Chapter 6
Classical models of quantum mechanics

This chapter gives an introduction to a chain of results attempting to exclude deeper layers underneath quantum mechanics that restore some form of classical physics:

> '[Such results] more or less illustrate the ways along which some opponents might hope to escape Bohr's reasonings and von Neumann's proof and the places where they are dangerously near breaking their necks.' (Groenewold, 1946, p. 454)

In so far as they are mathematically precise, such no-go results have their roots in von Neumann's 1932 book, which gave rise to two traditions that were often in polemical opposition to each other. Mathematically minded authors typically admired von Neumann's exclusion of hidden variables, yet tried to strengthen his theorem by weakening its assumptions; this sparked, for example, *Gleason's Theorem* (1957) as well as the *Kochen–Specker Theorem* (1967). Certain physicists (led by Bell), on the other hand, tried to circumvent (and later even ridicule) von Neumann's work. A high point of this tradition was *Bell's Theorem* from 1964, which was informed not only by von Neumann, but even more so by the famous Einstein–Podolsky–Rosen (EPR) paper from 1935, as well as by Bohm's deterministic pilot wave reformulation of quantum mechanics (1952). However, at the end of the day these traditions turned out to be not really divergent after all: Bell not only independently (and earlier) obtained a version of the Kochen–Specker Theorem, but, more importantly, his results from 1964 turn out to be very closely related to the culmination of the first tradition in the form of the so-called *Free Will Theorem* (FWT), which was published by Conway and Kochen during 2006–2008. Indeed, although its validity is uncontroversial, this theorem has been criticized on the following grounds:

1. Lack of novelty compared with the famous paper by Bell (1964), whose assumptions and conclusions are at least quite similar to those of the FWT (although the underlying proofs are mathematically quite distinct from those in the FWT).
2. Lack of novelty even within its own terms: versions of the FWT had actually been around for decades under less illustrious titles and authorships, e.g. Heywood & Redhead (1983), Stairs (1983), Brown & Svetlichny (1990), and Clifton (1993).
3. Circularity, in that indeterminism is presupposed (namely in the assumption that 'experimenters have a certain freedom') instead of derived.

© The Author(s) 2017

K. Landsman, *Foundations of Quantum Theory*,
Fundamental Theories of Physics 188, DOI 10.1007/978-3-319-51777-3_6

One aim of this chapter is to clarify these matters, with the following conclusions:

1. The difference between earlier literature in the same direction and the FWT is largely one of emphasis, namely on free will (!), exemplifying a recent trend (also found elsewhere) in emphasizing free choice of the settings of experiments. Unfortunately, like Bell, Conway and Kochen even mathematically use an informal way of talking about free settings, not to speak of the complete absence of any serious philosophical analysis of free will among all three authors (for which perhaps Bell, but certainly not Conway and Kochen may be excused).
2. Granting the informal characterization of free settings, both Bell's (1964) Theorem and the FWT establish a contradiction between quantum mechanics, determinism, and locality (in the sense of Bell, which in the presence of determinism reduces to a no-signaling condition called parameter independence).
3. The technical difference between Bell's Theorem and the FWT lies in four facts:

   a. Bell's arguments rely on probability theory (whereas the FWT does not).
   b. The (optical) corner of quantum mechanics used in Bell's Theorem may be replaced by the corresponding experimental results, whereas the FWT uses uncontroversial yet untested predictions about massive spin-1 particles.
   c. The FWT must assume perfect (EPR) correlations, which are difficult to realize and hence are avoided by later versions of Bell's Theorem (i.e. through the CHSH inequalities rather than the original Bell inequalities).
   d. Like EPR, Bell and his followers focused on locality right from the beginning, and hence in Bell (1964) the inference is from locality to determinism. Conway and Kochen, on the other hand, resolve the contradiction their FWT established by inferring randomness of outcomes from freedom of settings.

We start with a very simple treatment of both von Neumann's argument against linear hidden variables and Kochen & Specker's refinement of it, in which von Neumann's controversial linearity assumption is decisively weakened so as to only apply to *commuting* operators; the Kochen–Specker Theorem excludes what are called *non-contextual quasi-linear hidden variables*. We then present what we see as a more transparent version of the FWT, whose key ingredient of replacing the non-contextuality assumption in the Kochen–Specker Theorem by a locality condition is preserved, but where this time the setting is completely deterministic. Freedom of choice then arises as a very natural independence assumption, and any threat of circularity is avoided: the conclusion is simply a contradiction between determinism, freedom of choice (i.e. of apparatus settings), locality, and quantum mechanics. Moreover, as we argue in §6.3, the philosophically precise concept of free will used in the assumptions of the FWT is what Lewis coined 'local miracle compatibilism'.

Following an interlude on the GHZ Theorem, which seamlessly fits into the given framework, we then turn to Bell's Theorems, which we compare with the FWT.

Finally, we give our own rigorous version of an argument first proposed by Colbeck and Renner to the effect that, under suitable freeness of choice and no-signaling conditions (similar to those in Bell's Theorem and the FWT), as long as they are compatible with quantum mechanics, hidden variables are at best irrelevant. In fact, this can only be proved under much stronger assumptions, obscuring the claim.

6.1 From von Neumann to Kochen–Specker

Von Neumann's Theorem 6.2 below was the first technical result excluding some class of hidden variables underneath quantum mechanics, namely (in current parlance) *linear non-contextual hidden variables*. This terminology requires some explanation. First, theorems of this kind apparently accept the mathematical structure of the observables prescribed by the usual formalism of quantum theory, i.e., observables are identified with elements of the self-adjoint part

$$H_n(\mathbb{C}) \equiv M_n(\mathbb{C})_{\mathrm{sa}} = \{a \in M_n(\mathbb{C}) \mid a^* = a\} \tag{6.1}$$

of the algebra $M_n(\mathbb{C})$ of $n \times n$ matrices (this simple case suffices to make all points of conceptual interest). Short of introducing "hidden" *observables*, hidden variable theories propose the existence of hidden *states*, which either replace or supplement the usual quantum states (which in the case at hand would be density operators). Mimicking classical (statistical) physics, such states are interpreted as probability measures on some phase space X, whose points $x \in X$ assign sharp values to quantum-mechanical observables. Naively, this is done through associated functions

$$V_x : H_n(\mathbb{C}) \to \mathbb{R}, \tag{6.2}$$

but in fact this choice already commits us to the first of two possibilities, which we pragmatically present as theories predicting measurement outcomes:

- In ***non-contextual*** deterministic theories of measurement, the outcome solely depends on the observable a that is being measured and on the (possibly 'hidden') state of the system. Theorem 6.2 below, then, rules out such theories in which values are sharp (i.e., dispersion-free), and V_x in (6.2) is *linear*. The Kochen–Specker Theorem subsequently proves the same impossibility under a weaker (and physically more reasonable) assumption called *quasi-linearity*.
- ***Contextual*** deterministic theories of measurement, on the other hand, allow the outcome of some measurement of a to depend on the ***measurement context*** (as well as on the state), which in this case is understood as the choice of possible other (compatible) observables b measured together with a (i.e., $ab = ba$). This seems a reasonable assumption, well within the spirit of quantum mechanics, though perhaps not so in the extreme form later held by Heisenberg, according to which measurement outcomes (or even "reality") are "created" by the measurement. Under a weakened non-contextuality assumption, Bell's Theorem (cf. §6.5) and the Free Will Theorem (§6.2) rule out such theories, too.

Definition 6.1. *A* **non-contextual hidden variable** *is a map $V : H_n(\mathbb{C}) \to \mathbb{R}$ that for each $a \in H_n(\mathbb{C})$, and in terms of the $n \times n$ unit matrix 1_n, satisfies*

$$V(a^2) = V(a)^2; \tag{6.3}$$

$$V(1_n) = 1. \tag{6.4}$$

That is, V is **dispersion-free** *as well as* **normalized***, respectively.*

Theorem 6.2. *For $n \geq 2$, non-zero linear dispersion-free maps $V : H_n(\mathbb{C}) \to \mathbb{R}$ do not exist. In particular, linear non-contextual hidden variables do not exist.*

Proof. Such maps extend to complex-linear dispersion-free maps $V : M_n(\mathbb{C}) \to \mathbb{C}$ by complex linearity, so that theorem is equivalent to Proposition 2.10. □

As von Neumann perfectly well understood himself, his seemingly natural linearity assumption (given the mathematical structure of quantum mechanics unearthed by none other than he!) is unwarranted physically (and even mathematically, since eigenvalues and eigenstates, which should be the hallmark of dispersion-free states, are by no means linear in the underlying operator). This suggests the following:

Definition 6.3. *A map $V : H_n(\mathbb{C}) \to \mathbb{R}$ is called* **quasi-linear** *if for all $s,t \in \mathbb{R}$ and all $a,b \in H_n(\mathbb{C})$* that commute *(i.e., $ab = ba$) one has*

$$V(sa+tb) = sV(a)+tV(b). \tag{6.5}$$

As in the linear case, such a map uniquely extends to a map $V : M_n(\mathbb{C}) \to \mathbb{C}$ that is precisely a quasi-state in the sense of Definition 2.26. The following lemma will be useful, also showing that the above objections to linearity have been met.

Lemma 6.4. *Let $V : H_n(\mathbb{C}) \to \mathbb{R}$ be a quasi-linear non-contextual hidden variable.*

1. *For each $a \in H_n(\mathbb{C})$, the number $\lambda = V(a)$ is an eigenvalue of a.*
2. *If $(a_1,\ldots,a_k)$ pairwise commute, and $b = f(a_1,\ldots,a_k)$ for some polynomial f, then $V(b) = f(V(a_1),\ldots,V(a_k))$.*

More generally, it follows from Theorem C.24 that if H is a Hilbert space and $V : B(H)_{\mathrm{sa}} \to \mathbb{R}$ is a quasi-linear non-contextual hidden variable (or, equivalently, its complexification $V_{\mathbb{C}} : B(H) \to \mathbb{C}$ is a dispersion-free quasi-state), then $V(a) \in \sigma(a)$ (provided $a^* = a$). This implies the above lemma, but we also provide a direct proof.

Proof. For any $b \in H_n(\mathbb{C})$ with $ab = ba$, eq. (6.3) and quasi-linearity imply that

$$V(ab) = V(a)V(b); \tag{6.6}$$

just evaluate $V((a \pm b)^2) = (V(a) \pm V(b))^2$. Taking $b = a^2$ etc. and also invoking (6.4) then yields $V(p(a)) = p(V(a))$ for any polynomial in a. If λ_i are the eigenvalues of a, its characteristic polynomial $p(a) = \prod_{i=1}^n (a-\lambda_i)$ satisfies $p(a) = 0$, so that $V(p(a)) = 0$ and hence $p(V(a)) = 0$, or $\prod_{i=1}^n (\lambda - \lambda_i) = 0$. This implies that $\lambda = \lambda_i$ for some i. The second claim is proved in a similar way. □

Theorem 6.5. *For $n \geq 3$, quasi-linear non-contextual hidden variables do not exist.*

This is the ***Kochen–Specker Theorem***. It follows from Gleason's Theorem 2.28 and von Neumann's Theorem 6.2, since according to Corollary 2.29 to the former, quasi-states on $M_n(\mathbb{C})$ are actually states (in other words, quasi-linear non-contextual hidden variables are linear). However, Kochen and Specker also gave a direct proof of their theorem, subsequently somewhat simplified along the following lines.

