfactual) measurement of the finest sharp observable would have yielded ω.

It is important to note that while each experiment has a Boolean structure (the subsets of the index set I form a Boolean algebra), in general these Boolean algebras do *not* correspond to the Boolean subalgebras of $\mathcal{B}$. It is only the Boolean algebras associated with measurements of sharp observables that correspond to Boolean subalgebras of $\mathcal{B}$. (As we shall have again occasion to remark in Section 4, sharp observables have a number of useful properties not shared by general observables.)

There are two further notions we wish to introduce with an eye to the analogy with quantum probability. One is a notion of *compatibility* of observables. To this end, we first introduce the *coarse-graining* of observables: the observable $\{e_i\}_{i\in I}$ is a coarse-graining of the observable $\{g_k\}_{k\in K}$ iff there is a partition of the index set $K = \cup_{i\in I} K_i$ such that for all $i \in I$,

$$e_i = \sum_{k\in K_i} g_k \tag{18}$$

(note that every sharp observable is indeed a coarse-graining of the finest sharp observable (14)). Clearly, any experiment that measures $\{g_k\}$ also measures $\{e_i\}$. We now call two observables $\{e_i\}$ and $\{f_j\}$ *compatible* iff there is an observable $\{g_k\}$ such that $\{e_i\}$ and $\{f_j\}$ are both coarse-grainings of $\{g_k\}$. The observable $\{g_k\}$ is called a *joint observable* for (or a *joint fine-graining* of) $\{e_i\}$ and $\{f_j\}$.

Obviously, *any two classical observables are compatible.* Indeed, given any two observables $\{e_i\}$ and $\{f_j\}$, we can define a joint observable $\{g_k\}$ simply as $\{e_i f_j\}_{(i,j)}$ (the indices ranging over those pairs (i, j) for which the product $e_i f_j \neq 0$). In the special case of sharp observables, the product $e_i f_j$ is of course the characteristic function of the intersection of the sets defined by e_i and f_j, and the joint observable corresponds simply to the Boolean sub-algebra of $\mathcal{B}$ generated by the union of the two subalgebras corresponding to $\{e_i\}$ and $\{f_j\}$.[5]

The other notion we introduce with an eye to quantum probability is that of a *state*, defined as a *family of overlapping probability measures* over the outcomes of all possible experiments, i. e. a mapping p from the response functions to the reals, such that:

(C1$''$) For all e, $p(e) \in [0, 1]$,

[5] The definition of the joint of any two observables $\{e_i\}$ and $\{f_j\}$ extends to that for finite sets of observables. In the case of the sharp observables it is easy to see that arbitrary sets of observabes are jointly compatible, because they are all Boolean subalgebras of the same Boolean algebra $\mathcal{B}$ (corresponding to the finest sharp observable (14)).

(C2″) $p(1) = 1$,

(C3″) For all (finite or denumerable) families $\{e_i\}$ of effects with $\sum_i e_i \leq 1$,

$$p(\sum_i e_i) = \sum_i p(e_i) \ .$$

We have already seen that a classical probability measure p induces a probability measure on every classical observable and thus defines a state in this sense. Conversely, the family of the probability measures on all observables fixes the original probability measure uniquely, because the original probability measure is nothing other than the probability measure associated with the finest sharp observable (14). Note, finally, that if we consider all possible states p, we can identify the response functions e with the affine mappings from the states into $[0, 1]$ defined by

$$e : p \mapsto p(e) \ . \tag{19}$$

3 Quantum mechanics (with an eye to probability)

As mentioned in Section 1, the formalism of quantum mechanics is based on the fairly familiar structure of a vector space with scalar product, or technically a Hilbert space (because it is complete in the norm induced by the scalar product – vacuously so in finite dimensions).[6] What interests us in particular (with an eye to highlighting the probabilistic structure of quantum mechanics) are the notions of 'state' and 'observable' we find in the formalism.

States are associated in the first place with (unit) vectors in the space.[7] As in our example of a Stern–Gerlach measurement and the associated 'collapse' of the state (9–10), experiments are generally and abstractly associated with probabilistic transformations of the states, corresponding to the

[6]The Hilbert spaces used in standard quantum mechanics are over the complex numbers, and they are separable (i. e. they always have either a finite or a denumerable basis).

[7]The notation we use is the so-called Dirac notation, in which scalar products are denoted by angle brackets, $\langle\varphi|\psi\rangle$, and vectors ('kets') are denoted by right half-brackets, $|\psi\rangle$ (a left half-bracket, $\langle\varphi|$, denotes the linear functional ('bra') assigning to each vector $|\psi\rangle$ the complex number $\langle\varphi|\psi\rangle$).

idea that 'measurements' induce an irreducible disturbance of a quantum system:

$$|\psi\rangle \mapsto A|\psi\rangle \tag{20}$$

(where the right-hand side should be thought of as suitably renormalised).[8] The probability for a transition of the form (20) is given by

$$\langle\psi|A^*A|\psi\rangle \ , \tag{21}$$

which is the scalar product of the vector $A|\psi\rangle$ with itself.[9] The operator A in (20) is arbitrary (in particular not necessarily unitary or self-adjoint), the only restriction being that (21) be no greater than 1. Note that the probability (21) depends only on the product A^*A and not on the specific transformation A (indeed, one can construct infinitely many other operators B such that $B^*B = A^*A$). Operators of the form $E = A^*A$ for some transformation A are called 'effects'.[10]

Each experiment will include a number of alternative transformations that could possibly take place:

$$|\psi\rangle \mapsto A_i|\psi\rangle \ , \tag{22}$$

whose combined probability equals 1:

$$\sum_i \langle\psi|A_i^*A_i|\psi\rangle = 1 \ . \tag{23}$$

This is required to hold for all possible unit vectors $|\psi\rangle$, so in fact we have

$$\sum_i A_i^*A_i = \mathbf{1} \ , \tag{24}$$

[8] As we point out at the end of this section, one can consider even more general transformations, but that does not in the least affect the generality of what follows.

[9] The operator A^*, called the adjoint of A, is the operator defined by $\langle\varphi|A^*\psi\rangle = \langle A\varphi|\psi\rangle$ (ignoring niceties about domains of definition, which become vacuous if the Hilbert space is finite-dimensional). Self-adjoint operators are operators with $A^* = A$. Note that $(AB)^* = B^*A^*$, so that in particular an operator of the form A^*A is self-adjoint. Note also that for A self-adjoint, the mapping $A \mapsto \langle\psi|A\psi\rangle$ (normally written $A \mapsto \langle\psi|A|\psi\rangle$) is a linear functional onto the positive reals that is normalised, in the sense that $\langle\psi|\mathbf{1}|\psi\rangle = 1$ (with $\mathbf{1}$ the identity operator), and is continuous with respect to the so-called operator norm (again vacuously so in finite dimensions).

[10] More explicitly, effects are operators E such that both $E = A^*A$ and $1 - E = B^*B$ for some operators A, B (thus ensuring that the expression (21) is between 0 and 1).

or, writing $E_i := A_i^* A_i$,

$$\sum_i E_i = \mathbf{1} \ , \tag{25}$$

where $\mathbf{1}$ is the identity operator. Each such (finite or denumerable) family of effects $\{E_i\}_{i \in I}$, or 'resolution of the identity' (25) is called an *observable*, quantum mechanical effects being the formal analogue of classical response functions.[11]

Note that the probability of an effect (in any given state) is independent of which observable the effect is part of, i. e. which family of alternative transformations is being implemented in a particular experiment. We shall return to this 'non-contextuality' of probabilities in Sections 5 and 6 below.[12] Suffice it to say now that it is a non-trivial feature because, unlike the classical case, the same effect could be part of two mutually *incompatible* observables.

There is more than one definition of (in)compatibility in the literature, but the following one (on which we have modelled the definition of Section 2) is the most suited to our purposes (see e. g. Cattaneo *et al.* 1997). Define an observable $\{E_i\}_{i \in I}$ to be a *coarse-graining* of the observable $\{G_k\}_{k \in K}$ iff there is a partition of the index set $K = \cup_{i \in I} K_i$ such that for all $i \in I$,

$$E_i = \sum_{k \in K_i} G_k \ . \tag{26}$$

Any experiment that measures $\{G_k\}$ also measures $\{E_i\}$. As in the classical case, we call two observables $\{E_i\}$ and $\{F_j\}$ *compatible* iff there is an observable $\{G_k\}$ such that $\{E_i\}$ and $\{F_j\}$ are both coarse-grainings of $\{G_k\}$. The observable $\{G_k\}$ is called a *joint observable* for (or a *joint fine-graining* of) $\{E_i\}$ and $\{F_j\}$. Compatibility of two observables can be easily generalised to joint compatibility of arbitrary sets of observables.

[11] As in the classical case, one could consider more general observables (even in finite dimensions), in which the sum in (25) is replaced or supplemented by an integral. For simplicity, however, we consider only discrete resolutions of the identity.

[12] More generally, the term non-contextuality is used to denote independence not only of the observable measured but also of any details of the measurement context (Shimony (1984) calls these, respectively, 'algebraic' and 'environmental' contextuality). Unless we have a very weird theory, we can presumably assume that the observable measured is fixed by the details of the experiment (which in quantum mechanics also determine which transformation implements any particular effect). Cf. also our distinction between 'observables' and 'experiments' in Section 4.

The definition of an observable in any (older) textbook on quantum mechanics is as a self-adjoint operator A, i.e. an operator with $A^* = A$. But this traditional definition corresponds to a special case of the one above. Self-adjoint operators are diagonalisable, in the sense that $A = \sum_i a_i P_i$ with real a_i (the eigenvalues) and $\{P_i\}$ a family of projections (self-adjoint operators with $P_i^2 = P_i$, or $P_i(\mathbf{1} - P_i) = \mathbf{0}$, where $\mathbf{0}$ is the zero operator) that are mutually orthogonal ($P_i P_j = 0$ for $i \neq j$).[13] Thus, each self-adjoint operator is associated with a unique 'projection-valued observable', i.e. a resolution of the identity (25), in which the effects E_i are in fact projections (they are 'sharp', meaning $E_i(\mathbf{1} - E_i) = \mathbf{0}$), and which is finite if the Hilbert space is finite-dimensional. Note also that a measurement of such a 'sharp observable' can be implemented by taking $A_i = P_i$, since $P_i^* P_i = P_i$, i.e. each state is transformed to an eigenstate of the measured observable. This is the usual 'collapse postulate' or 'projection postulate' of textbook quantum mechanics, corresponding to a 'minimally disturbing' measurement of a sharp observable.[14]

Compatibility of two sharp observables A and B is traditionally defined as their *commutativity*, i.e. $AB = BA$. This is equivalent to the commutativity of the elements of the respective resolutions of the identity, $\sum_i P_i = \mathbf{1}$ and $\sum_j Q_j = \mathbf{1}$, i.e.

$$P_i Q_j = Q_j P_i \qquad \text{for all } i, j \ . \tag{27}$$

In this case, a joint (projection-valued) observable $\{R_k\}$ is given by $\{P_i Q_j\}_{(i,j)}$ (the indices (i, j) ranging over those pairs for which $P_i Q_j \neq 0$). Indeed, trivially,

$$P_i = \sum_n P_i Q_n \qquad \text{and} \qquad Q_j = \sum_m P_m Q_j \ . \tag{28}$$

This argument generalises to show that finite sets of pairwise compatible

[13] This is the so-called spectral theorem. In infinite dimensions the sum is generally to be replaced or supplemented by an integral. The set of (generalised) eigenvalues is called the spectrum of the operator. Each discrete eigenvalue a_i has corresponding eigenvectors, i.e. vectors such that $A|\psi\rangle = a_i|\psi\rangle$, and if the whole spectrum is discrete there exists an orthonormal basis of the Hilbert space consisting of eigenvectors of A. In finite dimensions, the number of terms in the diagonal decomposition is bounded by the dimension n of the Hilbert space, because there cannot be a set of more than n mutually orthogonal (eigen)vectors. Note that the spectrum of a projection operator is just the set $\{0, 1\}$.