Proof. We prove the claim for $n = 3$, which (by restricting V to any self-adjoint subalgebra of $M_n(\mathbb{C})$ isomorphic to $H_3(\mathbb{C})$) implies the result for all $n > 3$ also. To prove Theorem 6.5 for $n = 3$, we interpret $H_3(\mathbb{C})$ as the algebra of observables of a spin-1 particle and introduce the well-known angular momentum matrices

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \; J_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \; J_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6.7}$$

In what follows, we will heavily use the squares

$$J_1^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, J_2^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, J_3^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.8}$$

each of which has eigenvalues 0 and 1. The J_i^2 commute by inspection, and satisfy

$$J_1^2 + J_2^2 + J_3^2 = 2 \cdot 1_3. \tag{6.9}$$

The (matrix-valued) angular momentum vector is given by

$$\mathbf{J} = J_1 \mathbf{e}_1 + J_2 \mathbf{e}_2 + J_3 \mathbf{e}_3, \tag{6.10}$$

where $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is the standard basis of $\mathbb{R}^3$ (seen as a vector space with the usual inner product $\langle \cdot, \cdot \rangle$), i.e., $\mathbf{e}_1 = (1,0,0)$, etc., and the angular momentum $J_{\mathbf{u}}$ along an arbitrary unit vector $\mathbf{u} = \sum_i u_i \mathbf{e}_i$ in $\mathbb{R}^3$ is given by

$$J_{\mathbf{u}} = \langle \mathbf{J}, \mathbf{u} \rangle = \sum_{i=1}^{3} J_i u_i. \tag{6.11}$$

This brings us to the crucial point: a map $V : H_3(\mathbb{C}) \to \mathbb{R}$ induces a map $\tilde{V} : S^2 \to \mathbb{R}$ on the set S^2 of all unit vectors $\mathbf{u}$ in $\mathbb{R}^3$, via

$$\tilde{V}(\mathbf{u}) = V(J_{\mathbf{u}}^2). \tag{6.12}$$

As usual, a *basis* of $\mathbb{R}^3$, denoted by $a = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, is always assumed *orthonormal*.

Lemma 6.6. *Let $V : H_3(\mathbb{C}) \to \mathbb{R}$ be a non-contextual quasi-linear hidden variable, with associated map $\tilde{V} : S^2 \to \{0,1\}$ given by* (6.12)*. Then:*

1. *$\tilde{V}(-\mathbf{u}) = \tilde{V}(\mathbf{u})$ for each $\mathbf{u} \in S^2$ (so that $\tilde{V}$ is defined on the real projective plane);*
2. *If $a = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ is a basis, then the triple $\tilde{V}(a) \equiv (\tilde{V}(\mathbf{u}_1), \tilde{V}(\mathbf{u}_2), \tilde{V}(\mathbf{u}_3))$ must contain a single 0 and two 1's, i.e., $\tilde{V}(a)$ must be one of the triples*

$$\begin{aligned} \lambda^{(1)} &= (0,1,1); \\ \lambda^{(2)} &= (1,0,1); \\ \lambda^{(3)} &= (1,1,0). \end{aligned} \tag{6.13}$$

In Gleason-like language, $\tilde{V}$ is a $\underline{2}$-valued frame function of weight $w(\tilde{V}) = 2$.

Proof. If $a = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ is a basis, then $J_{\mathbf{u}_i} = uJ_iu^*$ for $i = 1,2,3$, where u is the 3×3 matrix with entries $u_{ij} = \langle \mathbf{u}_i, \mathbf{e}_j \rangle$. Since u is unitary, the matrices $J_{\mathbf{u}_i}$ and their squares have the same eigenvalues and satisfy the same relations as the J_i and their squares. Thus the eigenvalues of $J^2_{\mathbf{u}_i}$ are 0 and 1, for fixed a the squares $J^2_{\mathbf{u}_i}$ mutually commute, and they satisfy the sum rule (6.9), i.e., $J^2_{\mathbf{u}_1} + J^2_{\mathbf{u}_2} + J^2_{\mathbf{u}_3} = 2 \cdot 1_3$, so $\tilde{V}(\mathbf{u}_1) + \tilde{V}(\mathbf{u}_2) + \tilde{V}(\mathbf{u}_3) = 2$. The claim then follows from Definition 6.3 and Lemma 6.4. □

Now define a ***coloring*** of $\mathbb{R}^3$ as any map $\tilde{V} : S^2 \to \{0,1\}$ satisfying the two properties in Lemma (6.6). The proof of Theorem 6.5 then reduces to the following lemma.

Lemma 6.7. *There exists no coloring of* $\mathbb{R}^3$.

Proof. Take the following unit vectors (some identical), grouped into 11 bases (for simplicity we use unnormalized vectors, e.g., $(1,0,1)$ stands for $(1/\sqrt{2}, 0, 1/\sqrt{2})$):

basis	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$
a_1	$(0,0,1)$	$(1,0,0)$	$(0,1,0)$
a_2	$(1,0,1)$	$(-1,0,1)$	$(0,1,0)$
a_3	$(0,1,1)$	$(0,-1,1)$	$(1,0,0)$
a_4	$(1,-1,2)$	$(-1,1,2)$	$(1,1,0)$
a_5	$(1,0,2)$	$(-2,0,1)$	$(0,1,0)$
a_6	$(2,1,1)$	$(0,-1,1)$	$(-2,1,1)$
a_7	$(2,0,1)$	$(0,1,0)$	$(-1,0,2)$
a_8	$(1,1,2)$	$(1,-1,0)$	$(-1,-1,2)$
a_9	$(0,1,2)$	$(1,0,0)$	$(0,-2,1)$
a_{10}	$(1,2,1)$	$(-1,0,1)$	$(1,-2,1)$
a_{11}	$(1,0,0)$	$(0,2,1)$	$(0,-1,2)$.

We will show that one cannot even color this particular finite set of vectors (let alone all unit vectors in $\mathbb{R}^3$). We denote a vector $\mathbf{u}_i$ in a basis a_μ by

$$\mathbf{u}_i^{(\mu)}, i = 1,2,3, \mu = 1,\ldots,11,$$

and write e.g. $\tilde{V}(a_\mu) = (0,1,1)$ for the three conditions

$$\tilde{V}(\mathbf{u}_1^{(\mu)}) = 0,\ \tilde{V}(\mathbf{u}_2^{(\mu)}) = 1),\ \tilde{V}(\mathbf{u}_3^{(\mu)}) = 1.$$

The main point is that if some coloring $\tilde{V}$ maps a specific vector $\mathbf{u}$ to 0, then all vectors orthogonal to $\mathbf{u}$ must go to 1. In particular, two orthogonal vectors can never both be sent to 0. To find a contradiction (to the assumption that $\tilde{V}$ exists), we try to assign values $\tilde{V}(\mathbf{u}_i^{(\mu)})$ one after the other, starting in row 1. Here some specific choices will be made, but by symmetry other choices lead to similar contradictions.

1. Suppose that $\tilde{V}(a_1) = (0,1,1)$ (i.e., $\tilde{V}(\mathbf{u}_1^{(1)}) = 0$ and $\tilde{V}(\mathbf{u}_2^{(1)}) = \tilde{V}(\mathbf{u}_3^{(1)}) = 1$). In a_2 this forces $\tilde{V}(\mathbf{u}_3^{(2)}) = 1$, so that either $\mathbf{u}_1^{(2)}$ or $\mathbf{u}_2^{(2)}$ must be mapped to 0 (and the other to 1). Let $\tilde{V}(\mathbf{u}_1^{(2)}) = 0$, so that $\tilde{V}(\mathbf{u}_2^{(2)}) = 1$, i.e., $\tilde{V}(\mathbf{u}_2) = (0,1,1)$. In a_3 one has $\mathbf{u}_3^{(3)} = \mathbf{u}_2^{(1)}$, so $\tilde{V}(\mathbf{u}_3^{(3)}) = 1$. We choose $\tilde{V}(\mathbf{u}_1^{(3)}) = 0$ and hence $\tilde{V}(\mathbf{u}_2^{(3)}) = 1$, so $\tilde{V}(\mathbf{u}_2) = (0,1,1)$. In a_4, the vector $\mathbf{u}_3^{(4)}$ is orthogonal to $\mathbf{u}_1^{(1)}$, which has been mapped to zero already, so that $\tilde{V}(\mathbf{u}_3^{(4)}) = 1$. The remaining free choice is arbitrarily made as $\tilde{V}(\mathbf{u}_1^{(4)}) = 0$, so that $\tilde{V}(\mathbf{u}_2^{(4)}) = 1$ and hence $\tilde{V}(a_4) = (0,1,1)$.
2. But now everything is fixed for a_5 t/m a_{11}, as follows. From a_5, the vector $\mathbf{u}_3^{(5)}$ already occurred in $\mathbf{u}_1$, and moreover, $\mathbf{u}_2^{(5)}$ is orthogonal to $\mathbf{u}_1^{(4)}$ from a_4. Because $\tilde{V}(\mathbf{u}_1^{(4)}) = 0$, one must have $\tilde{V}(\mathbf{u}_2^{(4)}) = 1$. And so on and so forth, yielding $\tilde{V}(a_\mu) = (0,1,1)$ voor $\mu = 5,\ldots,10$ (as was the case also for $\mu = 1,2,3,4$).
3. In a_{11} one has $\mathbf{u}_1^{(11)} = \mathbf{u}_2^{(1)}$, so $\mathbf{u}_1^{(11)}$ is mapped to 1. Furthermore, $\mathbf{u}_2^{(11)}$ is orthogonal to $\mathbf{u}_1^{(4)}$, which was mapped to 0; hence $\mathbf{u}_2^{(11)}$ goes to 1. Finally, $\mathbf{u}_3^{(11)}$ is orthogonal to $\mathbf{u}_1^{(10)}$, which was mapped to 0, so that $\mathbf{u}^{(11)}$ must go to 1. Thus

$$\tilde{V}(a_{11}) = (1,1,1). \tag{6.14}$$

But $(1,1,1)$ is not an admissible value of $\tilde{V}$! So $\tilde{V}$ and hence V cannot exist. □

Corollary 6.8. *There is no function $\tilde{V}$ with the two properties stated in Lemma 6.6.*

The Kochen–Specker Theorem is often stated in the following way.