[14] Note that the corresponding transition probabilities reduce to the form $\langle\psi|P_i|\psi\rangle$, i.e. the scalar product between the initial state $|\psi\rangle$ and final state $P_i|\psi\rangle$. The value of the scalar product $\langle\psi|\varphi\rangle$ is often referred to as *the* transition probability between the vectors $|\psi\rangle$ and $|\varphi\rangle$.

sharp observables possess a joint projection-valued resolution of the identity.[15] Indeed, since in finite dimensions all diagonal decompositions are discrete, one can generalise it further to arbitrary sets of pairwise commuting operators.

Effect-valued and projection-valued observables are the analogues, respectively, of the general and sharp classical observables of Section 2, and, at least in some cases, also effect-valued observables can be interpreted as 'unsharp' versions of projection-valued ones. Since every effect E is a self-adjoint operator, it is itself diagonalisable as

$$E = \sum_k e_k P_k \ , \tag{29}$$

where all the eigenvalues e_k are positive and lie in the interval $[0, 1]$.[16] Now suppose all effects E_i in an observable (25) *commute*: there will then exist a single projection-valued resolution of the identity $\{R_k\}$, such that every E_i can be written as

$$E_i = \sum_k e^i_k R_k \ . \tag{30}$$

with suitable coefficients e^i_k. The probability of E_i in the state $|\psi\rangle$ is

$$\langle\psi|E_i|\psi\rangle = \sum_k e^i_k \langle\psi|R_k|\psi\rangle \ , \tag{31}$$

and we can again at least formally identify the coefficient e^i_k as the *conditional probability* that the measurement of $\{E_i\}$ yields i given that a measurement of $\{R_k\}$ would have yielded k.[17] Thus, we can think of a commutative

[15] This property of the sharp observables in quantum mechanics is known as 'coherence'. See also Section 5 below.

[16] Indeed, an equivalent definition of an effect is as a self-adjoint operator with spectrum in the interval $[0, 1]$.

[17] We have the correct normalisation because the E_i form a resolution of the identity (25), so that

$$\sum_i \sum_k e^i_k \langle\psi|R_k|\psi\rangle = \sum_i \langle\psi|E_i|\psi\rangle = \langle\psi| \sum_i E_i|\psi\rangle = 1 \ ,$$

and choosing $|\psi\rangle$ to be an eigenstate with eigenvalue 1 of, say, R_{k_0}, we get

$$1 = \sum_i \sum_k e^i_k \langle\psi|R_k|\psi\rangle = \sum_i \sum_k e^i_k \delta_{kk_0} = \sum_i e^i_{k_0} \ ,$$

independently of the choice of k_0.

effect-valued observable as (at least probabilistically equivalent to) a 'noisy' or 'fuzzy' or 'unsharp' measurement of an associated projection-valued observable.

Further, given our definition of joint observables, we can also understand the general case of a *non-commutative* effect-valued observable $\{E_i\}_{i\in I}$ as a *joint observable* for the generally denumerably many *commutative* effect-valued observables $\{E_i, \mathbf{1}-E_i\}$, one for each $i \in I$.[18] This gives us a further insight on compatibility and incompatibility — namely that incompatible observables can be made compatible if one is willing to introduce enough 'noise' in one's measurements.[19]

While in a sense we can thus reduce the effect-valued observables to the projection-valued ones (and the question of incompatibility essentially to that of incompatibility for projection-valued observables), it makes very good sense to work with the more general effect-valued ones. To quote three reasons: effect-valued observables are needed for modelling realistic experiments; the concatenation of two experiments is clearly an experiment, but cannot generally be represented by a projection-valued observable; and no measurement of a projection-valued observable can fully determine a quantum state, while — precisely because effect-valued observables can combine probabilistic information from incompatible projection-valued ones, even if noisily — there are so-called 'informationally complete' effect-valued observables, whose measurement statistics completely determine the quantum state.[20]

Let us now return to the quantum states themselves. We can think of a quantum state as defining a *family of overlapping probability measures* over the outcomes of all possible experiments. More precisely, we can identify states $|\psi\rangle$ with mappings p_ψ from the effects to the reals, such that:

[18] See Cattaneo *et al.* (1997) for a fuller discussion of both the commutative and the non-commutative case (in finite dimensions).

[19] Indeed, let $\{E_i\}$ and $\{F_j\}$ be incompatible observables. Then the observable $\{\frac{1}{2}E_1, \frac{1}{2}F_1, \frac{1}{2}E_2, \frac{1}{2}F_2, \ldots\}$ is trivially a joint observable for $\{\frac{1}{2}\mathbf{1}, \frac{1}{2}E_1, \frac{1}{2}E_2, \ldots\}$ and $\{\frac{1}{2}\mathbf{1}, \frac{1}{2}F_1, \frac{1}{2}F_2, \ldots\}$, which are 'noisy' versions of $\{E_i\}$ and $\{F_j\}$.

[20] Any effect-valued observable can in fact be implemented by letting the system interact appropriately with an ancillary system and then performing an appropriate measurement of a projection-valued observable on the ancillary system (so-called Naimark dilation), much like one measures spin by coupling it to the spatial degrees of freedom and then measuring position. But we shall not need this in the following.

(Q1) For all E, $p_\psi(E) = \langle\psi|E|\psi\rangle \in [0,1]$,

(Q2) $p_\psi(\mathbf{1}) = \langle\psi|\mathbf{1}|\psi\rangle = 1$,

(Q3) For all (finite or denumerable) families $\{E_i\}$ of effects with $\sum_i E_i \leq \mathbf{1}$,

$$p_\psi(\sum_i E_i) = \sum_i p_\psi(E_i) \ .$$

This definition can of course be restricted to projection operators (with (Q3) restricted to families of mutually orthogonal projections). If one does so, a famous theorem due to Gleason (1957) (valid for Hilbert spaces of dimension $n \geq 3$) shows that the most general state ρ on the projections is an *arbitrary convex combination* (i. e. a weighted average) of vector states p_ψ. *A fortiori* this is true for states defined on effects (and the direct proof for this case is much simpler (Busch 2003)). The most general quantum states thus form a convex set, with the vector states as its extremal points.[21]

Perhaps surprisingly, these general quantum states cannot *uniquely* be decomposed as convex combinations of vector states. (We shall not have the space to develop this point, but it is a very important disanalogy with the classical case.) This can be seen very easily in the special case of a spin-1/2 system (described by a 2-dimensional Hilbert space), using a rather beautiful geometric representation, the so-called Poincaré sphere (or Bloch sphere). The unit vectors form a 2-dimensional complex sphere, and this is affinely isomorphic to a 3-dimensional real sphere. That is, one can map the two bijectively in a way that preserves convex combinations. This allows one to associate the abstract 'spin' states with directions in 3-dimensional space: an electron has spin 'up' in the direction r iff its abstract spin state is mapped to the spatial vector r under this affine mapping, and 'down' iff it is mapped to the spatial vector $-r$. One can then see directly that any point in the interior of the sphere (corresponding to a general quantum state) can be written in infinitely many ways as a convex combination of points on the surface of the sphere (corresponding to vector states). Indeed, any straight line through a point in the interior will intersect the surface of the sphere in two points, thus defining a convex decomposition of the interior point, but there are infinitely many such straight lines. The centre of the

[21] General quantum states are formally represented by so-called 'density operators' ρ, and the probabilities they define on the effects E can be expressed using the 'trace functional', $\mathrm{Tr}(\rho E)$ (see footnote 2 for standard references).

sphere corresponds to the so-called maximally mixed state, which assigns equal probabilities $\frac{1}{2}$ to spin up or down in any spatial direction r.

Note that, in turn, effects can be identified with the mappings from the states into $[0, 1]$ defined by

$$E : |\psi\rangle \mapsto p_\psi(E) \ . \tag{32}$$

These mappings are also affine, i. e. map convex combinations of states to the same convex combination of the corresponding probabilities.[22]

We are now ready to sketch a generalisation of the notion of probability encompassing both classical and quantum probability (Section 4), and to discuss whether it is indeed a non-trivial generalisation of the classical notion (Sections 5 and 6).

4 Generalised probability (a sketch)

There are different (but largely convergent) approaches to defining a theory of probability generalising both classical and quantum probability. The seminal work in this area is due to Mackey (1957), who inaugurated what is known today as the 'convex set' approach to quantum and generalised probability, which axiomatises directly pairs of state–observable structures (Mackey 1963, Varadarajan 1968, Beltrametti and Cassinelli 1981).

Alternative more concrete routes to generalised probability have been pursued in particular by Ludwig (1954, 1985) and his group, and by Foulis and Randall with their work on test spaces (starting with Randall and Foulis (1970, 1973), and Foulis and Randall (1972, 1974); see also the review by Wilce (2000)).

Research in quantum logic opened yet further routes into generalised probability by providing various generalisations of Boolean algebras (orthomodular lattices, orthomodular posets, partial Boolean algebras, orthoalgebras, effect algebras etc.), thus generalising the event spaces of classical

[22] Once the notion of state has been generalised, it makes sense to generalise also the notion of state transformation, which in (20) was restricted to those transformations that map vector states to vector states. This does not lead, however, to a further generalisation of the notion of observable: effect-valued observables are still the most general ones, and can be used to completely determine a general quantum state.

probability. The main lines of research in this tradition stem, respectively, from the classic paper by Birkhoff and von Neumann (1936), which inaugurated the lattice-theoretic version of quantum logic (with its emphasis on weakening the distributive law, originally in favour of modularity and then of orthomodularity), and from the work on partial Boolean algebras by Specker (1960) and Kochen and Specker (1965a,b, 1967) (with its emphasis on partial operations).[23]

I shall deliberately ignore these distinctions and sketch instead a somewhat pedagogical version of generalised probability theory — drawing from elements of these various approaches, and having the advantage of being fairly simple and of leading rather naturally to the abstract notions of effect algebras and orthoalgebras (possibly the best current candidates for providing an abstract setting for a generalised probability theory). While much of what follows can be generalised or ought to be generalisable to the denumerable case, in this section I shall focus exclusively on the finite case.

Let us start with the quasi-operational idea of a set $\mathcal{A}$ of *experiments*. Each experiment $A \in \mathcal{A}$ is characterised by a Boolean algebra $\mathcal{B}_A$ of experimental outcomes $e_A \in \mathcal{B}_A$. We further consider *states* $p \in \mathcal{P}$, which define probability measures over the outcomes of all possible experiments, $p(e_A)$. If an outcome e_A of one experiment can be somehow identified with an outcome e_B of a different experiment, one will obviously require that $p(e_A) = p(e_B)$ for all states p, i. e. the probability measures induced by the states can generally overlap. Less obviously, we will identify *all* pairs of experimental outcomes that are equiprobable in all states. This will be the first of only very few substantial requirements. Thus we will define *effects* as equivalence classes $e = [e_A]$ of experimental outcomes under the equivalence relation $\sim$ of equiprobability for all p:

$$e_A \sim e_B \quad :\Leftrightarrow \quad p(e_A) = p(e_B) \text{ for all } p \ . \tag{33}$$

We denote the set of all effects by $\mathcal{E}$. In Section 6 we shall return to the question of identifying outcomes of different experiments. Suffice it to say at this stage that identifying equiprobable outcomes means we are thinking of effects as characterised by what they can tell us about the states.[24] Note

[23]For useful discussions and reviews of aspects of quantum logic and generalised probability, see also Hardegree and Frazer (1981), Hughes (1985), Coecke, Moore and Wilce (2001), Wilce (2012), Darrigol (forthcoming), and the collections edited by Hooker (1975), Marlow (1978) and Coecke, Moore and Wilce (2000).