Definition 6.9. *For any finite-dimensional Hilbert space H, a* **coloring** *of the set $\mathscr{P}_1(H)$ of one-dimensional projections on H is a function*

$$W : \mathscr{P}_1(H) \to \{0,1\}$$

such that for any resolution of the identity (e_i) with $e_i \in \mathscr{P}_1(H)$, i.e.,

$$e_i e_j = \delta_{ij} e_i; \tag{6.15}$$

$$\sum_i e_i = 1_H, \tag{6.16}$$

one has

$$\sum_i W(e_i) = 1, \tag{6.17}$$

so that there is exactly one member e_i of the family such that $W(e_i) = 1$.

Note that if $e \in \mathscr{P}_1(H)$ then $e = e_\psi = |\psi\rangle\langle\psi|$ for some unit vector $\psi \in H$, so that each basis (υ_i) of H defines such a family by $e_i = |\upsilon_i\rangle\langle\upsilon_i|$, and *vice versa*, up to phase factors. The setting of Gleason's Theorem is similar, with the crucial difference that the function on $\mathscr{P}_1(H)$ in question then takes values in $[0,1]$ instead of $\{0,1\}$ and hence can be shown to exist, even amply so (as there are many states).

Theorem 6.10. *If* $\dim(H) > 2$*, there exists no coloring of* $\mathscr{P}_1(H)$.

Proof. For $H = \mathbb{C}^3$, the existence of W would yield the existence of $\tilde{V}$ through

$$\tilde{V}(\mathbf{u}) = 1 - W(e_{\mathbf{u}}), \tag{6.18}$$

where $\mathbf{u} \in \mathbb{R}^3$ is regarded as a vector in $\mathbb{C}^3$. Property 1 in Lemma 6.6 is obviously satisfied. To prove property 2, we note that for any unit vector $\mathbf{u} \in \mathbb{R}^3 \subset \mathbb{C}^3$, we have

$$J_{\mathbf{u}}^2 \mathbf{u} = 0, \tag{6.19}$$

since an explicit computation based on (6.11) shows that, with $\mathbf{u} = (u_1, u_2, u_3)$,

$$J_{\mathbf{u}}^2 = \begin{pmatrix} u_2^2 + u_3^2 & -u_1 u_2 & -u_1 u_3 \\ -u_1 u_2 & u_1^2 + u_3^2 & -u_2 u_3 \\ -u_1 u_3 & -u_2 u_3 & u_1^2 + u_2^2 \end{pmatrix}. \tag{6.20}$$

It follows from rotation invariance that the eigenvalues of $J_{\mathbf{u}}^2$ are the same as those of each J_i^2, cf. (6.8), i.e., $\lambda = 0$ with multiplicity one and $\lambda = 1$ with multiplicity two. Hence (6.19) gives the projection e_0 onto the eigenspace of $J_{\mathbf{u}}^2$ for $\lambda = 0$ as

$$e_0 = |\mathbf{u}\rangle\langle\mathbf{u}| \equiv e_{\mathbf{u}}. \tag{6.21}$$

Property 2 in Lemma 6.6 then follows from the assumption that W is a coloring. Since $\tilde{V}$ cannot exist by Lemma 6.7, neither can W. This proves the claim for $\mathbb{C}^3$.

We finish by induction. Suppose $\mathbb{C}^n$ contains some set $\{\mathbf{u}_k\}_{k\in K}$ of unit vectors that cannot be colored, assuming that $\mathbf{u}_0 = (1, 0, \ldots, 0)$ lies in this set. We embed each $\mathbf{u}_k$ into $\mathbb{C}^{n+1}$ by adding a zero *at the end*, calling the image $\mathbf{u}'_k$. Adding $\mathbf{v} = (0, \ldots, 0, 1)$, the only possible coloring of the set $\{\mathbf{u}'_k, \mathbf{v}\}_{k\in K}$ in $\mathbb{C}^{n+1}$ is given by $W(\mathbf{u}'_k) = 0$ for each $k \in K$ and $W(\mathbf{v}) = 1$. Indeed, if $W(\mathbf{u}'_{k_0}) = 1$ for some k_0, then, since $\mathbf{v}$ is orthogonal to each $\mathbf{u}'_k$, we must have $W(\mathbf{v}) = 0$, which means that the original set $\{\mathbf{u}_k\}_{k\in K}$ should be colorable in $\mathbb{C}^n$, but this is impossible by assumption.

We now embed each $\mathbf{u}_k$ into $\mathbb{C}^{n+1}$ by adding a zero *at the beginning*, denoting its image by $\mathbf{u}''_k$, and add $\mathbf{u}'_0 = (1, 0, \ldots, 0, 0)$. By the same token, the only coloring of the set $\{\mathbf{u}''_k, \mathbf{u}'_0\}_{k\in K}$ is given by $W(\mathbf{u}''_k) = 0$ for each $k \in K$ and $W(\mathbf{u}'_0) = 1$. But this leaves the set $\{\mathbf{u}'_k, \mathbf{u}''_k, \mathbf{v}\}_{k\in K}$ in $\mathbb{C}^{n+1}$ uncolorable, since colorability of $\{\mathbf{u}'_k, \mathbf{v}\}_{k\in K}$ gave $W(\mathbf{u}'_0) = 0$, whereas colorability of $\{\mathbf{u}''_k, \mathbf{u}'_0\}_{k\in K}$ gave $W(\mathbf{u}'_0) = 1$. □

The set thus obtained is larger than necessary. For example, already for $H = \mathbb{C}^4$ the following bases cannot be colored (again writing down unnormalized vectors):

basis	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$\mathbf{u}_4$
a_1	$(0,0,0,1)$	$(0,0,1,0)$	$(1,1,0,0)$	$(1,-1,0,0)$
a_2	$(0,0,0,1)$	$(0,1,0,0)$	$(1,0,1,0)$	$(1,0,-1,0)$
a_3	$(1,-1,1,-1)$	$(1,-1,-1,1)$	$(1,1,0,0)$	$(0,0,1,1)$
a_4	$(1,-1,1,-1)$	$(1,1,1,1)$	$(1,0,-1,0)$	$(0,1,0,-1)$
a_5	$(0,0,1,0)$	$(0,1,0,0)$	$(1,0,0,1)$	$(1,0,0,-1)$
a_6	$(1,-1,-1,1)$	$(1,1,1,1)$	$(1,0,0,-1)$	$(0,1,-1,0)$
a_7	$(1,1,-1,1)$	$(1,1,1,-1)$	$(1,-1,0,0)$	$(0,0,1,1)$
a_8	$(1,1,-1,1)$	$(-1,1,1,1)$	$(1,0,1,0)$	$(0,1,0,-1)$
a_9	$(1,1,1,-1)$	$(-1,1,1,1)$	$(1,0,0,1)$	$(0,1,-1,0)$

The proof is the following observation: if we present the coloring condition as

$$W(0,0,0,1)+W(0,0,1,0)+W(1,1,0,0)+W(1,-1,0,0)=1; \qquad (a_1)$$

$$\cdots \qquad (a_\bullet)$$

$$W(1,1,1,-1)+W(-1,1,1,1)+W(1,0,0,1)+W(0,1,-1,0)=1, \qquad (a_9)$$

then since there are nine such equations the sum of the right-hand sides is odd, whereas the sum of the left-hand sides is even, since each vector appears twice.

To bridge the gap between the Kochen–Specker Theorem and the Free Will Theorem, as well as the one between mathematics and physics, we now rephrase the former as a "mini FWT". We build an experiment consisting of a box containing a spin-1 particle and a device capable of measuring all of the three observables

$$(J^2_{\mathbf{u}_1}, J^2_{\mathbf{u}_2}, J^2_{\mathbf{u}_3})$$

for an arbitrary basis a of $\mathbb{R}^3$; since the operators in question commute, this simultaneous measurement is allowed by quantum theory. The choice of a is called the ***setting*** of the experiment, traditionally denoted by A (in honor of Alice, who is supposed to perform the experiment), with possible values $A = a$. In "phenomenological" notation, the observable measured in an experiment like this is called F, which in the case at hand has three components $F = (F_1, F_2, F_3)$: given the setting a, the observable F_i corresponds to $J^2_{\mathbf{u}_i}$. The notation $F = \lambda$ for $\lambda = (\lambda_1, \lambda_2, \lambda_3)$, i.e., $F_i = \lambda_i$, then expresses the fact that the outcome of a measurement of F is λ.

According to both quantum mechanics and our quasi-linear non-contextual hidden variable theory, either $\lambda_i = 0$ or $\lambda_i = 1$, and λ must lie in the value space

$$\Lambda = \{(0,1,1),(1,0,1),(1,1,0)\}; \qquad (6.22)$$

cf. Lemma 6.6 for the hidden variable theory, while in quantum mechanics (6.22) follows from the fact that λ must lie in the joint spectrum of the three operators $J^2_{\mathbf{u}_i}$.

This, in turn means that there must be a joint eigenvector ψ such that $J^2_{\mathbf{u}_i} = \lambda_i \psi$ for each $i = 1,2,3$. There are three such joint eigenvectors, namely $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ (initially defined as vectors in $\mathbb{R}^3$ but now seen as vectors in $\mathbb{C}^3$), with joint eigenvalues $(0,1,1)$, $(1,0,1)$, and $(1,1,0)$, respectively.

Otherwise, quantum mechanics and our quasi-linear non-contextual hidden variable theory provide a different picture of the experiment. According to the former theory, a given spin-1 particle may be prepared in a (pure) quantum state ψ, which is a unit vector in $\mathbb{C}^3$. Quantum theory then merely predicts probabilities

$$P_\psi(F = \lambda | A = a) \equiv p_{J^2_{\mathbf{u}_1}, J^2_{\mathbf{u}_2}, J^2_{\mathbf{u}_3}}(\lambda_1, \lambda_2, \lambda_3), \tag{6.23}$$

for the possible outcomes λ, which according to the Born rule (2.21) are given by

$$P_\psi(F = \lambda^{(i)} | A = a) = |\langle \mathbf{u}_i, \psi \rangle|^2. \tag{6.24}$$

So if $\psi = \mathbf{u}_i$, then the outcome will be $\lambda = \lambda^{(i)}$ with probability one, but in a superposition $\psi = \sum_i c_i \mathbf{u}_i$ (with $\sum_i |c_i|^2 = 1$), quantum theory predicts a random sequence of outcomes $\lambda^{(i)}$, each with probability $|c_i|^2$.

Let us note that quantum mechanics is non-contextual in the following (probabilistic) sense. Alice could decide to perform just one measurement instead of three, say F_1, with setting $a_1 = \mathbf{u}_1$, or perhaps she may not know if the other two are performed. Fortunately, this does not matter, since for any unit vector $\psi \in \mathbb{C}^3$,

$$P_\psi(F_1 = \lambda_1 | A_1 = \mathbf{u}_1) = \sum_{\lambda_2, \lambda_3} P_\psi(F = \lambda | A = a), \tag{6.25}$$

so that according to quantum mechanics, it does not matter for the Born probabilities of the first measurement if the other two are performed or not.