[24]To take a fully operationalist stance on this question would mean to identify events

that each effect e defines an affine mapping from the states to the unit interval:

$$e : \mathcal{P} \to [0,1], \quad p \mapsto p(e) \ . \tag{34}$$

It may be convenient to require that every such mapping corresponds to an effect, but for our limited purposes we shall not do so.

Now, for any two experiments $A, B \in \mathcal{A}$,

$$0_A \sim 0_B \quad \text{and} \quad 1_A \sim 1_B \tag{35}$$

(where 0_A and 1_A are the 0 and 1 elements of the Boolean algebra $\mathcal{B}_A$, and similarly for $\mathcal{B}_B$). Indeed, $p(0_A) = 0$ and $p(1_A) = 1$ for all p, independently of A. We can thus define an effect 0 and an effect 1 as

$$0 := [0_A] \quad \text{independently of } A \tag{36}$$

and

$$1 := [1_A] \quad \text{independently of } A \ . \tag{37}$$

Similarly, for any $A, B \in \mathcal{A}$, if $e_A \sim e_B$ then $\neg e_A \sim \neg e_B$ (where $\neg$ denotes negation in the relevant Boolean algebra). Indeed, if $p(e_A) = p(e_B)$ for all p, then

$$p(\neg e_A) = 1 - p(e_A) = 1 - p(e_B) = p(\neg e_B) \quad \text{for all } p \ , \tag{38}$$

and for any effect e we can define a unique effect $e^{\perp}$ as

$$e^{\perp} := [\neg e_A] \quad \text{independently of } A \ . \tag{39}$$

Clearly, for any e, we have $e^{\perp\perp} = e$. Note, however, that it is perfectly possible for some e that $e^{\perp} = e$ (this is the case for both the response function $\frac{1}{2}1$ in classical probability and the effect $\frac{1}{2}\mathbf{1}$ in quantum probability).

---

with results of laboratory procedures, such as letting an electron pass through an inhomogeneous magnetic field and observing it hit the upper (or lower) half of a screen; or opening a box and finding (or failing to find) a gem inside it. But this will clearly not do. Despite the lip service to operationalism, such procedures will always abstract away from aspects of the experimental setting deemed irrelevant, e. g. whether or not I am performing the experiment standing on one leg. But to deem any detail of the experimental arrangement irrelevant is already to make a theoretical decision. To quote two suggestive examples: it might be important whether the magnetic field is stronger at the north or the south pole of the magnet (this detail is irrelevant in standard quantum mechanics, but makes all the difference in the description of spin measurements in de Broglie and Bohm's pilot-wave theory, as discussed e. g. by Albert (1992)); or it might be important whether one opens box A together with box B or together with box C (we shall discuss this example explicitly in Section 6).

The states naturally induce a *partial ordering* on the effects — which will be a useful tool in the following — defined as

$$e \leq f \quad :\Leftrightarrow \quad p(e) \leq p(f) \text{ for all } p \ . \tag{40}$$

Note that

$$e \leq f \Leftrightarrow f^\perp \leq e^\perp \ . \tag{41}$$

We now introduce two important notions: compatibility and orthogonality of effects.

Two effects e and f are *compatible*, written $e\$f$, iff there is an experiment A and outcomes $e_A \in \mathcal{B}_A$ and $f_A \in \mathcal{B}_A$ such that $e = [e_A]$ and $f = [f_A]$. That is, two effects are compatible iff they can be measured in a single experiment. The definition of compatibility can be trivially extended to finite sets of effects.

Two effects e and f are *orthogonal* (or *disjoint*), written $e \perp f$, iff $e \leq f^\perp$, i. e. $p(e) \leq 1 - p(f)$ for all p. Note that this relation is symmetric, but generally not irreflexive (since it is possible that $e^\perp = e$). We can generalise also orthogonality to finite sets of effects, by defining a family $\{e_i\}$ of effects to be jointly orthogonal iff $\sum_i p(e_i) \leq 1$ for all p.

Note that if there are experimental outcomes e_A and f_A in some $A \in \mathcal{A}$ such that $e_A \leq f_A$ with respect to the partial order on the Boolean algebra $\mathcal{B}_A$, then $p(e_A) \leq p(f_A)$ for all p, and therefore $[e_A] \leq [f_A]$ with respect to the partial order on the effects.

We shall *require*, conversely, that if $e \leq f$ for two effects e and f, then there exist at least one experiment A and experimental outcomes $e_A \in e$ and $f_A \in f$, such that $e_A \leq f_A$ in the Boolean algebra $\mathcal{B}_A$. This is the second of our substantive requirements.

It follows in particular that *comparable effects are compatible*, that is, that $e \leq f$ implies that e and f are in fact compatible. Since $e_A \leq f_A$ for some A means that $p(f_A \wedge \neg e_A) = p(f_A) - p(e_A)$ for all p, it also follows that if $e \leq f$, there is an effect $f \ominus e := [f_A \wedge \neg e_A]$ (independently of any particular A with $e_A \leq f_A$), which is jointly compatible with e and f, and such that $p(f \ominus e) = p(f) - p(e)$ for all p.

By the requirement above we also have that *orthogonal effects are compatible*. Indeed, $e \perp f$ means that e and $f^\perp$ are comparable, and thus compatible. But if $e_A \in e$ and $\neg f_A \in f^\perp$ are in the same Boolean algebra $\mathcal{B}_A$, then so are e_A and f_A, thus e and f are compatible. It also follows that if $e \perp f$, i. e. $p(e) + p(f) \leq 1$ for all p, there is an effect $e \oplus f$ jointly compatible with e and f, such that $p(e \oplus f) = p(e) + p(f)$ for all p. Indeed, given the above, we can define

$$e \oplus f := (f^\perp \ominus e)^\perp = [e_A \vee f_A] \tag{42}$$

(independently of any particular A with $e_A \leq f_A$), which is jointly compatible with e and f, and such that

$$p(e \oplus f) = p(e) + p(f) \quad \text{for all } p \ . \tag{43}$$

Note that for $e \leq f$ we thus have

$$f = e \oplus (f \ominus e) \ , \tag{44}$$

or

$$f = e \oplus (e \oplus f^\perp)^\perp \tag{45}$$

(the so-called 'effect algebra orthomodular identity').

As our third and last substantive requirement, we shall strengthen the above so that for (finite) *ordered chains* of effects $e^1 \leq e^2 \leq e^3 \leq \ldots$, there exist an experiment A and experimental outcomes $e^1_A \in e^1$, $e^2_A \in e^2$, etc., such that $e^1_A \leq e^2_A \leq e^3_A \leq \ldots$ in the Boolean algebra $\mathcal{B}_A$, so that in particular the effects in the chain are jointly compatible. (In the rest of this section, whenever we write 'ordered chain' we shall mean 'finite ordered chain', and similarly for 'jointly orthogonal set'.)

If ordered chains are jointly compatible it follows that also *jointly orthogonal sets* of effects are jointly compatible. Indeed, given a jointly orthogonal set of effects $\{e^i\}_{i=1}^N$, the sequence of effects

$$e^1, \ e^1 \oplus e^2, \ (e^1 \oplus e^2) \oplus e^3, \ \ldots \tag{46}$$

is an ordered chain, and so is a jointly compatible set. But if $e^1_A, e^1_A \vee e^2_A, \ldots$ are in the same Boolean algebra $\mathcal{B}_A$, so are $e^1_A, e^2_A, \ldots$, and the original set $\{e^i\}$ is jointly compatible.

A (finite) observable on $\mathcal{E}$ can now be defined simply as a jointly orthogonal set (for which one automatically has $\bigoplus_i e_i \leq 1$), and a state $p \in \mathcal{P}$ can be identified with a mapping from the effects to the reals, such that:

(G1) For all $e \in \mathcal{E}$, $p(e) \in [0,1]$,

(G2) $p(1) = 1$,

(G3) For all jointly orthogonal sets $\{e_i\}$ of effects,

$$p(\bigoplus_i e_i) = \sum_i p(e_i) \ .$$

Since jointly orthogonal sets are jointly compatible, to each such observable on $\mathcal{E}$ there corresponds at least one experiment $A \in \mathcal{A}$. Coarse-graining and compatibility of observables can be defined as above, and compatibility of two effects e and f is trivially equivalent to compatibility of the two observables $\{e, 1-e\}$ and $\{f, 1-f\}$.

We are now in a position to show that the structure $(\mathcal{E}, 0, 1, \oplus)$ is an *effect algebra*, that is, a structure with two distinguished elements 0 and 1 and a partial operation $\oplus$ (defined on a subset of $\mathcal{E} \times \mathcal{E}$), satisfying the following axioms:[25]

(E1) The partial operation $\oplus$ is commutative, i. e. if $e \oplus f$ is defined, so is $f \oplus e$, and $e \oplus f = f \oplus e$.

(E2) The partial operation $\oplus$ is associative, i. e. if $e \oplus f$ and $(e \oplus f) \oplus g$ are defined, so are $f \oplus g$ and $e \oplus (f \oplus g)$, and $(e \oplus f) \oplus g = e \oplus (f \oplus g)$.

(E3) For any e, there is a unique element $e^\perp$ such that $e \oplus e^\perp = 1$.

(E4) If $e \oplus 1$ is defined, then $e = 0$.

(The elements of an abstract effect algebra are also called effects.)

Proof:
Define $\oplus$ as above. Since the relation $\perp$ is symmetric, if $e \oplus f$ is defined, so is $f \oplus e$, and (E1) follows because

$$p(e) + p(f) = p(f) + p(e) \quad \text{for all } p \ . \tag{47}$$

[25] There are some slightly different but equivalent axiomatisations of effect algebras, which were introduced more or less independently by a number of authors: by Giuntini and Greuling (1989) under the name 'weak orthoalgebras' or 'generalised orthoalgebras', by Kôpka (1992) under the name 'D-posets', and by Foulis and Bennett (1994) under the name 'effect algebras'. Here I follow the latter.

Next, assume that $e \oplus f$ and $(e \oplus f) \oplus g$ are defined, i. e.

$$p(e) \leq 1 - p(f) \quad \text{and} \quad p(e) + p(f) \leq 1 - p(g) \quad \text{for all } p \ . \tag{48}$$

Then also

$$p(f) \leq 1 - p(g) \quad \text{and} \quad p(f) + p(g) \leq 1 - p(e) \quad \text{for all } p \ , \tag{49}$$

i. e. also $f \oplus g$ and $e \oplus (f \oplus g)$ are defined, and (E2) follows because

$$\Big(p(e) + p(f)\Big) + p(g) = p(e) + \Big(p(f) + p(g)\Big) \ . \tag{50}$$

(Associativity of $\oplus$ was already implicit when we showed above that orthogonal sets are jointly compatible, rather than only orthogonal sequences.)

Further, for each e the unique element satisfying (E3) is the element $e^{\perp}$ defined by (39): clearly $e \oplus e^{\perp}$ is defined and $e \oplus e^{\perp} = 1$; and because of (43), if there are two effects f and f' both satisfying

$$p(e \oplus f) = p(e \oplus f') = 1 \qquad \text{for all } p \ , \tag{51}$$

then $p(f) = p(f')$ for all p, hence $f = f'$, and (E3) follows.

Finally, if $e \oplus 1$ is defined then $p(e \oplus 1) = p(e) + p(1) = p(e) + 1$ for all p, but since $0 \leq p(e), p(e \oplus 1) \leq 1$, we have $p(e) = 0$ for all p, and (E4) follows. *QED.*

Note that in any effect algebra, we can abstractly define a partial order $e \leq f$ as: there is a g such that $e \oplus g = f$. Given our definition of $\oplus$, it follows from (43) and (44) above that our previous definition of the partial order on $\mathcal{E}$ coincides with the abstract one.

Similarly, in any effect algebra one can abstractly define relations of orthogonality and compatibility. Two effects e and f are orthogonal in the abstract sense iff $e \oplus f$ is defined, and a finite set of effects $\{e_i\}$ is jointly orthogonal iff $\bigoplus_i e_i$ is defined. Our definition of orthogonality for $\mathcal{E}$ clearly coincides with the abstract one.

As for compatibility, two or finitely many effects $\{f_j\}$ are compatible in the abstract sense iff there is an orthogonal set $\{e_i\}_{i \in I}$ and subsets $I_j \subset I$ such that $f_j = \bigoplus_{i \in I_j} e_i$ for all j (we shall say the family of effects $\{f_j\}$ has an orthogonal decomposition).

It is easy to see that also our definition of compatibility for $\mathcal{E}$ coincides with the abstract one. For instance (and similarly for finitely many effects), if two effects e and f are compatible in our sense above, there is an experiment $A \in \mathcal{A}$ and experimental outcomes $e_A \in e$ and $f_A \in f$ in $\mathcal{B}_A$. In this case the effects $g := [e_A \wedge f_A]$, $h := [e_A \wedge \neg f_A]$ and $i := [\neg e_A \wedge f_A]$ form a jointly orthogonal set with $e = g \oplus h$ and $f = g \oplus i$ (such a 'minimal' orthogonal decomposition is called a Mackey decomposition).[26] Conversely, if two effects e and f are compatible in the abstract sense, they have an orthogonal decomposition (in fact a Mackey decomposition). But orthogonality in the abstract sense coincides with orthogonality in our sense above, and we have already seen that this implies compatibility also in our sense.

Using the partial order, we can finally define *sharp* elements of the effect algebra ('sharp effects' or 'projections'), as those satisfying $e \wedge e^{\perp} = 0$ (meaning that the greatest lower bound of e and $e^{\perp}$ exists and is 0). Since 0 is the minimal element of the partially ordered set (poset) $\mathcal{E}$, this in turn means that *every* lower bound of e and $e^{\perp}$ is 0, i. e.

$$p(f) \leq \min[p(e), 1 - p(e)] \text{ for all } p \quad \Rightarrow \quad f = 0 \ . \tag{52}$$

We can now show that the sharp elements of $\mathcal{E}$ form an *orthoalgebra* $\mathcal{L}$, i. e. in addition to (E1)–(E4) they satisfy also:

(E5) If $e \oplus e$ is defined, then $e = 0$

(in this case, (E4) in fact becomes redundant).

Proof:
Let e be sharp, and let $e \oplus e$ be defined. Then, by (43),

$$p(e \oplus e) = 2p(e) \quad \text{for all } p \ , \tag{53}$$

and thus $p(e) \leq \frac{1}{2}$ for all p. But then $1 - p(e) \geq \frac{1}{2}$ for all p, therefore

$$p(e) = \min[p(e), 1 - p(e)] \quad \text{for all } p \ . \tag{54}$$

Since e is sharp, $e = 0$, and (E5) follows. *QED.*

[26] Note that, once g is given, h and i are uniquely defined by $h = e \ominus g$ and $i = f \ominus g$. However, $g = [e_A \wedge f_A]$ itself is generally not independent of the choice of A, i. e. Mackey decompositions are generally not unique, so that one cannot define a partial operation $\wedge$ in this way. (Some pairs of effects may nevertheless have greatest lower bounds in the sense of the partial order.)

Note that the sharp elements of an arbitrary effect algebra need not form an orthoalgebra in general, so this last result depends in fact on how we have constructed $\mathcal{E}$.

Without further requirements on the experiments and the states, however, we cannot garantee that general observables can be somehow reduced to sharp observables. (Indeed, nothing forces the orthoalgebra of sharp elements of $\mathcal{E}$ to be an interestingly rich structure — there might even be no sharp effects besides 0 and 1!)

One could ask further what conditions one might impose on $\mathcal{E}$ or $\mathcal{L}$ in order to recover classical or quantum probability. In the case of classical probability, it is obvious that $\mathcal{L}$ needs to be a Boolean algebra. In the case of quantum probability, there are some classic partial results going some way towards ensuring that the orthoalgebra of sharp effects be isomorphic to the projections on a complex Hilbert space. For instance, one might impose the following conditions on an orthoalgebra, in increasing order of strength:

(α) Unique Mackey Decomposition (UMD), i. e. compatible pairs are required to have *unique* Mackey decompositions. This ensures that the Boolean structure of sharp experiments coincides where experiments overlap, in particular allowing conjunction and disjunction to be defined globally as partial operations on the orthoalgebra (thereby turning an orthoalgebra into a so-called Boolean manifold).

(β) Orthocoherence, i. e. pairwise orthogonal sets are jointly orthogonal. This ensures that an orthoalgebra is an orthomodular poset.

(γ) Coherence, i. e. pairwise compatible sets are jointly compatible. This ensures that an orthoalgebra is a (transitive) partial Boolean algebra.

These are all properties of the orthoalgebras of sharp effects in both quantum and classical probability (indeed, we have seen this explicitly in the case of coherence).[27]

[27]For good discussions, see Hardegree and Frazer (1981) and Hughes (1985). Note that the *effect* algebras of quantum and classical probability are less well-behaved. For instance, ef and $\min(e, f)$ are alternative elements g defining non-unique Mackey decompositions for any two response functions e and f (and analogously for any two commuting quantum effects E and F). And $\frac{1}{4}\mathbf{1}$, $\frac{1}{3}\mathbf{1}$, $\frac{1}{2}\mathbf{1}$ are pairwise orthogonal quantum effects that fail to be jointly orthogonal (and analogously for multiples of the classical response function 1).

A well-known problem, however, relates to the existence of tensor products, i. e. the possibility of composing generalised probability structures. If enough states exist (in a well-defined sense), one can construct tensor products of orthoalgebras, but forming tensor products tends to destroy orthocoherence. In this respect, the theory still needs to be investigated further.[28]

Fortunately, these are not questions we need to address in order to discuss whether a generalisation of the notion of probability such as the above is indeed necessary, or whether generalised probabilities might after all be embeddable in some suitable classical probability space. This is what we discuss in the next section, and we can do it by looking at very simple examples.

5 Non-embeddability (and no-hidden-variables)

In Section 4, we have sketched a generalised theory of probability that is non-classical in the sense that it allows for incompatibility of observables. None of what we have said so far, however, shows that it is impossible to *embed* such a generalised probabilistic structure into some larger classical probability space, at least under some minimal assumptions.

For instance, if we require that the orthoalgebra $\mathcal{L}$ of sharp effects have the UMD property, then we can define partial Boolean operations on the sharp effects. One might now naïvely imagine *formally* extending the Boolean operations even to pairs of incompatible ones, and defining a probability measure on the thus enlarged event space that should return the original measures as marginals on each of the original Boolean algebras.

[28] See in particular Foulis and Randall (1979, 1981). Other programmes for reconstructing quantum mechanics have been developed in more recent years, some of which have had striking successes (Comte 1996, Hardy 2002, Goyal 2008a,b, Dakić and Brukner 2011, Masanes and Müller 2011, Chiribella, D'Ariano and Perinotti 2011), the most influential of these arguably being Hardy (2002). (See Darrigol (forthcoming) for analysis and comparison of most or all of these approaches.) The same is true of programmes for reconstructing specific striking aspects of quantum mechanics, such as the bounds in the Bell inequalities (perhaps of particular note among these is the research by Cabello and co-workers (Cabello, Severini and Winter 2010, Cabello 2012, 2013)). Making connections between quantum logic and quantum probability on the one hand and some of these newer approaches on the other might eventually prove very fruitful. For a suggestion in this direction, see Bacciagaluppi and Wilce (in preparation).

Indeed, if two sharp effects e and f are not compatible, the joint probability of $e \wedge f$ might not be experimentally meaningful, but by the same token no experiment would constrain us in choosing a measure that specified also these joint probabilities (e. g. by considering e and f as independent) — or so it might seem. And if the orthoalgebra of sharp events is rich enough, one might get back all the original observables by considering unsharp realisations of sharp observables, as in the classical case of Section 2. Even if we considered this a purely formal construction, it would still mean that our 'generalised probability spaces' are indeed embeddable into classical probability spaces, and thus provide only a fairly trivial generalisation of the formalism of classical probability.

Where this argument goes wrong, however, is in failing to realise the richness of the compatibility structure in a general orthoalgebra. The relation of compatibility is clearly reflexive and symmetric, but it is not transitive, so that observables do not fall neatly into different equivalence classes of mutually compatible ones. Put slightly differently, if compatibility is not transitive, it is possible for the same observable to be a coarse-graining of two mutually incompatible observables — which in this sense can be said to be *partially* compatible. And the interlocking structure of partially compatible observables (and the corresponding partially overlapping probability measures) can be surprisingly rich. In such cases the question of whether general probabilistic states might be induced by classical probability measures becomes non-trivial.

We shall now construct a simple master example showing explicitly that probabilistic states cannot in general be induced by classical probability measures. We shall then see how the example relates to classic impossibility theorems ruling out various kinds of 'hidden variables' theories in quantum mechanics (in particular Bell's theorem and the Kochen–Specker theorem).[29]

Imagine that we have three boxes, A, B and C. We can open any box, and find (or fail to find) a gem in it. Let e, f, g be the outcomes 'finding a gem in the box' in each of the three experiments. We further imagine that we can open any two but *not* all three boxes simultaneously. We finally imagine that we have probabilistic states specifying for any pair of boxes the probabilities for finding gems in neither, either or both of the boxes.[30]

[29]In fact our example generalises one given by Albert (1992) for the purpose of illustrating the Bell inequalities.

[30]Cf. Specker's tale of the Sage of Nineveh in Section 6 below.

Now let us take a special state, such that for any ordered pair (x, y) with $x, y \in \{e, f, g\}$:

$$p(x|y) = p(\neg x|\neg y) = \alpha \ , \tag{55}$$

with $\alpha \in [0, 1]$. For $\alpha = 1$, such a state obviously has the form

$$\begin{aligned} p(x \wedge y) &= a \ , \\ p(\neg x \wedge \neg y) &= 1 - a \ , \\ p(x \wedge \neg y) &= 0 \ , \\ p(\neg x \wedge y) &= 0 \end{aligned} \tag{56}$$

for all pairs (x, y), for some $a \in [0, 1]$. And it is equally obvious that for any a, the state is (uniquely) induced by the probability measure defined by

$$p(e \wedge f \wedge g) = a \ , \quad p(\neg e \wedge \neg f \wedge \neg g) = 1 - a \ , \tag{57}$$

and

$$\begin{aligned} p(e \wedge f \wedge \neg g) = p(e \wedge \neg f \wedge g) = p(e \wedge \neg f \wedge \neg g) = \\ p(\neg e \wedge f \wedge g) = p(\neg e \wedge f \wedge \neg g) = p(\neg e \wedge \neg f \wedge g) = 0 \ . \end{aligned} \tag{58}$$

For the case $\alpha \neq 1$ we have instead:

Lemma*:*
A state satisfying (55) *with* $\alpha \neq 1$ *is uniquely given by*

$$p(x \wedge y) = p(\neg x \wedge \neg y) = \frac{\alpha}{2} \ , \tag{59}$$

and

$$p(x \wedge \neg y) = p(\neg x \wedge y) = \frac{1 - \alpha}{2} \tag{60}$$

for any (x, y)*. Thus in particular,*

$$p(x) = p(\neg x) = p(y) = p(\neg y) = \frac{1}{2} \ . \tag{61}$$

We then have a rather striking result:

Proposition*:*
Under the assumptions of the Lemma, p *is induced by a joint probability*

measure on the Boolean algebra (formally) generated by $\{e, f, g\}$ *if and only if* $\alpha \geq \frac{1}{3}$.