The question now arises if some quasi-linear non-contextual hidden variable theory theory could improve on this, in that the *probabilities* quantum theory assigns to various outcomes are replaced by *predictions*. In the sprit of determinism (whilst avoiding the appearance of circularity), such a theory should also predict the settings of the experiment. Accordingly, the assumptions leading to our "mini FWT" are:

Definition 6.11. *In the context of the experiment on spin-1 particles just discussed:*

- **Determinism** firstly *means that there is a state space X with associated functions*

$$A : X \to X_A; \tag{6.26}$$

$$F : X \to \Lambda, \tag{6.27}$$

where X_A is the set of all bases in $\mathbb{R}^3$ (i.e. $a \in X_A$), and Λ is some set of possible outcomes; these functions completely describe the experiment in the sense that each state $x \in X$ determines both its settings $a = A(x)$ and its outcome $\lambda = F(x)$. Here $A = (A_1, A_2, A_3)$, where the functions $A_i : X \to S^2$ (seen as the space of unit vectors in $\mathbb{R}^3$) combine to define a basis, and $F = (F_1, F_2, F_3)$, where $F_i : X \to \mathbb{R}$.

Secondly, *there exists some set X_Z and an additional function*

$$Z : X \to X_Z, \tag{6.28}$$

such that

$$F = F(A, Z). \tag{6.29}$$

More precisely, for each $x \in X$ one has

$$F(x) = \hat{F}(A(x), Z(x)) \tag{6.30}$$

for a certain function $\hat{F} : X_A \times X_Z \to \Lambda$. Also this function is, of course, a triple $\hat{F} = (\hat{F}_1, \hat{F}_2, \hat{F}_3)$, where $\hat{F}_i : X_A \times X_Z \to \underline{2}$. In terms of (6.28)*, then:*

- **Nature** *then requires that Λ is given by* (6.22) *(so that $F_i : X \to \underline{2}$).*
- **Freedom** *states that A and Z are* independent *in the sense that the function*

$$\begin{aligned} A \times Z : X &\to X_A \times X_Z \\ x &\mapsto (A(x), Z(x)) \end{aligned} \tag{6.31}$$

is surjective; in other words, for each $(a, z) \in X_A \times X_Z$ there is an $x \in X$ for which $A(x) = a$ and $Z(x) = z$ (making a and z free variables).

- **Non-contextuality** *(cf. Lemma 6.6) finally stipulates that $\hat{F}$ take the form*

$$\hat{F}((\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3), z) = (\tilde{F}(\mathbf{u}_1, z), \tilde{F}(\mathbf{u}_2, z), \tilde{F}(\mathbf{u}_3, z)), \tag{6.32}$$

for a single function $\tilde{F} : S^2 \times X_Z \to \underline{2}$ that also satisfies

$$\tilde{F}(-\mathbf{u}, z) = \tilde{F}(\mathbf{u}, z). \tag{6.33}$$

"Nature" may be taken to be either an experimental result or an uncontroversial prediction of (some corner of) quantum mechanics. The function Z (including its domain X_Z) describes anything relevant to the experiment (such as the behaviour of the particle) *except* the variables determining the settings (which do form part of X). The goal of the freedom assumption is to remove any potential dependencies between the variables (a, z), and hence between the physical system Alice perform her measurements *on*, and the devices she performs her measurements *with*.

Corollary 6.12. *Determinism, Nature, Freedom, and Non-contextuality are contradictory.*

Proof. For each $z \in X_Z$, define a function $\tilde{V}_z : S^2 \to \underline{2}$ by $\tilde{V}_z(\mathbf{u}) = \tilde{F}(\mathbf{u}, z)$. The assumptions combine to give $\tilde{V}_z$ the same properties as $\tilde{V}$ in Lemma 6.6 (where z "goes along for a free ride"). According to Corollary 6.8 (which applies because by *Freedom* one can freely vary a for any given z), the function $\tilde{V}_z$ cannot exist. □

This "mini FWT" is a good exercise for the Free Will Theorem in the next section. For example, let us note, as a warning, that if Determinism is seen as the culprit (and hence falls), then the other assumptions in the (min) FWT are no longer defined. This blocks a direct inference from Freedom to Indeterminism à la Conway & Kochen.

6.2 The Free Will Theorem

The Free Will Theorem is similar in spirit to Corollary 6.12, with the difference that the experiment now has two wings and the *non-contextuality* assumption is replaced by a certain *locality* condition. This condition relates to the setting introduced by Einstein, Podolsky, and Rosen in 1935 and further studied by Bohm, Bell, and others, in which (in current jargon) two physicists, called Alice and Bob, are far apart whilst performing simultaneous experiments on some correlated two-particle state (technically speaking, their measurements need to be *spacelike separated*). In the situation considered by EPR each particle had a spatial degree of freedom and hence required the infinite-dimensional Hilbert space $L^2(\mathbb{R}^3)$ for its description, but, as recognized by Bohm, the thrust of the argument comes out more clearly if each particle merely has an internal degree of freedom (and is "frozen" otherwise).

Bell (1964) considered a pair of spin $\frac{1}{2}$ particles (cf. §6.5), each of which has Hilbert space $\mathbb{C}^2$ (although the famous experiments of Aspect testing the violation of Bell's inequalities used photons, which have the "same" Hilbert space), but because of its reliance on the Kochen–Specker Theorem (which fails for $\mathbb{C}^2$) the Free Will Theorem requires one dimension more, i.e., $H = \mathbb{C}^3$. As before, we see this as the state space of a massive spin-1 particle. The price of this extra dimension is that the pertinent experiment whose outcome provides the *Nature* input for the Free Will Theorem has not actually been performed, but, as in the Bell case, the predictions of quantum mechanics are uncontroversial and will serve as input instead.

These predictions are as follows. Alice and Bob measure on the correlated state

$$\psi_0 = (\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3)/\sqrt{3}, \tag{6.34}$$

where we recall that $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is the standard basis of $\mathbb{R}^3$, now seen as a basis of $\mathbb{C}^3$. This state is rotation-invariant, which means that nonzero angular momentum in one particle must be compensated for in the other, creating the desired correlations.

As before, we denote Alice's setting by $A = a$, which remains the choice of some basis of $\mathbb{R}^3$, but this time also Bob picks some basis b, so that we write $B = b$ for his choice. Similar to Alice's outcome $F = \lambda$ we denote Bob's by $G = \gamma$, and quantum mechanics provides all (Born) probabilities

$$P_{\psi_0}(F = \lambda, G = \gamma | A = a, B = b) \equiv p_{J^2_{\mathbf{u}_1}, J^2_{\mathbf{u}_2}, J^2_{\mathbf{u}_3}, J^2_{\mathbf{v}_1}, J^2_{\mathbf{v}_2}, J^2_{\mathbf{v}_3}}(\lambda_1, \lambda_2, \lambda_3, \gamma_1, \gamma_2, \gamma_3),$$

which are well defined because Alice's squared angular momentum operators $J^2_{\mathbf{u}_1}$ commute with Bob's $J^2_{\mathbf{v}_1}$ as a consequence of Einstein locality (stating that spacelike separated observables commute). Note that similarly to $a = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ for Alice's basis, we write $b = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ for Bob's. If Alice merely measures F_i whilst Bob measures G_j, then, as in the previous section, it does not matter which other (commuting) operators are measured and/or whether Alice and Bob know about this, cf. (6.25). Thus we may write either $(A = a, B = b)$ or $A_i = \mathbf{u}_i, B_i = \mathbf{v}_i$ for the settings, and simple calculations show that the Born probabilities are given by:

$$P_{\psi_0}(F_i = 1, G_j = 1 | A = a, B = b) = \tfrac{1}{3}(1 + \langle \mathbf{u}_i, \mathbf{v}_j \rangle^2); \tag{6.35}$$
$$P_{\psi_0}(F_i = 0, G_j = 0 | A = a, B = b) = \tfrac{1}{3}\langle \mathbf{u}_i, \mathbf{v}_j \rangle^2; \tag{6.36}$$
$$P_{\psi_0}(F_i = 1, G_j = 0 | A = a, B = b) = \tfrac{1}{3}(1 - \langle \mathbf{u}_i, \mathbf{v}_j \rangle^2); \tag{6.37}$$
$$P_{\psi_0}(F_i = 0, G_j = 1 | A = a, B = b) = \tfrac{1}{3}(1 - \langle \mathbf{u}_i, \mathbf{v}_j \rangle^2), \tag{6.38}$$

where $\langle \mathbf{u}_i, \mathbf{v}_j \rangle^2 = |\langle \mathbf{u}_i, \mathbf{v}_j \rangle|^2$, etc., since the vectors are real, In terms of the notation

$$P_{\psi_0}(F_i = G_j | \cdot) = P_{\psi_0}(F_i = 0, G_j = 0 | \cdot) + P_{\psi_0}(F_i = 1, G_j = 1 | \cdot); \tag{6.39}$$
$$P_{\psi_0}(F_i \neq G_j | \cdot) = P_{\psi_0}(F_i = 0, G_j = 1 | \cdot) + P_{\psi_0}(F_i = 1, G_j = 0 | \cdot), \tag{6.40}$$

this yields

$$P_{\psi_0}(F_i = G_j | A = a, B = b) = \tfrac{1}{3}(1 + 2\langle \mathbf{u}_i, \mathbf{v}_j \rangle^2); \tag{6.41}$$
$$P_{\psi_0}(F_i \neq G_j | A = a, B = b) = \tfrac{2}{3}(1 - \langle \mathbf{u}_i, \mathbf{v}_j \rangle^2). \tag{6.42}$$

The crucial point for the Free Will Theorem is that this implies ***perfect correlation***:

$$P_{\psi_0}(F_i = G_j | A_i = B_j) = 1, \tag{6.43}$$

in agreement with the intuition about angular momentum expressed earlier.

We now move to a (possibly counterfactual) deterministic description of this experiment along the lines of the previous section. It is straightforward to adapt all of Definition 6.11 except Non-contextuality (which after all is the assumption we would like to get rid of!). With the obvious changes, we obtain:

- ***Determinism*** again *first* claims there is a state space X with associated functions

$$A : X \to X_A; \tag{6.44}$$
$$B : X \to X_B; \tag{6.45}$$
$$F : X \to \Lambda; \tag{6.46}$$
$$G : X \to \Lambda, \tag{6.47}$$

  where $X_A = X_B$ is the set of all bases in $\mathbb{R}^3$, and Λ is some set of possible outcomes, which completely describe the experiment in the sense that each state $x \in X$ determines both its settings $(a = A(x), b = B(x))$ and its outcome $(\lambda = F(x), \gamma = G(x))$. Here $A = (A_1, A_2, A_3)$ and $B = (B_1, B_2, B_3)$ where the functions $A_i : X \to S^2$ (where S^2 is seen as the space of unit vectors in $\mathbb{R}^3$) combine to define a basis (similarly for $B_j : X \to S^2$), and $F = (F_1, F_2, F_3)$. *Secondly*, there exists some set X_Z and an additional function $Z : X \to X_Z$ such that

$$F = F(A, B, Z); \tag{6.48}$$
$$G = G(A, B, Z), \tag{6.49}$$

  in that for each $x \in X$ one has the functional relationships

$$F(x) = \hat{F}(A(x), B(x), Z(x)); \tag{6.50}$$
$$G(x) = \hat{G}(A(x), B(x), Z(x)), \tag{6.51}$$

for certain functions $\hat{F} : X_A \times X_B \times X_Z \to \Lambda$ and $\hat{G} : X_A \times X_B \times X_Z \to \Lambda$, each of which is a triple $\hat{F} = (\hat{F}_1, \hat{F}_2, \hat{F}_3)$ with $\hat{F}_i : X_A \times X_B \times X_Z \to \mathbb{R}$, etc. The value $z = Z(x)$ is just the traditional "hidden variable" (which is often denoted by λ).