The proofs are left for the Appendix. Intuitively, the case $\alpha = 1$ corresponds to perfect correlations between finding or not finding gems in any two boxes, and it is indeed obvious that this state can be extended to a probability measure in which there are perfect correlations between finding gems in all three boxes. To take an intermediate case, $\alpha = \frac{1}{2}$ is the uncorrelated case, in which any two boxes are independent, and is again obviously extendable to the case in which all three boxes are independent. The case $\alpha = 0$ instead is the perfectly anti-correlated case, and is clearly *not* classically reproducible: if whenever there is a gem in the first box there is no gem in the second, and whenever there is no gem in the second there is one in the third, then whenever there is a gem in the first box there also is one in the third, contradicting the hypothesis.[31] In fact, every state with $\alpha < 1$ is a convex combination of the (unique) perfectly anti-correlated state and the (special) perfectly correlated state with $p(e) = p(f) = p(g) = \frac{1}{2}$. While all states with positive correlations, the uncorrelated state, and even some with negative correlations are reproducible classically (all states with $\alpha \geq \frac{1}{3}$), if the perfectly anti-correlated component comes to dominate too strongly, the negative correlations can no longer be reproduced by a classical probability measure.

As we shall now see, a larger set of states *can*, however, be reproduced quantum mechanically (all states with $\alpha \geq \frac{1}{4}$). This shows, indeed, that already the case of quantum probabilities requires generalised probabilites that cannot be embedded in classical probability spaces. It also shows explicitly that quantum probabilities are only a subset of all possible generalised probabilities.[32]

[31] More precisely: the Proposition states the non-existence of any joint probability measure exhibiting perfect anti-correlations between the outcomes of three binary experiments, and this obviously implies the non-existence of trivial probability measures (i. e. ones assigning the probabilities 0 or 1 to all events) exhibiting the same anti-correlations; but the converse is also true, because if a probability measure did exist, it would at least formally be a convex combination of trivial probability measures, all of which would have to exhibit perfect anti-correlations (since any weaker ones would spoil the perfect anti-correlations of any of their convex combinations); the argument in the text establishes the non-existence of such trivial probabiity measures.

[32] Note that by the remark on convex combinations in the previous paragraph, the states that are specifically quantum ($\frac{1}{4} \leq \alpha < \frac{1}{3}$) can be simulated by a convex combination of a classical and a non-classical but non-quantum state. Such techniques can be put to use for instance to analyse the non-locality of a quantum state in terms of the maximally

Take two spin-$\frac{1}{2}$ systems in the so-called singlet state,

$$\frac{1}{\sqrt{2}}(|+\rangle|-\rangle + |-\rangle|+\rangle) \ . \tag{62}$$

In this state, results of spin measurements in the same direction on the two electrons are perfectly anti-correlated. The singlet state (62) is also rotationally symmetric, so that perfect anti-correlations are obtained for pairs of parallel measurements in whatever direction.[33] Pairs of spin measurements on different particles are always compatible, and the joint probability for spin up in direction r on the left and $-r'$ on the right is equal to $\cos^2(\vartheta/2)$, where ϑ is the angle between r and r'. Note that taking the two directions r, r'' on the left and the two directions $-r', -r$ on the right, one obtains four compatible pairs comprising each one direction on the left and one on the right.

Now, if one is attempting to construct a joint probability measure for all four spin observables, then, given the perfect correlations for the pair of measurements in the directions $(r, -r)$, the constraint (55) with $\alpha = \cos^2(\vartheta/2)$ will hold of the joint probabilities for $(r, -r')$ and $(-r', r'')$ (in both cases defined on different sides) and the putative joint probabilities for (r'', r) (defined on the same side).

It is obvious that one can have three spatial directions pairwise spanning the same angle ϑ iff this angle is between 0 (when the three directions are collinear) and 120 degrees (when they are coplanar). This correponds exactly to values of $\cos^2(\vartheta/2)$ between 1 and $\frac{1}{4}$. And it gives us a quantum model for states satisfying (55) with $\frac{1}{4} \leq \alpha \leq 1$.[34]

non-local non-quantum states ('Popescu–Rohrlich boxes') needed for its simulation (see e. g. Cerf *et al.* 2005).

[33] Here we are going slightly beyond the brief description of entangled states given in Section 1. The point to grasp is that measurements on different subsystems are always compatible with each other, so we can consider the probabilities for such joint measuremenst in an entangled state, and it turns out that some such states provide perfect examples of correlations that cannot be reproduced classically.

[34] A related quantum model of the same correlations is the following. Take a single spin-$\frac{1}{2}$ system and take spin states corresponding to spin up in three directions r, r', r'' pairwise spanning the same angle ϑ. The transition probability $\langle\psi|\varphi\rangle$ between any two such spin states is also equal to $\cos^2(\vartheta/2)$. Of course, sharp spin observables in non-collinear directions are not compatible, because the projections on the spin states in different directions do not commute. So these transition probabilities only correspond to conditional probabilities for outcomes of *sequential* minimally disturbing spin measurements. But if we assume the initial state of the electron to be the maximally mixed state (which assigns

We can now make explicit the connection with classic results in quantum mechanics about ruling out various kinds of 'hidden variables' models, in particular the Bell inequalities and the Kochen–Specker theorem.

The best-known Bell inequality is the Clauser–Holt–Shimony–Holt (CHSH) inequality (Clauser *et al.* 1969),

$$-2 \leq E(AB) - E(AB') + E(A'B) + E(A'B') \leq 2 \ , \tag{63}$$

where A, A', B, B' are two-valued observables with the values ± 1, each of A, A' is compatible with each of B, B', and

$$E(XY) = p(X=1, Y=1) - p(X=-1, Y=1) - \\ p(X=1, Y=-1) + p(X=-1, Y=-1) \tag{64}$$

is the correlation coefficient of X and Y. As is well-known, the CHSH inequality (63) can be derived from the assumption of a local hidden variables model when A, A' and B, B' are interpreted, respectively, as pairs of observables pertaining to two (space-like) separated systems, e. g. spin-$\frac{1}{2}$ observables in various directions for two different particles (Bell 1971). (One readily recognises that the quantum version of our example above uses the same set-up, with a special choice of directions.)

It was Fine (1982) who first pointed out that (63) is also the necessary and sufficient condition for the existence of a joint probability measure for the observables A, A', B, B' when the marginals for the four compatible pairs are given. Such a joint probability measure is known as a 'non-contextual hidden variables' model of the experimental situation, since the same measure returns the correct marginals irrespective of how an observable is paired with an observable on the other side, or indeed on how an observable is assumed to be measured.[35] Pitowsky (1989a,b, 1991, 1994) then gave a general

probability $\frac{1}{2}$ to spin in any direction), one can see that the statistics of such sequential measurements on any two spin observables are *independent* of the order of measurement (and the marginal statistics are the same as the statistics for single measurements). We can thus say that any two spin observables are compatible *in the maximally mixed state* (with either sequential observable playing the role of a joint fine-graining of the two single observables). With this understanding, we can reproduce the example also using a single spin system. (It is a general fact that using the perfect correlations of the singlet state, one can always translate between results about the (im)possibility of modelling certain correlations in two systems ('non-locality' results) and results about the (im)possibility of modelling certain correlations in a single system ('non-contextuality') results.)

[35]More generally (and, indeed, in the case of Bell's derivation), one could consider

and systematic treatment of necessary and sufficient conditions for the existence of joint probability measures in terms of such inequalities, further pointing out that these results had already been anticipated more than a century earlier by George Boole (1862).[36]

In this sense, our discussion of the master example above must be a special case of a Bell inequality, and in fact it is a special case of (63). Setting $A = B'$ in (63) we get

$$\begin{aligned} -2 \leq E(AB) - E(AA) + E(A'B) + E(A'A) = \\ E(AB) - 1 + E(A'B) + E(A'A) \leq 2 \quad (65) \end{aligned}$$

for any three two-valued observables. Interpreting any two of them as 'finding or not finding a gem in the box', and substituting the probabilities (55) into (64), we have

$$E(XY) = 2\,\frac{\alpha}{2} - 2\,\frac{1-\alpha}{2} = \alpha - (1-\alpha) = 2\alpha - 1 \tag{66}$$

for any distinct X, Y, and (65) becomes

$$-2 \leq 6\alpha - 3 - 1 \leq 2 \ , \tag{67}$$

that is

$$2 \leq 6\alpha \leq 6 \ , \tag{68}$$

thus $\alpha \in [\frac{1}{3}, 1]$, as above.

'contextual hidden variables' models of the correlations, i. e. allow the use of a different classical probability measure in each experimental context, thus treating for instance the outcomes of a measurement of A as *different events* in the two cases in which A is measured together with B or together with B'. The relation between Bell's and Fine's derivations of the CHSH inequalities can then be seen as follows. Take the case of the singlet state, and assume that the two systems are spatially (or space-like) separated, call them the left-hand and right-hand system. Assume that the probabilities for the outcomes of a spin measurement on the left-hand system depend on the experimental context on the left. But then, given the perfect (anti-)correlations, also the probabilities for outcomes of spin measurements in the same direction on the *right-hand* side will (non-locally!) depend on the experimental context on the left. Thus, a local hidden variables model that reproduces the correlations for the singlet state must be non-contextual, and so by Fine's theorem necessarily obey the CHSH inequality. (We shall return to the idea of contextual classical models of general probabilistic structures in our final Section 6.)

[36]See also Beltrametti and Bugajski (1996) for further discussion of the case in which probabilities fail to be induced by a joint probability measure (which they call the 'Bell phenomenon').

The Kochen–Specker theorem instead takes finite sets of projection-valued observables in a Hilbert space of dimension at least 3, that may pairwise share a projection (partially compatible observables), and considers the question of whether values 1 and 0 may be assigned to the projections in such a way that exactly one projection from each observable is assigned the value 1. (The theorem was first announced in Specker (1960), and its proof was published in Kochen and Specker (1967).)

One already knows that making such assignments to *all* projection-valued observables in a Hilbert space of dimension at least 3 must lead to a contradiction. Indeed, such assignments are simply trivial probability measures over the projections in Hilbert space, and by Gleason's theorem (which we discussed in Section 3) the most general such probability measures are the quantum mechanical states, which in fact always assign non-trivial probabilities to *some* observables. By the compactness theorem of first-order logic, one then knows also that there must be a finite set of observables for which such an assignment leads to a contradiction. But Kochen and Specker offer a constructive proof that such a finite set of observables exists (the original proof involved 117 one-dimensional projections in 3 dimensions[37]).

The case $\alpha = 0$ in our example can now be seen as a 'two-dimensional' analogue of the Kochen–Specker theorem. Indeed, it can be seen as comprised of three interlocking pairs of projections, such that exactly one element in each pair is assigned the value 1, and the other one the value 0.

The analogy goes both ways: any Kochen–Specker construction (of finite sets of orthonormal bases that cannot be assigned values 1 and 0 in such a way that exactly one vector in each basis is assigned 1) is equivalent to the non-existence of trivial probability measures satisfying suitable constraints (in three dimensions, these are $p(x \vee y|z) = 0$ and $p(\neg x|\neg y \wedge \neg z) = 0$ for all orthonormal triples). But the existence of trivial probability measures satisfying such constraints is in fact equivalent to the existence of non-trivial probability measures satisfying the same constraints. Thus, indeed, every Kochen–Specker theorem can be translated into the violation of some Bell–Pitowsky inequality.[38]

[37] More economical proofs are now available even in 3 dimensions, using as little as 31 one-dimensional projections (in an unpublished proof by Conway and Kochen), and in 4 dimensions, using as little as 18 (Cabello, Estebaranz and García Alcaine, 1996). For details and references, see e. g. Bub (1997) and Held (2013).