- ***Freedom*** then states that A, B, and Z are *independent* in that for each $(a,b,z) \in X_A \times X_B \times X_Z$ there is an $x \in X$ for which $A(x) = a$, $B(x) = b$, and $Z(x) = z$.
- ***Nature*** requires that:
  - Λ is given by (6.22), i.e. F_i and G_j, and hence $\hat{F}_i$ and $\hat{G}_j$ take values in $\{0,1\}$;
  - The experiment measures *squares* of angular momenta, so that

$$\hat{F}(a', b', z) = \hat{F}(a, b, z); \tag{6.52}$$
$$\hat{G}(a', b', z) = \hat{G}(a, b, z), \tag{6.53}$$

    whenever (a', b') differ from (a, b) by changing the sign of any basis vector;
  - *Perfect correlation* obtains, cf. (6.43), i.e., writing $a = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ for Alice's basis and $b = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ for Bob's, one has

$$\mathbf{u}_i = \mathbf{v}_j \Rightarrow \hat{F}_i(a, b, z) = \hat{G}_j(a, b, z). \tag{6.54}$$

We now come to the locality condition that is to replace *Non-contextuality*. This condition was first clearly stated by Bell (1964, p. 196), who attributes it to Einstein:

> 'The vital assumption is that the result G for particle 2 does not depend on the setting a of the magnet for particle 1, nor F on b.'

Noting various other notions of locality (such as *Einstein locality* in local quantum physics, which requires spacelike separated operators to commute, or *Bell locality*, discussed below), the above idea might be called *Context locality*, but we will simply refer to it as *Locality*. In our deterministic setting, a precise formulation is this:

- ***Locality*** means that $F(A,B,Z)$ is independent of B and $G(A,B,Z)$ is independent of A. In other words, we have $F = F(A,Z)$ and $G = G(B,Z)$, so that (with slight abuse of notation) $\hat{F} : X_A \times X_Z \to \Lambda$ and $\hat{G} : X_B \times X_Z \to \Lambda$, or, then again, $F(x) = \hat{F}(A(x), Z(x))$ and $G(x) = \hat{G}(B(x), Z(x))$, for each $x \in X$.

This finally brings us to (our reformulation of) the ***Free Will Theorem***:

Theorem 6.13. *Determinism, Freedom, Nature, and Locality are contradictory.*

Proof. The *Freedom* assumption allows us to treat (a,b,z) as free variables, a fact that will tacitly be used all the time. First, taking $i = j$ in (6.54) shows that $\hat{F}_i(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, z)$ only depends on $(\mathbf{u}_i, z)$, whilst $\hat{G}_j(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, z)$ only depends on $(\mathbf{v}_j, z)$. Hence we write $\hat{F}_i(a,z) = \tilde{F}_i(\mathbf{u}_i, z)$, etc. Next, taking $i \neq j$ in (6.54) shows that $\tilde{F}_1(\mathbf{u}, z) = \tilde{F}_2(\mathbf{u}, z) = \tilde{F}_3(\mathbf{u}, z)$. Consequently, the function $\hat{F} : X_A \times X_Z \to X_F$ is given by (6.32). We are now back to the proof of Corollary 6.12, concluding that such a function does not exist by Corollary 6.8. □

6.3 Philosophical intermezzo: Free will in the Free Will Theorem

> 'The determinism-free will controversy has all of the earmarks of a dead problem. The positions are well staked out and the opponents manning them stare at each other in mutual incomprehension.' (Earman, 1986, p. 235)

The question arises which specific notion of free will is among the assumptions of the FWT (in the reformulation just given). To put this question in perspective, let us briefly recall the main point of the debate about free will. This concept has two poles. One is the "will" itself, requiring a sense of ***agency***, deliberation, and control. This pole seems to require some form of determinism. A powerful expressions is:

> 'Fürst! Was Sie sind, sind Sie durch Zufall und Geburt. Was ich bin, bin ich durch mich.'[1] (Beethoven, to his benefactor (!) Prince Lichnowsky)

The other pole of free will is the adjective "free", i.e., ***the ability to do otherwise***, which at first sight requires indeterminism. *The problem of free will is that these poles seem contradictory*. Many authors conflate free will with moral responsibility:

> 'free will can be defined as the unique ability of persons to exercise control over their conduct in the manner necessary for moral responsibility.' (McKenna & Coates, 2015)

This aspect is irrelevant to our discussion, concerned as it is with the question what it would mean for Alice and Bob to choose their settings "freely" if determinism is assumed (it would have been different if one setting launched a nuclear missile).

Even in our narrow context, the traditional philosophical stances are relevant:

- ***Compatibilism*** denies the contradiction, claiming that free will and determinism coexist. This position may be defended in many ways, among which one finds:
  - Reconceptualizing "the ability to do otherwise" in a deterministic world. This will be our focus in what follows, especially in a version inspired by Lewis.
  - Belittling the relevance of "the ability to do otherwise", as e.g. by Dennett:

    > 'So if anyone at all is interested in the question of whether one could have done otherwise in *exactly* the same circumstances (and internal state) this will have to be a particularly pure metaphysical curiosity—that is to say, a curiosity so pure as to be utterly lacking in any ulterior motive, since the answer could not conceivably make any noticeable difference to the way the world went.' (Dennett, 1984, p. 559).

- ***Incompatibilism*** accepts the contradiction, once again branching off into:
  - *Libertarianism*, arguing that free will requires an indeterministic world.
  - *Hard determinism*, claiming determinism (which is assumed) blocks free will:

    > 'Ein Mensch kann zwar tun was er will, aber nicht wollen was er will.'[2] (Schopenhauer)

  - *Hard incompatibilism*, asserting that 'every way you look at it you lose': free will makes no sense in either a deterministic or an indeterministic world.

[1] 'Lord! What you are, you are through chance and birth. What I am, I am because of myself.'

[2] 'One can admittedly do what one wants, but one cannot want what one wants.'

Although hard incompatibilism has our sympathy, our opening question concerning the notion of free will in the FWT drives us into the compatibilist direction, since determinism is among the assumptions shown to be contradictory by Theorem 6.13. Within compatibilism, we will be close to the well-known 'local miracle' variant thereof proposed by the philosopher David Lewis. Like other compatibilists before him (starting at least with G.E. Moore), Lewis attempts to make sense of the intuition that even in a deterministic world one in principle has the ability to act differently from the way one actually does, despite the fact that the latter was predetermined. A simple example is Alice's choosing setting a by moving her hand in a certain way, although she was able to choose a'. On the other hand, she could not have moved her hand with a speed greater than that of light, so her ability remains constrained by the laws of nature. Lewis asks us to distinguish between:

- 'I am able to do something such that, if I did it, a law would be broken.'
- 'I am able to break a law.'

The latter is impossible, but the former is not on Lewis's own theory of counterfactuals, according to which the phrase 'if I did it' leads us to consider the possible world in which doing 'something' is actually true, whilst in the possible worlds under consideration as many other features as possible are kept the same as in the actual world (the precise underlying measure of similarity is not important here). Thus the phrase 'a law would be broken' refers to the laws of the actual world (in which the alternative action is not realized). It seems to be of great importance to Lewis that in the first case it is not the agent who would break a law; instead, it is the breaking of some law of our actual world at an earlier time that enables the subject to do in an alternative possible world what she could not do in our actual world, .

By making this distinction, Lewis claims that he invalidates the seemingly lethal ***Consequence Argument*** against compatibilist free will, of which a simple version reads (assuming determinism, on which compatibilist free will is predicated):

1. Alice's actions are a necessary consequence of the laws of nature plus the state of the universe (or the relevant part thereof) at any earlier time;
2. Alice is unable to render both (laws and earlier states) false;
3. Alice is unable to render the consequences of laws and earlier states false;
4. *Ergo*: Alice is unable to do otherwise than what she actually does.

Lewis claims that statement 3 is ambiguous, in that it fails to distinguish between the two senses in his two bullet points above. The Consequence Argument requires the latter (which is false), whereas this argument itself is unsound on the former (which is true). This disambiguation of assumption 3 in the Consequence Argument, then, is supposed to save (compatibilist) free will. However, a considerable philosophical literature suggests that the tension between Lewis's denying the second bullet point whilst accepting the first is pretty uncomfortable, reflecting the corresponding tension between the conjunction of determinism and freedom in general; indeed, this is what the FWT makes precise! Let us first point out that, at least in his terminology Lewis fails to make a clear distinction between *laws of nature* and *initial states*; from the point of view of modern physics, this distinction is absolutely fundamental (although it may dispappear in post-modern physics based in e.g. quantum gravity).

Lewis's examples of law-breaking events in our actual world typically refer to violations of some law of nature (like exceeding the speed of light), whereas the (alleged) law-breaking in his counterfactuals, such as choosing a' (where in fact Alice did not do so) amounts to a change in some earlier state. Thus it might have been more appropriate if the paper in which Lewis laid out his version of compatibalism had been entitled *Are we free to change the states?* instead of *Are we free to break the laws?*. On this revision, his distinction of the two cases takes the following form:

- I am able to do something such that, if I did it, the state of the actual world at some earlier time would have been different.
- I am able to change the actual state of the world.

The latter remains impossible, while it is the former that enables free will. Applied to Alice, the former should mean (still in the compatibilist spirit of Lewis):

- A slight alteration in the state of the actual world (which would have made it a different but very similar world according to Lewis) would have led Alice to do something (such as choosing a') that she did not do in the actual world (because according to determinism its actual state at any earlier time—as opposed to the counterfactual alternative state in the discussion—led her to choose a).