[38] If I am not mistaken, this is the intuitive way of understanding the 'non-contextuality'

6 Is probability empirical (and quantum)?

The title of this section recalls (tongue-in-cheek) the title of the classic paper by Putnam (1968) in which he notoriously argued that quantum mechanics requires a fundamental revision of logic. Empirical considerations alone presumably cannot decide the question of whether logic is an empirical or an *a priori* discipline (as forcefully pointed out in another classic paper by Dummett (1976)). But if one is already sympathetic to the idea that logic is an empirical discipline, then it does make sense to ask what kind of empirical evidence might suggest adopting this or that logic, and in particular whether the evidence we have for quantum mechanics suggests adopting a non-classical one (e. g. one based on Kochen and Specker's partial Boolean algebras). Essentially, the question boils down to whether quantum logic should be seen as a derivative construct that is definable in terms of and alongside classical logic, or whether classical logic should be seen as an instance of quantum logic restricted to certain special 'well-behaved' cases.[39]

A somewhat similar question might be asked with regard to probability. We have seen in Sections 3–4 that quantum mechanics suggests introducing a notion of probabilistic state generalising that of a probability measure, and the non-embeddability result we have derived in Section 5 states that in general the joint probability distributions defined by a state (in particular a quantum mechanical state) for certain pairs of observables cannot be recovered as marginals of a single classical probability measure. In this final section, we shall discuss whether these results should compel us to see classical probabilities as a special case of generalised probabilities (for the case in which all observables are compatible), or whether generalised probability theory could after all be derivative of classical probability.

There is a sense in which the latter question can indeed be trivially answered in the affirmative, by taking seriously the idea that a general probabilistic state should be seen as a family of classical probability measures, but denying that they in fact overlap. For instance, in our master example above (say, with $\alpha = 0$), this means simply that instead of describing the relevant

inequalities first introduced in Cabello (2008), and shown to describe any Kochen–Specker contradiction in Badziąg *et al.* (2008).

[39] This would be analogous to, say, the application of intuitionistic logic to finitary problems in mathematics, for which also *tertium non datur* becomes an intuitionistically valid principle. For a recent review, emphasising that the answer might depend rather sensitively on the interpretation of quantum mechanics, see Bacciagaluppi (2009).

probabilistic structure using a single state that assigns the probabilities

$$p(e) = p(f) = p(g) = \frac{1}{2} \tag{69}$$

to the outcomes e, f and g (and the appropriate joint probabilities to pairs), we describe it using *three different* classical probability measures, which are to be applied respectively to the experiments in which we measure e and f together, or f and g together, or g and e together, and that assign, respectively, the probabilities

$$p_{ef}(e) = p_{ef}(f) = \frac{1}{2}, \quad p_{fg}(f) = p_{fg}(g) = \frac{1}{2}, \quad p_{ge}(g) = p_{ge}(e) = \frac{1}{2}\ , \tag{70}$$

to the single outcomes (and the appropriate joint probabilities to the three pairs). We see now that these three probability measures can be derived from a single classical probability measure if we assign probabilities also to performing each of the experiments ef, fg and ge. This is just the 'naïve' argument we rehearsed at the beginning of Section 5, but which is now no longer blocked, because we resist identifying the two events e_{ef} and e_{ge} as one (and similarly for f and for g). In this (formal) sense, a 'contextual hidden variables theory' is always possible.[40] (Note, however, that if we then imagine performing the three joint measurements ef, fg and ge in sequence, then the measured value of at least one observable, say e, must be different in the two measurements containing it, in this case ef and ge. Thus we have some mysterious form of 'disturbance through measurement', much like in our simple discussion of spin in Section 1.)

What is crucial here is that instead of insisting that experimental outcomes belonging to the same effect be identified as the same event, we rather insist that experimental outcomes belonging to different experiments are *different events*. This suggestion should not be too hastily dismissed. Identifying experimental outcomes that are equiprobable in all states might after all be thought of only as a convenient book-keeping device of no fundamental importance. (Even in quantum mechanics, as we pointed out in Section 3, the same effect can correspond to different physical transformations of the state in different experiments, so one could very well argue that these be considered different events.)

In quantum mechanics, these questions are played out in the context of the debate on hidden variables theories (see e. g. Shimony 1984). If effects (or at least projections) correspond directly to physical properties of a

[40] For an explicit construction in the case of quantum mechanics, see Gudder (1970).

quantum system, which are then measured in various ways, then projections in common to different observables (resolutions of the identity) should, indeed, be identified. If instead the properties of the system are some 'hidden variables', which in the context of specific experimental arrangements lead to certain experimental outcomes (perhaps with certain probabilities), then projections no longer represent intrinsic properties of the system in general, but only aspects of how systems can be probed in the context of specific experimental situations.

These brief remarks suggest that the question of whether different experimental outcomes ought to be identified should not be decided abstractly, but rather in relation to specific theoretical commitments. We shall not attempt a general discussion of this point, nor even an exhaustive one of hidden variables theories in quantum mechanics. What we shall do instead is illustrate the point in some concrete implementations of our master example (mainly for $\alpha = 0$), which will enable us to see the possibility of underlying mechanisms providing us with a rationale for deciding when different experimental outcomes should be treated as different events.

The original setting of our case $\alpha = 0$ (but without the probabilistic structure) is the tale of the Sage of Nineveh from Specker (1960) (my translation):[41]

> At the Assyrian school for prophets in Arba'ilu, there taught, in the age of king Asarhaddon, a sage from Nineveh. He was an outstanding representative of his discipline (solar and lunar eclipses), who, except for the heavenly bodies, had thoughts almost only for his daughter. His teaching success was modest; the discipline was seen as dry, and did furthermore require previous mathematical knowledge that was rarely available. If in his teaching he thus failed to gain the interest he would have wanted from the students, he received it overabundantly in a different field: no sooner had his daughter reached the marriageable age, than he was flooded with requests for her hand from students and young graduates. And even though he did not imagine wishing to keep her with him forever, yet she was still far too young, and the suitors in

[41] This is the paper that first (informally) introduced the notion of partial Boolean algebras, later developed in detail by Kochen and Specker (1965a,b). The tale can in fact be analysed as an example of a (non-transitive, hence non-quantum) partial Boolean algebra. Specker was a great story-teller, and I personally heard him tell this particular story I think in the spring of 1985.

> no way worthy of her. And so that each should himself be assured of his unworthiness, he promised her hand to the one who could perform a set prophecy task. The suitor was led in front of a table on which stood three boxes in a row, and urged to say which boxes contained a gem and which were empty. Yet, as many as would try it, it appeared impossible to perform the task. After his prophecy, each suitor was in fact urged by the father to open two boxes that he had named as both empty or as both not empty: it always proved to be that one contained a gem and the other did not, and actually the gem lay now in the first, now in the second of the opened boxes. But how should it be possible, out of three boxes, to name no two as empty or as not empty? Thus indeed the daughter would have remained unmarried until her father's death, had she not upon the prophecy of a prophet's son swiftly opened two boxes herself, namely one named as full and one named as empty — which they yet truly turned out to be. Upon the father's weak protest that he wanted to have two different boxes opened, she tried to open also the third box, which however proved to be impossible, upon which the father, grumbling, let the unfalsified prophecy count as successful.

The question we wish to address is whether opening box A in the context of also opening box B is the same event as opening box A in the context of also opening box C. Given our usual intuitions, i. e. background theoretical assumptions, it would seem that we do have to identify the two events. But we can also imagine the situation as follows.

What the father wants to establish is whether any of the suitors are better prophets than himself (only then would he willingly surrender his daughter's hand in marriage). Whenever a suitor is set the task, the father predicts which two boxes will be opened, and places exactly one gem at random in one of the two boxes. (Note that, in this form, the example bears some analogy to Newcomb's paradox!) If we now assume that the father possesses a genuine gift for clairvoyant prophecy, the action of opening boxes A and B, or of opening A and C, has a retrocausal effect on whether the father has placed the gem in either A or B, or has placed it in either A or C.

Now we have an explanation of why to each of the three experimental situations corresponds a *different* classical probability measure, and there appears to be no longer a motivation for describing the situation using a single probabilistic state irrespective of the experimental situation. (Note that should probabilities be defined also for which two of the boxes will be opened, one could again introduce a single classical measure from which the three probability measures arise through conditionalisation.)

We can imagine a different mechanism (and arrive at the same conclusion) by considering another classic illustration of our example, the so-called 'firefly box'.[42] Imagine that the three boxes are in fact chambers at the corners of a single box in the shape of an equilateral triangle, all three chambers being accessible from the centre. Assume that we are in darkness, and hold up a lantern to any one of the sides of the triangle. And assume that what we observe is always that (at random) one chamber on the illuminated side faintly starts to glow. Probabilistically, this is exactly the same example as above. But we can now imagine a different explanation for this phenomenon, as follows.

At the centre of the box sits a firefly, which is attracted to the light of our lantern, and thus enters at random one of the two chambers on the side from which we are approaching. And mistaking our lantern for a potential mating partner, the firefly starts to glow!

We have again a mechanism explaining the statistics of our experiments, and we can give the same classical probabilistic model of the situation as before, i. e. we have three different experiments, each of which is described by a different classical probability measure. And if we so wish, we can again introduce probabilities for our approaching from any particular side.

Quantum mechanics provides us with non-classical, non-contextual probabilistic models of various phenomena, and several impossibility theorems show that there fail to be any classical non-contextual probabilistic models reproducing the quantum mechanical statistics (so-called non-contextual hidden variables theories). In these examples instead we see the analogues of various strategies used in quantum mechanics to introduce *classical but contextual* probabilistic models (so-called contextual hidden variables theories).

Indeed, only few retrocausal models of quantum mechanics have been

[42] It appears that this illustration was devised by Dave Foulis to explain the work of his group to Eugene Wigner, who happened to be visiting. (Thanks to Alex Wilce for relating the anecdote.)

developed in detail, but retrocausality has long been recognised as a possible strategy to deal with the puzzles of quantum mechanics, in particular in the face of the Bell inequalities.[43] The firefly model instead more closely resembles a theory like de Broglie and Bohm's pilot-wave theory, in which experimental outcomes depend on both the initial configuration of the system (e.g. the position of an electron) and the details of the experimental arrangement. This last point can probably best be seen in another slight variant of the example.

Imagine that instead of the firefly we have a small metal ball in the centre of the box, and that each experiment consists of tilting the box towards one of the sides, say AB. The ball rolls towards the side AB and bounces off a metal pin either into chamber A or chamber B, depending on its exact initial location to the left or the right of the symmetry axis perpendicular to the side AB. It is now clear that the *same* initial position of the ball might lead it to fall or not to fall into, say, chamber A, depending on whether the whole box is tilted towards the side AB or the side CA (namely if the ball is on the left of the symmetry axis through AB *as well as* to the left of the symmetry axis through CA). Thus, depending on which way the box is tilted, the ball ending up in A corresponds to a *different* random variable on the probability space of initial positions of the ball. If the initial position of the ball is uniformly distributed in a symmetric neighbourhood of the centre of the triangle, the equal probabilities of the non-classical state are reproduced. But if the initial position is not in such an 'equilibrium' distribution, deviations from the probabilities in the Lemma can occur — so that if one allows also such 'disequilibrium' hidden states, different experimental outcomes are in fact no longer equiprobable in all states.[44]

Our examples can be easily generalised to include e. g. the original Kochen–Specker example — for which we need a spherical firefly box with 117 subchambers, only three of which are made accessible to the firefly every time we approach the box (depending on how exactly we approach it). Or in-

[43] This has been recognised at least since the work of Costa de Beauregard (1977). For a detailed retrocausal model, see Sutherland (2008), and for in-depth discussion of both retrocausality and its possible role in quantum mechanics, see Price (1996).