We now make this revised version of Lewis's local miracle compatibilism mathematically precise, in a way that has the additional advantage of involving not only "the ability do do otherwise", but also the other component free will, i.e. agency. Here the intuition is that free will involves a separation between the agent, Alice, (who is to exercise it) and the rest of the world, under whose influence she acts. Namely, as in the FWT, let X be the state space of the Universe, and let

$$a = A(x) \tag{6.55}$$

again be Alice's setting, where $A : X \to X_A$, as before. We now assume that a is determined by her "inner state" I as well as the "outer state" O of the rest of the world, under whose influence she acts. These, in turn, are determined by the state $x \in X$ of the world. That is, $A = A(O,I)$, which expresses the existence of functions

$$O : X \to X_O; \tag{6.56}$$

$$I : X \to X_I; \tag{6.57}$$

$$\hat{A} : X_O \times X_I \to X_A, \tag{6.58}$$

where X_O and X_I are certain sets, such that for each $x \in X$ one has

$$A(x) = \hat{A}(O(x), I(x)). \tag{6.59}$$

In other words, for some given state x of the world we have

$$o = O(x); \tag{6.60}$$

$$i = I(x); \tag{6.61}$$

$$a = \hat{A}(o,i). \tag{6.62}$$

Note that, in the spirit of Conway and Kochen, in the above analysis Alice (whose free choice they after all believe to be ultimately a consequence of the free choice of elementary particles) now plays the role of the spin-1 particles in the bipartite experiment. Thus the analogy is between the triples:

$$(a, z, \lambda) \in X_A \times Z \times \Lambda; \tag{6.63}$$

$$(o, i, a) \in X_O \times X_I \times X_A. \tag{6.64}$$

- The first triple is defined in the experimental context of the FWT, where a is the setting of Alice's wing of the experiment (which from the perspective of the spin-1 particle plays the role of the outer state of the world), z is the inner state of the particle, and λ is the outcome of Alice's measurement.
- The second pertains to the analysis of Alice's "free" choice of the setting of her experiment, where o is the outer state of the world, i is her inner state, and a is her actual setting, given $x \in X$ and hence $(o, i) = (O(x), I(x))$.

Beyond ***Determinism***, which is expressed by the above framework, our fundamental assumption underpinning compatibilist free will is ***Freedom***, defined exactly as in the FWT: O and I are *independent* in that the following function is surjective:

$$\begin{aligned} O \times I : X &\to X_O \times X_I \\ x &\mapsto (O(x), I(x)), \end{aligned} \tag{6.65}$$

i.e., for each pair $(o, i) \in X_I \times X_O$ there is $x \in X$ for which (6.60) and (6.61) hold.

Rephrasing our earlier analysis in this elementary mathematical language, Lewis wants to make sense of the idea that although Alice's choice (6.62) at some fixed time t was determined by the state x of the Universe at that time through (6.60) - (6.61), or, equivalently, through (6.59), and hence—and this is the whole point of the Consequence Argument Lewis challenges—by any earlier state x_p of the Universe at time t_p, *nonetheless* Alice was "able to act otherwise" at time t, e.g. in choosing

$$a' = \hat{A}(o', i'), \tag{6.66}$$

but did not do so, since choosing a' would illegally have changed the state x to x' (both at time t), and, equivalently (given determinism), would have changed x_p to x'_p. On our reading of Lewis's theory of counterfactuals, Alice's ability to choose a' simply means that there exists a state x' of the world close to x in the sense that

$$O(x') = O(x) = o, \tag{6.67}$$

making the environment in which Alice acts the same as in the actual world, but

$$i' = I(x') \neq I(x) = i, \tag{6.68}$$

where i' should be close to i in some appropriate sense (such as a slight change in the state of Alice's brain), such that (6.66) holds, with $o' = o$ as required by (6.67).

The point, then, is that according to our *Freedom* assumption, there indeed *is* such a nearby state x', for any given i' and (o, i). Thus the freedom Alice has is precisely what we have formalized as *Freedom*: even *given* the state o of the causal influences on her behaviour (and possibly even the entire state of the rest of the world), there is a different admissible state x' of the world such that, had this state been actual, she would have chosen a' (although she in fact, necessarily, picked a).

It should be clear now that at least in the context of the Free Will Theorem, our precise technical formulation of all assumptions implies that the freedom Alice and Bob have in choosing their settings is an instance of the local miracle compatibilist form of free will proposed by Lewis (1981), at least if one accepts our reformulation thereof. The theorem then establishes a contradiction between:

- the physics assumptions, i.e., *Nature*, and *Locality*;
- the compatibilist free will assumption, i.e., *Determinism* and *Freedom*.

Accepting the former, the latter must fall. Making this choice, one should realize that the physics assumptions on the one hand just form a small corner of modern physics (from which point of view they are weak), but on the other hand have singled out the corner in which the two fundamental theories of quantum mechanics and special relativity meet and are brought to a head (from which perspective they are strong).

The challenge their theorem puts to compatibalism was recognized by Conway & Kochen (2009), who write:

> 'The tension between human free will and physical determinism has a long history. Long ago, Lucretius made his otherwise deterministic particles swerve unpredictably to allow for free will. It was largely the great success of deterministic classical physics that led to the adoption of determinism by so many philosophers and scientists, particularly those in fields remote from current physics. (This remark also applies to "compatibilism", a now unnecessary attempt to allow for human free will in a deterministic world.)'

This quotation does not use a precise version of compatibilism, but, as Conway explains elsewhere, what they mean is that compatibilism in whatever form was a desperate pre-twentieth-century attempt to save the notion of free will for e.g. Christianity in the face of the physics of the time, which assumed that the universe was a mechanical clockwork. Such attempts, then, would no longer be necessary if the world is, in fact, indeterministic (as Conway and Kochen claim to have at last proved). Our reformulation of their theorem (which removes the threat of circularity) gives a more subtle picture: the FWT uses modern physics to challenge one particular version of *compatibilist free will*. As such, it only provides indirect support for *libertarian free will*, namely by weakening one of its competitors.

To close this philosophical intermezzo, let us note that determinism is seen as a property of *theories*. Since it is the job of a deterministic theory to predict the outcome of any experiment, whether or not it is performed, this obviates the need for assumptions like counterfactuality in the sense that 'unperformed experiments have results' (which was famously denied by Asher Peres). Such controversial notions of counterfactuality have effectively been replaced by the considerably more refined modal counterfactuality of Lewis (at least in our slight reformulation thereof).

6.4 Technical intermezzo: The GHZ-Theorem

The essence of the proof of the Free Will Theorem lies in the argument that perfect correlation together with context-locality implies non-contextuality. Remarkably, context-locality is at the same time a special case of non-contextuality, as the following example illustrates. We take $H = \mathbb{C}^2 \otimes \mathbb{C}^2$, equipped with the ***Bell basis***

$$\upsilon_0 = (|01\rangle - |10\rangle)/\sqrt{2}; \tag{6.69}$$
$$\upsilon_1 = (|01\rangle + |10\rangle)/\sqrt{2}; \tag{6.70}$$
$$\upsilon_2 = (|00\rangle - |11\rangle)/\sqrt{2}; \tag{6.71}$$
$$\upsilon_3 = (|00\rangle + |11\rangle)/\sqrt{2}, \tag{6.72}$$

where we use the physicists' notation

$$|1\rangle = (1,0); \tag{6.73}$$
$$|0\rangle = (0,1); \tag{6.74}$$
$$|ij\rangle = |i\rangle \otimes |j\rangle. \tag{6.75}$$

Of course, $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$ contains the spin-1 Hilbert space $\mathbb{C}^3$ of the Kochen–Specker Theorem as the subspace orthogonal to the vector υ_0. Thus we identify $\mathbb{C}^3$ with the subspace $\tilde{\mathbb{C}}^3$ of $\mathbb{C}^4$ spanned by the basis vectors $\upsilon_1, \upsilon_2, \upsilon_3$. The operators

$$\tilde{J}_{\mathbf{u}} = \tfrac{1}{2}(\sigma_{\mathbf{u}} \otimes 1_2 + 1_2 \otimes \sigma_{\mathbf{u}}), \tag{6.76}$$

where $\mathbf{u} \in \mathbb{R}^3$ is a unit vector as before, and

$$\sigma_{\mathbf{u}} = \sum_{i=1}^{3} \sigma^i u_i \tag{6.77}$$

in terms of the Pauli matrices σ^i, map υ_1 to zero and leave its orthogonal complement $\tilde{\mathbb{C}}^3$ stable. Elementary group theory or direct calculation then shows that the operator $J_{\mathbf{u}}$ on $\mathbb{C}^3$ in (6.11) is (unitarily) equivalent to the operator $\tilde{J}_{\mathbf{u}}$ on $\tilde{\mathbb{C}}^3$. Since

$$\tilde{J}^2_{\mathbf{u}} = \tfrac{1}{2}(\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}} + 1_2 \otimes 1_2), \tag{6.78}$$

the Kochen–Specker argument can be rephrased in terms of the operators $\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{u}}$. In particular, for each frame $a = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, the three operators

$$(\sigma_{\mathbf{u}_1} \otimes \sigma_{\mathbf{u}_1}, \sigma_{\mathbf{u}_2} \otimes \sigma_{\mathbf{u}_2}, \sigma_{\mathbf{u}_3} \otimes \sigma_{\mathbf{u}_3}) \tag{6.79}$$

commute, they each square to one, and their joint eigenvalues are one of the triples:

$$(-1,-1,-1), (-1,1,1), (1,-1,1), (1,1,-1).$$

The eigenvector corresponding to the first one is υ_0, and hence the others must lie in $\tilde{\mathbb{C}}^3$. Hence by Lemma 6.4 any quasi-linear non-contextual hidden variable must also assign these values, which by Lemma 6.7 is impossible for arbitrary bases.

The key mathematical property of the three operators (6.79) is that they commute, and together with the unit $1_2 \otimes 1_2$ form a maximal set of commuting self-adjoint matrices on $\mathbb{C}^4$. But other such sets could have been chosen by Alice (under whose sole control the situation so far has been assumed to be), such as a triple of the kind

$$(\sigma_{\mathbf{u}} \otimes 1_2, 1_2 \otimes \sigma_{\mathbf{v}}, \sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{v}}),$$

where $\mathbf{u}$ and $\mathbf{v}$ are arbitrary unit vectors in $\mathbb{R}^3$. Since the third operator is the product of the first two, the joint eigenvalues of this triple, and hence also the assignments by a quasi-linear non-contextual hidden variable, must be one of the four triples

$$(1,1,1), (-1,1,-1), (1,-1,-1), (-1,-1,1).$$

The non-contextuality assumption would then dictate that the outcome of Alice's measurement of $\sigma_{\mathbf{u}} \otimes 1_2$ be independent of her choice of the setting $\mathbf{v}$ in a possible simultaneous measurement of $1_2 \otimes \sigma_{\mathbf{v}}$, and *vice versa*. Therefore, in a (non-local) bipartite setting where Alice is only able to measure operators of the type $a \otimes 1_2$, whilst Bob can measure $1_2 \otimes b$, on the above choice of (commuting) operators, *non-contextuality in the situation where Alice controls everything is mathematically equivalent to (context) locality in the bipartite Alice & Bob setting*.

Further constraints then arise if the system is prepared in a correlated state like ψ_0, which is an eigenstate of $\sigma_{\mathbf{u}} \otimes \sigma_{\mathbf{v}}$ with eigenvalue -1 whenever $\mathbf{u} = \mathbf{v}$. So in that case the values of $(\sigma_{\mathbf{u}} \otimes 1_2, 1_2 \otimes \sigma_{\mathbf{v}})$ can only be $(1,-1)$ or $(-1,1)$, yielding perfect anti-correlation. This is not enough, however, to derive a Free Will Theorem; to do so with the small single-site Hilbert space $\mathbb{C}^2$, one needs a third (non-local) party.

Indeed, the well-known tripartite GHZ-argument may be rephrased as a Free Will Theorem, as follows. The underlying Hilbert space is

$$H = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^8, \tag{6.80}$$

and hence as a warm-up we first (re)prove Theorem 6.5 for $n = 8$. Suppose we have a map $V : H_8(\mathbb{C}) \to \mathbb{R}$ as in Definition 6.1. Write

$$\lambda_1^{(a)} = V(\sigma_a \otimes 1_2 \otimes 1_2), \lambda_2^{(b)} = V(1_2 \otimes \sigma_b \otimes 1_2), \lambda_3^{(c)} = V(1_2 \otimes 1_2 \otimes \sigma_c),$$

where a, b, c can be $1, 2, 3$. From Lemma 6.4 we then have

$$V(\sigma_1 \otimes \sigma_2 \otimes \sigma_2) = \lambda_1^{(1)} \lambda_2^{(2)} \lambda_3^{(2)}; \tag{6.81}$$

$$V(\sigma_2 \otimes \sigma_1 \otimes \sigma_2) = \lambda_1^{(2)} \lambda_2^{(1)} \lambda_3^{(2)}; \tag{6.82}$$

$$V(\sigma_2 \otimes \sigma_2 \otimes \sigma_1) = \lambda_1^{(2)} \lambda_2^{(2)} \lambda_3^{(1)}; \tag{6.83}$$

$$V(\sigma_1 \otimes \sigma_1 \otimes \sigma_1) = \lambda_1^{(1)} \lambda_2^{(1)} \lambda_3^{(1)}. \tag{6.84}$$

Furthermore, the four operators on the left-hand side commute and turn out to satisfy

$$\sigma_1 \otimes \sigma_2 \otimes \sigma_2 \cdot \sigma_2 \otimes \sigma_1 \otimes \sigma_2 \cdot \sigma_2 \otimes \sigma_2 \otimes \sigma_1 = -\sigma_1 \otimes \sigma_1 \otimes \sigma_1, \tag{6.85}$$

so that again by Lemma 6.4,

$$\lambda_1^{(1)}\lambda_2^{(2)}\lambda_3^{(2)} \cdot \lambda_1^{(2)}\lambda_2^{(1)}\lambda_3^{(2)} \cdot \lambda_1^{(2)}\lambda_2^{(2)}\lambda_3^{(1)} = -\lambda_1^{(1)}\lambda_2^{(1)}\lambda_3^{(1)}, \tag{6.86}$$

i.e. $(\lambda_1^{(1)}\lambda_2^{(2)}\lambda_3^{(2)})^2 = -1$. Since $\lambda_j^{(i)} = \pm 1$, this is impossible, so that V cannot exist.

Now, using the notation in the preceding discussion, consider the unit vector

$$\psi_{GHZ} = (|111\rangle - |000\rangle)/\sqrt{2}, \tag{6.87}$$

which is a joint eigenstate of each of the four operators on the left-hand side of (6.81) - (6.84), with eigenvalue $+1$ for the first three, and hence eigenvalue -1 for the fourth, i.e., $\sigma_1 \otimes \sigma_1 \otimes \sigma_1$. So if setting $A = a$ for Alice (where $a \in \{1,2\}$) means that she measures $F = \sigma_a \otimes 1_2 \otimes 1_2$ with outcome $\lambda_1^{(a)} = \pm 1$, and similarly $B = b$ for Bob and $C = c$ for Cindy mean that they measure $G = 1_2 \otimes \sigma_b \otimes 1_2$ and $H = 1_2 \otimes 1_2 \otimes \sigma_c$ with outcomes $\lambda_2^{(b)} = \pm 1$ and $\lambda_3^{(c)} = \pm 1$, respectively, then in the state ψ_{GHZ} each of the settings gives the correlation

$$\text{settings } (a,b,c) = (1,2,2),(2,1,2),(2,2,1) \Rightarrow \lambda_1^{(a)}\lambda_2^{(b)}\lambda_3^{(c)} = 1; \tag{6.88}$$

$$\text{setting } (a,b,c) = (1,1,1) \Rightarrow \lambda_1^{(a)}\lambda_2^{(b)}\lambda_3^{(c)} = -1. \tag{6.89}$$

Theorem 6.14. *The conjunction of the following assumptions is contradictory:*

- **Determinism***: there is a state space X with associated functions*

$$A,B,C : X \to \{1,2\}, F,G,H : X \to \Lambda,$$

  which completely describes the experiment, in that $x \in X$ determines both settings (a,b,c) and outcomes $(\lambda_1,\lambda_2,\lambda_3) \in \Lambda^3$ through $a = A(x)$, $\lambda_1 = F(x)$, etc.
- **Nature:** *the experiment (performed in the state ψ_{GHZ}) has possible outcomes in $\Lambda = \{-1,1\}$, subject to the correlations* (6.88) - (6.89)*;*
- **Freedom***: there is a further function $Z : X \to X_Z$, in terms of which*

$$F = F(A,B,C,Z),\ \ G = G(A,B,C,Z),\ \ H = H(A,B,C,Z),$$

  and F, G, H, Z are independent, i.e. for each (a,b,c,z) there is $x \in X$ such that

$$A(x) = a,\ \ B(x) = b,\ \ C(x) = c,\ \ Z(x) = z.$$

- **Locality***: $F = F(A,Z)$, $G = G(B,Z)$, and $H = H(C,Z)$.*

Proof. Using notation as in the proof of Theorem 6.13, for fixed $z \in Z$ we obtain $\hat{F}(a,z) = \lambda_1^{(a)}$ etc. *Nature* then leads to the contradiction derived after (6.86). □

6.5 Bell's theorems

Two different results are known as "Bell's Theorem": the first, from his paper in 1964, is Theorem 6.15 below, and the second, dating from 1976, is Theorem 6.18. The first is similar to the Free Will Theorem in both its assumptions and its conclusion, and to make this similarity more obvious we first state it for $\mathbb{C}^3$ instead of $\mathbb{C}^2$. The difference lies in the probabilistic flavour of Bell's Theorem, whose empirical input is not given by the only non-probabilistic consequence to be drawn from the quantum-mechanical formulae (6.35) - (6.38), viz. the certainty (6.43) of perfect correlation on identical settings, but rather by the probabilistic formula (6.40), i.e.,

$$P_{\psi_0}(F_i \neq G_j | A_i = \mathbf{u}_i, B_j = \mathbf{v}_j) = \tfrac{2}{3}\sin^2\theta_{\mathbf{u}_i,\mathbf{v}_j} \quad (i,j = 1,2,3), \tag{6.90}$$

where $\theta_{\mathbf{u},\mathbf{v}}$ is the angle between two unit vectors $\mathbf{u}$ and $\mathbf{v}$. Furthermore, the state space X must be upgraded to a probability space (X,Σ,μ), carrying functions A and B (for the settings, which unlike Bell himself—who treated them as labels—we include among the random variables), F and G (for the outcomes) and finally Z (for the hidden variable traditionally called λ) as random variables, i.e., measurable functions. This also implies that the target spaces X_A to X_Z (which is traditionally called Λ) must be equipped with some σ-algebra of measurable subsets. But this is not a big deal, since $X_A = X_B$ carries a natural Borel structure and $X_F = X_G$ is finite. The probability measure μ is assumed independent of (A,B,F,G), and *vice versa*.

The measure μ, which gives the "hidden state" of the system that allegedly underlies its quantum-mechanical description, is chosen in such a way that empirical probabilities (typically obtained from long runs of repeated measurements) are recovered as joint conditional probabilities defined by μ and the random variables, i.e., assuming the settings (a,b) are possible in that $P(A = a, B = b) > 0$, we put

$$P(F = \lambda, G = \gamma | A = a, B = b) = \frac{P(F = \lambda, G = \gamma, A = a, B = b)}{P(A = a, B = b)}, \tag{6.91}$$

where the joint probabilities on the right-hand side are given by

$$P(A = a, B = b) = \mu(A = a, B = b\}; \tag{6.92}$$

$$P(F = \lambda, G = \gamma, A = a, B = b) = \mu(F = \lambda, G = \gamma, A = a, B = b\}, \tag{6.93}$$

where $\mu(A = a, B = b)$ is shorthand for $\mu(x \in X \mid A(x) = a, B(x) = b\}$, etc. This implies that μ depends on (but may not be determined by) the quantum state ψ_0.

On this understanding, the assumptions of ***Determinism*** and ***Locality*** are the same as for the Free Will Theorem (except that equations like $F(x) = \hat{F}(A(x), Z(x))$ are merely supposed to hold almost everywhere with respect to μ). ***Freedom*** is now taken to mean that (A,B,Z) are *probabilistically independent* relative to μ. By definition, this also means that the pairs (A,B), (A,Z), and (B,Z) are independent, so that for any $\mathsf{A} \subset X_A$, $\mathsf{B} \subset X_B$, and (measurable) $\mathsf{Z} \subset X_Z$, defining

$$P(A \in \mathsf{A}, B \in \mathsf{B}, Z \in \mathsf{Z}) = \mu(x \in X \mid A(x) \in \mathsf{A}, B(x) \in \mathsf{B}, Z(x) \in \mathsf{Z}), \tag{6.94}$$

and analogous expressions for $P(A \in \mathsf{A})$ and $P(A \in \mathsf{A}, B \in \mathsf{B})$, etc., we have

$$P(A \in \mathsf{A}, B \in \mathsf{B}) = P(A \in \mathsf{A})P(B \in \mathsf{B}); \tag{6.95}$$

$$P(A \in \mathsf{A}, Z \in \mathsf{Z}) = P(A \in \mathsf{A})P(Z \in \mathsf{Z}); \tag{6.96}$$

$$P(B \in \mathsf{B}, Z \in \mathsf{Z}) = P(B \in \mathsf{B})P(Z \in \mathsf{Z}); \tag{6.97}$$

$$P(A \in \mathsf{A}, B \in \mathsf{B}, Z \in \mathsf{Z}) = P(A \in \mathsf{A})P(B \in \mathsf{B})P(Z \in \mathsf{Z}). \tag{6.98}$$

If we finally define ***Nature*** as the claim that $\hat{F}$ and $\hat{G}$ are $\underline{2}$-valued and that

$$P(F_i \neq G_j | A_i = \mathbf{u}_i, B_j = \mathbf{v}_j) = \tfrac{2}{3}\sin^2\theta_{\mathbf{u}_i,\mathbf{v}_j} \;\; (i,j = 1,2,3), \tag{6.99}$$

where the left-hand side is the *conditional* probability defined by μ and the random variables in question (whereas the left-hand side of (6.90) is the *empirical* probability for the experiment in question, or, equivalently, the quantum-mechanical prediction thereof), then we obtain the following spin-1 version of ***Bell's first theorem***:

Theorem 6.15. *Determinism, Freedom, Nature, and Locality are contradictory.*

This formulation is literally the same as Theorem 6.13, but the terms have acquired a different technical meaning now, especially *Freedom* and *Nature*. Moreover, purists would add *Probability Theory* as an assumption in Bell's Theorem, as its formalism is decidedly non-tautological and its interpretation is far from obvious, even in a classical setting. In any case, the proof is practically the same as in the more familiar optical version of the EPR-experiment, to which we now turn.