[44] As remarked already, in a measurement of spin in pilot-wave theory the same initial position of a particle can lead to a spot on the screen corresponding to spin up or spin down depending on the relative orientation of the polarity and the gradient of the magnetic field that deflects the particle (this is a case of 'environmental' contextuality going beyond the 'algebraic' one). For the notion of disequilibrium in pilot-wave theory, see e. g. Valentini (2004) and Towler, Russell and Valentini (2012).

deed to include cases with $\alpha > 0$. For the latter, we need a cubical firefly box, which we approach from any of the six faces (counting opposite faces as equivalent). On each face, the four corners correspond to, say, $e \wedge f$ and $\neg e \wedge \neg f$ across one diagonal, and $e \wedge \neg f$ and $\neg e \wedge f$ across the other, and similarly with f and g, or g and e, on the other faces. The classical cases can be obtained if the firefly just sits somewhere in the box (maybe preferentially along one spatial diagonal — where food might be provided), and starts to glow when it sees the light from our lantern. We then observe the projections of the firefly's position on the face from which we approach. The non-classical cases can be obtained if the firefly moves towards the side from which we are approaching, and through various obstacles is channelled preferentially (although not always) along the planar diagonal corresponding to the opposite outcomes for that face (say $e \wedge \neg f$ and $\neg e \wedge f$). We can thus construct classical but contextual models that violate (our special case of) the Bell inequalities, reproducing the quantum violations, or even the non-quantum violations (reducing to the equilateral triangle in the limit).

In conclusion, while the results of Section 5 show that a generalised probabilistic model as introduced in Section 4 cannot always be embedded in a single classical probability space, the examples in this section indicate that it can always be reproduced using a family of classical probability measures indexed by different experimental contexts, *if* indeed we have a reason to resist the temptation to identify experimental outcomes across different experiments. Identifying experimental outcomes that are equiprobable in all states may be completely natural once a theoretical setting is given; but whether two events are to be judged the same is not a formal question, nor can it be decided purely on operational grounds. Instead it depends on the choice of theoretical setting. In the specific case of quantum probabilities, this question is closely related to the notorious question of the interpretation of quantum mechanics.

Appendix

We give the proofs of the Lemma and the Proposition from Section 5.

Proof of the Lemma:
Note first that if $\alpha \neq 1$, then for any pair (x, y) we have $p(\neg y) \neq 0$ and $p(x) \neq 0$. Indeed, if $p(\neg y) = 0$, then $p(\neg x|\neg y)$ is ill-defined, contrary to

assumption, and if $p(\neg y) \neq 0$ but $p(x) = 0$, then $p(\neg x) = 1$, and $p(\neg x|\neg y) = 1$, also contrary to assumption. Thus we can write

$$\frac{p(x \wedge \neg y)}{p(\neg y)} = p(x|\neg y) = 1 - p(\neg x|\neg y) = 1 - \alpha =$$

$$1 - p(y|x) = p(\neg y|x) = \frac{p(\neg y \wedge x)}{p(x)} \ . \quad (71)$$

Since $\alpha \neq 1$, also the numerators are non-zero, and we have

$$p(x) = p(\neg y) \ . \quad (72)$$

But if

$$p(e) = p(\neg f) \ , \quad p(f) = p(\neg g) \ , \quad p(g) = p(\neg e) \ , \quad (73)$$

it follows that

$$p(e) = p(\neg e) = p(f) = p(\neg f) = p(g) = p(\neg g) = \frac{1}{2} \ . \quad (74)$$

Finally, by (74) and assumption (55), we have

$$p(x \wedge y) = p(\neg x \wedge \neg y) = \frac{\alpha}{2} \ , \quad (75)$$

and

$$p(x \wedge \neg y) = p(\neg x \wedge y) = \frac{1-\alpha}{2} \quad (76)$$

for any (x, y). *QED.*

Proof of the Proposition:
'Only if' implication: Assume p is induced by a joint probability measure (also denoted by p). Then, by repeatedly applying the Lemma to the Boolean algebras generated by $\{e, f\}$, $\{f, g\}$ and $\{g, e\}$, we have:

$$p(e \wedge f \wedge g) + p(e \wedge f \wedge \neg g) = \frac{\alpha}{2} \quad (77)$$

$$p(\neg e \wedge \neg f \wedge g) + p(\neg e \wedge \neg f \wedge \neg g) = \frac{\alpha}{2} \ , \quad (78)$$

$$p(e \wedge f \wedge g) + p(\neg e \wedge f \wedge g) = \frac{\alpha}{2} \quad (79)$$

$$p(e \wedge \neg f \wedge \neg g) + p(\neg e \wedge \neg f \wedge \neg g) = \frac{\alpha}{2} \ , \quad (80)$$

and

$$p(e \wedge f \wedge g) + p(e \wedge \neg f \wedge g) = \frac{\alpha}{2} \tag{81}$$

$$p(\neg e \wedge f \wedge \neg g) + p(\neg e \wedge \neg f \wedge \neg g) = \frac{\alpha}{2} \ , \tag{82}$$

respectively. From (77), (79) and (81),

$$p(e \wedge f \wedge \neg g) = p(\neg e \wedge f \wedge g) = p(e \wedge \neg f \wedge g) = \\ \frac{\alpha}{2} - p(e \wedge f \wedge g) \leq \frac{\alpha}{2} \ , \tag{83}$$

and from (78) (80) and (82),

$$p(\neg e \wedge \neg f \wedge g) = p(e \wedge \neg f \wedge \neg g) = p(\neg e \wedge f \wedge \neg g) = \\ \frac{\alpha}{2} - p(\neg e \wedge \neg f \wedge \neg g) \leq \frac{\alpha}{2} \ . \tag{84}$$

But now, from (77), (78), (83) and (84),

$$1 = p(e \wedge f \wedge g) + p(e \wedge f \wedge \neg g) + p(e \wedge \neg f \wedge g) + p(e \wedge \neg f \wedge \neg g) + \\ p(\neg e \wedge f \wedge g) + p(\neg e \wedge f \wedge \neg g) + p(\neg e \wedge \neg f \wedge g) + p(\neg e \wedge \neg f \wedge \neg g) \leq 3\alpha \ , \tag{85}$$

Thus $\alpha \geq \frac{1}{3}$. *QED.*

'If' implication: let $\frac{1}{3} \leq \alpha < 1$ (note that this construction works also with $\alpha = 1$), and let

$$a := \frac{3\alpha - 1}{4} \qquad \text{and} \qquad b := \frac{1 - \alpha}{4} \ . \tag{86}$$

We have $a, b \in [0, 1]$. Set

$$p(e \wedge f \wedge g) = p(\neg e \wedge \neg f \wedge \neg g) = a \tag{87}$$

and

$$p(e \wedge f \wedge \neg g) = p(e \wedge \neg f \wedge g) = p(e \wedge \neg f \wedge \neg g) = \\ p(\neg e \wedge f \wedge g) = p(\neg e \wedge f \wedge \neg g) = p(\neg e \wedge \neg f \wedge g) = b \ . \tag{88}$$

Then the probability measure p induces a state satisfying both (74) (because $a + 3b = \frac{1}{2}$) and (77)–(82) (because $a + b = \frac{\alpha}{2}$). Thus the state satisfies (55). *QED.*

Note that this is not the unique probability measure inducing the given state. As in the case $\alpha = 1$, one need not have $p(e \wedge f \wedge g) = p(\neg e \wedge \neg f \wedge \neg g)$. Indeed, for any $\varepsilon \in [-\min(\frac{3\alpha-1}{4}, \frac{1-\alpha}{4}), \min(\frac{3\alpha-1}{4}, \frac{1-\alpha}{4})]$, one can set

$$p(e \wedge f \wedge g) = \frac{3\alpha - 1}{4} + \varepsilon \tag{89}$$

and

$$p(\neg e \wedge \neg f \wedge \neg g) = \frac{3\alpha - 1}{4} - \varepsilon \ , \tag{90}$$

and extend via (83)–(84) to a probability measure inducing the same state.

We see that we can construct a classical model of the given probabilistic state if and only if $\frac{1}{3} \leq \alpha \leq 1$. In the case of the non-unique states with $\alpha = 1$, this is given uniquely by (57)–(58), and in the case of the unique states with $\alpha < 1$, it is given non-uniquely by (89)–(90) and (83)–(84).

Acknowledgements

I would like to thank Alan Hájek and Crish Hitchcock for their invitation to contribute a version of this article to the *Oxford Handbook of Probability and Philosophy* and the opportunity to write on this topic, and for very helpful feedback on a previous draft. I am further grateful to Alex Wilce for some extremely useful discussions and hard-to-find references, to Jennifer Bailey for some stylistic advice, and to the audience of the Philosophy of Physics seminar at the University of Aberdeen, who heard preliminary versions of this material.

References

Albert, D. (1992), *Quantum Mechanics and Experience* (Cambridge, Mass.: Harvard University Press).

Bacciagaluppi, G. (2009), 'Is Logic Empirical?', in D. Gabbay, D. Lehmann and K. Engesser (eds.), *Handbook of Quantum Logic* (Amsterdam: Elsevier), pp. 49–78, `http://philsci-archive.pitt.edu/3380/`.

Bacciagaluppi, G. (2013), 'Measurement and Classical Regime in Quantum Mechanics', in R. Batterman (ed.), *Oxford Handbook of Philosophy of Physics* (Oxford: OUP), pp. 416–459 `http://philsci-archive.pitt.edu/8770/`.

Bacciagaluppi, G., and Wilce, A. (in preparation), 'Specker's Principle and (Ortho)coherence'.

Badziąg, P., Bengtsson, I., Cabello, A., and Pitowsky, I. (2008), 'Universality of State-independent Violation of Correlation Inequalities for Noncontextual Theories', *Physical Review Letters* **101**, 210401, http://arxiv.org/abs/0809.0430.

Bell, J. S. (1971), 'Introduction to the Hidden-Variable Question', *Proceedings of the International School of Physics 'Enrico Fermi', Course IL, Foundations of Quantum Mechanics* (New York: Academic), pp. 171–181. Reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: CUP), pp. 29–39.

Beltrametti, E. G., and Bugajski, S. (1996), 'The Bell Phenomenon in Classical Frameworks', *Journal of Physics A: Mathematical and General* **29**(2), 247–261.

Beltrametti, E. G., and Cassinelli, G. (1981), *The Logic of Quantum Mechanics* (Reading, Mass.: Addison-Wesley).

Birkhoff, G., and von Neumann, J. (1936), 'The Logic of Quantum Mechanics', *Annals of Mathematics* **37**, 823–843. Reprinted in Hooker (1975), pp. 1–26.

Boole, G. (1862), 'On the Theory of Probabilities', *Philosophical Transactions of the Royal Society of London* **152**, 225–252.

Bub, J. (1997), *Interpreting the Quantum World* (Cambridge: CUP).

Busch, P. (2003), 'Quantum States and Generalized Observables: A Simple Proof of Gleason's Theorem', *Physical Review Letters* **91**(12), 120403, http://arxiv.org/abs/quant-ph/9909073.

Busch, P., Lahti, P. J. and Mittelstaedt, P. (1991), *The Quantum Theory of Measurement* (Berlin: Springer).

Busch, P., Grabowski, M., and Lahti, P. J. (1995), *Operational Quantum Physics* (Berlin: Springer).

Cabello, A. (2008), 'Experimentally Testable State-Independent Quantum Contextuality', *Physical Review Letters* **101**, 210401, http://arxiv.org/abs/0808.2456.

Cabello, A. (2012), 'Specker's Fundamental Principle of Quantum Mechanics', http://arxiv.org/abs/1212.1756.

Cabello, A. (2013), 'Simple Explanation of the Quantum Violation of a Fundamental Inequality', *Physical Review Letters* **110**, 060402, http://arxiv.org/abs/1210.2988.

Cabello, A., Estebaranz, J. M., and García Alcaine, G. (1996), 'Bell–Kochen–Specker Theorem: A Proof with 18 Vectors', *Physics Letters* **A 212**(4), 183–187, http://arxiv.org/abs/quant-ph/9706009.

Cabello, A., Severini, S., and Winter, A. (2010), '(Non-)Contextuality of Physical Theories as an Axiom', `http://arxiv.org/abs/1010.2163`.

Cattaneo, G., Marsico, T., Nisticò, G., and Bacciagaluppi, G. (1997) 'A Concrete Procedure for Obtaining Sharp Reconstructions of Unsharp Observables in Finite-Dimensional Quantum Mechanics', *Foundations of Physics* **27**, 1323–1343.

Cerf, N. J., Gisin, N., Massar, S., and Popescu, S. (2005), 'Simulating Maximal Quantum Entanglement Without Communication', *Physical Review Letters* **94**(22), 220403.

Chiribella, G., D'Ariano, G. M., and Perinotti, P. (2011), 'Informational Derivation of Quantum Theory', *Physical Review* **A 84**, 012311, `http://arxiv.org/abs/1011.6451`.

Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. (1969), 'Proposed Experiment to Test Local Hidden-Variable Theories' *Physical Review Letters* **23**(15), 880–884.

Coecke, B., Moore, D. and Wilce, A. (eds.) (2000), *Current Research in Operational Quantum Logic* (Dordrecht: Kluwer).

Coecke, B., Moore, D. and Wilce, A. (2001), 'Operational Quantum Logic: An Overview', `http://arxiv.org/abs/quant-ph/0008019`.

Comte, C. (1996), 'Symmetry, Relativity and Quantum Mechanics', *Il Nuovo Cimento* **B 111**(8), 937–956.

Costa de Beauregard, O. (1977), 'Time Symmetry and the Einstein Paradox', *Il Nuovo Cimento* **B 42**(1), 41–64.

Dakić, B., and Brukner, Č. (2011), 'Quantum Theory and Beyond: Is Entanglement Special?', in H. Halvorson (ed.), *Deep Beauty: Understanding the Quantum World through Mathematical Innovation* (Cambridge: CUP), pp. 365–392, `http://arxiv.org/abs/0911.0695`.

Darrigol, O. (forthcoming), *Physics and Necessity: Rationalist Pursuits from the Cartesian Past to the Quantum Present* (Oxford: OUP).

Dummett, M. (1976), 'Is logic empirical?', in H. D. Lewis (ed.), *Contemporary British Philosophy*, 4th series (London: Allen and Unwin), pp. 45–68. Reprinted in M. Dummett, *Truth and other Enigmas* (London: Duckworth, 1978), pp. 269–289.

Fine, A. (1982), 'Hidden Variables, Joint Probability and Bell Inequalities', *Physical Review Letters* **48**, 291–295.

Foulis, D. J., and Bennett, M. K. (1994), 'Effect Algebras and Unsharp Quantum Logics', *Foundations of Physics* **24**, 1331–1352.

Foulis, D. J., and Randall, C. H. (1972), 'Operational Statistics. I. Basic Concepts', *Journal of Mathematical Physics* **13**, 1667–1675.

Foulis, D. J., and Randall, C. H. (1974), 'Empirical Logic and Quantum Mechanics', *Synthese* **29**, 81–111.

Foulis, D. J., and Randall, C. H. (1979), 'Tensor Products of Quantum Logics do not Exist', *Notices of the American Mathematical Society* **26**, 557.

Foulis, D. J., and Randall, C. H. (1981), 'Empirical Logic and Tensor Products', in H. Neumann (ed.), *Interpretations and Foundations of Quantum Theory* (Mannheim: Bibliographisches Institut), pp. 9–20.

Ghirardi, G.C. (1997), *Un'occhiata alle carte di Dio* (Milano: Il Saggiatore). Transl. by G. Malsbary as *Sneaking a Look at God's Cards* (Princeton: Princeton University Press, 2005).

Giuntini, R., and Greuling, H. (1989), 'Toward a Formal Language for Unsharp Properties', *Foundations of Physics* **19**, 931–945.

Gleason, A. M. (1957), 'Measures on the Closed Subspaces of a Hilbert Space', *Journal of Mathematics and Mechanics* **6**, 885–893. Reprinted in Hooker (1975), pp. 123–133.

Goyal, P. (2008a), 'An Information-Geometric Reconstruction of Quantum Theory, I: The Abstract Quantum Formalism', *Physical Review* **A 78**, 052120, `http://arxiv.org/abs/0805.2761`.

Goyal, P. (2008b), 'An Information-Geometric Reconstruction of Quantum Theory, II: The Correspondence Rules of Quantum Theory', `http://arxiv.org/abs/0805.2765`.

Gudder, S. P. (1970), 'On Hidden-Variable Theories', *Journal of Mathematical Physics* **11**, 431–436.

Hardegree, G. M., and Frazer, P. J. (1981), 'Charting the Labyrinth of Quantum Logics: A Progress Report', in E. G. Beltrametti and B. C. van Fraassen (eds.), *Current Issues in Quantum Logic* (New York: Plenum), pp. 53–76.

Hardy, L. (2002), 'Why Quantum Theory?', in T. Placek and J. Butterfield (eds.), *Non-locality and Modality*, NATO Science Series II, Vol. 64 (Dordrecht: Kluwer), pp. 61–73, `http://arxiv.org/abs/quant-ph/0111068`.

Held, C. (2013), 'The Kochen–Specker Theorem', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), http://plato.stanford.edu/archives/spr2013/entries/kochen-specker/.

Hooker, C. A. (1975), *The Logico-Algebraic Approach to Quantum Mechanics*, Vol. 1 (Dordrecht: Reidel).

Hughes, R. I. G. (1985), 'Review of S. Kochen and E. P. Specker, "Logical Structures Arising in Quantum Theory", etc.', *Journal of Symbolic Logic* **50**(2), 558–566.

Kochen, S., and Specker, E. P. (1965a), 'Logical Structures Arising in Quantum Theory', in L. Addison, L. Henkin and A. Tarski (eds.), *The Theory of Models* (Amsterdam: North-Holland), pp. 177–189. Reprinted in Hooker (1975), pp. 263–276.

Kochen, S., and Specker, E. P. (1965b), 'The Calculus of Partial Propositional Functions', in Y. Bar-Hillel (ed.), *Logic, Methodology, and Philosophy of Science* (Amsterdam: North-Holland), pp. 45–57. Reprinted in Hooker (1975), pp. 277–292.

Kochen, S., and Specker, E. P. (1967), 'The Problem of Hidden Variables in Quantum Mechanics', *Journal of Mathematics and Mechanics* **17**, 59–88. Reprinted in Hooker (1975), pp. 293–328.

Kôpka, F. (1992), 'D-posets of Fuzzy Sets', *Tatra Mountains Mathematical Publications* **1**, 83–87.

Ludwig, G. (1954), *Die Grundlagen der Quantenmechanik* (Berlin: Springer). Transl. by C. A. Hein as *Foundations of Quantum Mechanics* (Berlin: Springer, 1983).

Ludwig, G. (1985), *An Axiomatic Basis of Quantum Mechanics. 1. Derivation of Hilbert Space* (Berlin: Springer).

Mackey, G. W. (1957), 'Quantum Mechanics and Hilbert Space', *American Mathematical Monthly* **64**(2), 45–57.

Mackey, G. W. (1963), *The Mathematical Foundations of Quantum Mechanics* (New York: W. A. Benjamin).

Marlow, A. R. (ed.) (1978), *Mathematical Foundations of Quantum Physics* (New York: Academic).

Masanes, L., and Müller, M. (2011), 'A Derivation of Quantum Theory from Physical Requirements', *New Journal of Physics* **13**, 063001, http://arxiv.org/abs/1004.1483.

Pitowsky, I. (1989a), 'From George Boole to John Bell: The Origins of Bell's Inequality', in M. Kafatos (ed.), *Bell's Theorem, Quantum Theory and the Conceptions of the Universe* (Dordrecht: Kluwer), pp. 37–49.

Pitowsky, I. (1989b), *Quantum Probability, Quantum Logic*, Lecture Notes in Physics, Vol. 321 (Berlin: Springer).

Pitowsky, I. (1991), 'Correlation Polytopes, their Geometry and Complexity', *Mathematical Programming* **A 50**, 395–414.

Pitowsky, I. (1994), 'George Boole's "Conditions of Possible Experience" and the Quantum Puzzle', *The British Journal for the Philosophy of Science* **45**(1), 95–125.

Price, H. (1996), *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time* (Oxford: OUP).

Putnam, H. (1968), 'Is Logic Empirical?' in R. Cohen and M. Wartofsky (eds.), *Boston Studies in the Philosophy of Science*, Vol. 5 (Dordrecht: Reidel), pp. 216–241. Reprinted as 'The Logic of Quantum Mechanics' in H. Putnam, *Mathematics, Matter, and Method. Philosophical Papers*, Vol. 1 (Cambridge: CUP, 1975), pp. 174–197.

Randall, C. H., and Foulis, D. J. (1970), 'An Approach to Empirical Logic', *American Mathematical Monthly* **77**, 363–374.

Randall, C. H., and Foulis, D. J. (1973), 'Operational Statistics. II. Manuals of Operations and their Logics', *Journal of Mathematical Physics* **14**, 1472–1480.

Shimony, A. (1984), 'Contextual Hidden Variables Theories and Bell's Inequalities', *The British Journal for the Philosophy of Science* **35**(1), 25–45.

Specker, E. P. (1960), 'Die Logik nicht gleichzeitig entscheidbarer Aussagen', *Dialectica* **14**, 239–246. Transl. by A. Stairs as 'The Logic of Propositions which are not Simultaneously Decidable' in Hooker (1975), pp 135–140, and by M. P. Seevinck as 'The Logic of Non-simultaneously Decidable Propositions', `http://arxiv.org/abs/1103.4537`.

Sutherland, R. I. (2008), 'Causally Symmetric Bohm Model', *Studies in History and Philosophy of Modern Physics* **39**(4), 782–805, `http://arxiv.org/abs/quant-ph/0601095`.

Towler, M. D., Russell, N. J., and Valentini, A. (2012), 'Timescales for Dynamical Relaxation to the Born Rule', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* **468**(2140), 990-1013, `http://arxiv.org/abs/1103.1589`.

Valentini, A. (2004), 'Universal Signature of Non-quantum Systems', *Physics Let-*

ters **A 332**(3), 187–193, http://arxiv.org/abs/quant-ph/0309107.

Varadarajan, V. S. (1968), *The Geometry of Quantum Theory*, Vols. I–II (New York: Van Nostrand).

Wallace, D. (2008), 'Philosophy of Quantum Mechanics', in D. Rickles (ed.), *The Ashgate Companion to Contemporary Philosophy of Physics* (Aldershot: Ashgate), pp. 16–98. (Preliminary version available as 'The Quantum Measurement Problem: State of Play (December 2007)', http://philsci-archive.pitt.edu/3420/.)

Wilce, A. (2000), 'Test Spaces and Orthoalgebras', in Coecke, Moore and Wilce (2000), pp. 81–114.

Wilce, A. (2012), 'Quantum Logic and Probability Theory', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition), http://plato.stanford.edu/archives/fall2012/entries/qt-quantlog/.