In the classical (sic) form of the experiment, Alice and Bob perform measurements on incoming photons by letting them pass through a polaroid glass whose axis of polarization makes angle a (Alice) or b (Bob) with (say) the horizontal axis in the plane orthogonal to the direction of propagation of the photons. Considered in the light of the previous experiment on spin-1 particles, such a choice of settings may also be seen as a choice of basis for $\mathbb{R}^3$, with the proviso that, assuming (by convention) the photons move along the y-axis, one basis element $\mathbf{u}_2 = (0,1,0)$ is fixed so that the remaining two vectors $(\mathbf{u}_1, \mathbf{u}_3)$ must lie in the x-z plane (in which, on a naive picture, the photons may "vibrate"). This constraint gives rise to bases

$$\mathbf{u}_1 = (\cos a, 0, \sin a), \mathbf{u}_2 = (0,1,0), \mathbf{u}_3 = (-\sin a, 0, \cos a), \tag{6.100}$$

the first of which (say) gives the actual direction of the axis of polarization. In any case, Alice writes down $F = 1$ if her photon passes her glass at angle a, and $F = 0$ if it does not; similarly Bob writes $G = 1$ (pass) or $G = 0$ (fail) at setting b.

In a quantum-mechanical description of the experiment, the Hilbert space of the photon pair is $\mathbb{C}^2 \otimes \mathbb{C}^2$, and the correlated photon state is taken to be

$$\psi_0 = (\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2)/\sqrt{2}, \tag{6.101}$$

where $\mathbf{e}_1 = (1,0)$ and $\mathbf{e}_2 = (0,1)$ form the standard basis of $\mathbb{C}^2$. The probabilities (6.35) - (6.38) as predicted by quantum mechanics are now replaced by

$$P_{\psi_0}(F=1,G=1|A=a,B=b) = \tfrac{1}{2}\cos^2(a-b); \tag{6.102}$$
$$P_{\psi_0}(F=0,G=0|A=a,B=b) = \tfrac{1}{2}\cos^2(a-b); \tag{6.103}$$
$$P_{\psi_0}(F=1,G=0|A=a,B=b) = \tfrac{1}{2}\sin^2(a-b), \tag{6.104}$$
$$P_{\psi_0}((F=0,G=1|A=a,B=b) = \tfrac{1}{2}\sin^2(a-b), \tag{6.105}$$

which are also the experimentally measured ones. Instead of (6.90) we then obtain

$$P_{\psi_0}(F \neq G|A=a,B=b) = \sin^2(a-b); \tag{6.106}$$
$$P_{\psi_0}(F = G|A=a,B=b) = \cos^2(a-b). \tag{6.107}$$

In particular, if their settings are the same (i.e., $a=b$), then Alice and Bob will always find the same outcome (***perfect correlation***), whereas in case they are orthogonal (i.e., $a=b\pm\pi/2$), they obtain ***perfect anti-correlation***, in that Alice's photon passes whenever Bob's is blocked, and *vice versa*. However, this will not be used. Although it should be obvious from the previous case what the assumptions in Theorem 6.15 mean for this particular experiment, we make them explicit:

- ***Determinism*** means that there is a probability space (X,Σ,μ) with associated (measurable) functions

$$A : X \to [0,\pi], B : X \to [0,\pi], F : X \to \{0,1\}, G : X \to \{0,1\}, \tag{6.108}$$

  which completely describe the experiment in the sense that $x \in X$ determines *both* its settings $a = A(x), b = B(x)$ *and* its outcomes $\lambda = F(x), \gamma = G(x)$.
- ***Freedom*** stipulates that there is a (measurable) function $Z : X \to X_Z$ such that:
  - $F = F(A,B,Z)$ and $G = G(A,B,Z)$;
  - (A,B,Z) are probabilistically independent relative to μ.
- ***Locality*** means that $F(A,B,Z) = F(A,Z)$ and $G(A,B,Z) = G(B,Z)$.
- ***Nature*** states that the empirical as well as theoretical probabilities (6.106) for the experiment are reproduced as conditional joint probabilities given by μ through

$$P(F \neq G|A=a,B=b) = \sin^2(a-b). \tag{6.109}$$

Theorem 6.15 then holds *verbatim* for this situation, with the following proof.

Proof. *Determinism* and *Freedom* imply

$$P(F=\lambda,G=\gamma|A=a,B=b) = P_{ABZ}(\hat{F}=\lambda,\hat{G}=\gamma|\hat{A}=a,\hat{B}=b), \tag{6.110}$$

where we use the notation (6.50) - (6.51), the function $\hat{A} : X_A \times X_B \times X_Z \to X_A$ is projection on the first coordinate, likewise the function $\hat{B} : X_A \times X_B \times X_Z \to X_B$ is projection on the second, and P_{ABZ} is the joint probability on $X_A \times X_B \times X_Z$ induced by the triple (A,B,Z) and the probability measure μ; by independence, P_{ABZ} is a product measure on $X_A \times X_B \times X_Z$. According to *Locality*, $\hat{F}(a,b,z)$ does not depend on b, whilst $\hat{G}(a,b,z)$ does not depend on a.

For fixed settings (a,b), we may therefore define the following functions on X_Z:

$$\hat{F}_a(z) = \hat{F}(a,z); \tag{6.111}$$
$$\hat{G}_b(z) = \hat{G}(b,z). \tag{6.112}$$

A brief computation then yields

$$P_{ABZ}(\hat{F} = \lambda, \hat{G} = \gamma | \hat{A} = a, \hat{B} = b) = P_Z(\hat{F}_a = \lambda, \hat{G}_b = \gamma), \tag{6.113}$$

where P_Z is the joint probability on X_Z defined by Z and μ. Therefore, from (6.110),

$$P(F = \lambda, G = \gamma | A = a, B = b) = P_Z(\hat{F}_a = \lambda, \hat{G}_b = \gamma). \tag{6.114}$$

Nature then gives the crucial result

$$P_Z(\hat{F}_a \neq \hat{G}_b) = \sin^2(a-b). \tag{6.115}$$

Lemma 6.16. *Any four* $\{0,1\}$*-valued random variables* (F_1, F_2, G_1, G_2) *satisfy*

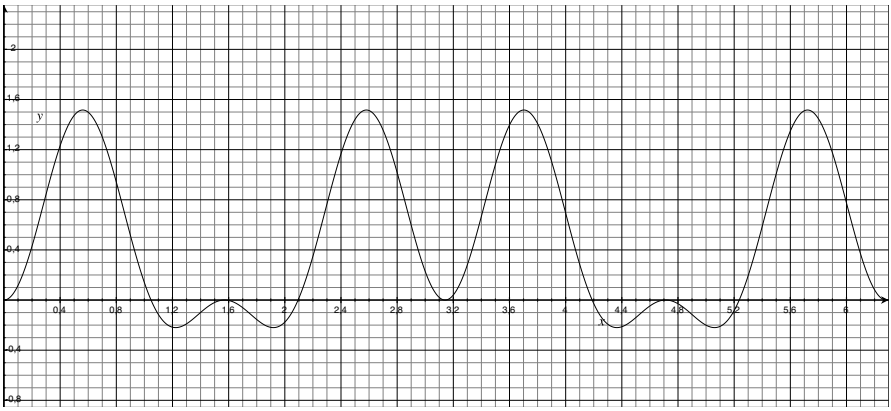
$$P(F_1 \neq G_1) \leq P(F_1 \neq G_2) + P(F_2 \neq G_1) + P(F_2 \neq G_2). \tag{6.116}$$

This lemma (said to go back to Boole) is very easy to prove directly, but for completeness's sake we mention that it also follows from Proposition 6.17 below.

Taking $F_1 = \hat{F}_{a_1}$, $F_2 = \hat{F}_{a_2}$, $G_1 = \hat{G}_{b_1}$, $G_2 = \hat{G}_{b_2}$, and $P = P_Z$, for suitable values of (a_1, a_2, b_1, b_2) this inequality is violated by (6.115). Take, for example, $a_2 = b_2 = 3x$, $a_1 = 0$, and $b_1 = x$. The inequality (6.116) then assumes the form $f(x) \geq 0$ for

$$f(x) = \sin^2(3x) + \sin^2(2x) - \sin^2(x).$$

But in fact, $f(x) < 0$ for continuously many values of $x \in [0, 2\pi]$, see plot. □



Graph of $x \mapsto \sin^2(3x) + \sin^2(2x) - \sin^2(x)$*, showing (in the region where it is negative) that quantum mechanics violates the Bell inequality* (6.116).

Lemma 6.16 is a special case of a more general result.

Proposition 6.17. *Let $F_i : X \to [-1,1]$ and $G_j : X \to [-1,1]$, where (X,Σ,μ) is some probability space, be two parametrized random variables, $i,j = 1,2$. Then the* **two-point function** $\langle F_i G_j \rangle = \int_X d\mu\, F_i G_j$ *satisfies the* CHSH**-inequality**

$$|\langle F_1 G_1 \rangle + \langle F_1 G_2 \rangle + \langle F_2 G_1 \rangle - \langle F_2 G_2 \rangle| \leq 2. \tag{6.117}$$

If F_i and G_j just take the values ± 1, then (6.116) is a special case of (6.117).

Proof. In terms of the function $\Phi = F_1 \cdot (G_1 + G_2) + F_2 \cdot (G_1 - G_2)$, we may write

$$\langle F_1 G_1 \rangle + \langle F_1 G_2 \rangle + \langle F_2 G_1 \rangle - \langle F_2 G_2 \rangle = \int_X d\mu\, \Phi. \tag{6.118}$$

Since $|F_i(x)| \leq 1$ and $|G_j(x)| \leq 1$ by assumption, we have $|\Phi(x)| \leq 2$ and hence

$$\left| \int_X d\mu(x)\, \Phi(x) \right| \leq \int_X d\mu(x)\, |\Phi(x)| \leq 2, \tag{6.119}$$

since μ is a probability measure. To prove the the last claim, we just note that

$$\begin{aligned} &P(F_i = G_j) - P(F_i \neq G_j) = \langle F_i G_j \rangle; \\ &P(F_i = G_j) + P(F_i \neq G_j) = 1. \end{aligned}$$

□

In Bell's second (1976) theorem on ***stochastic hidden variables***, the assumption of *Determinism* is dropped, and all we have is a theory stating conditional probabilities $P(F = \lambda, G = \gamma | A = a, B = b, x)$ for the outcomes of the above bipartite experiment given some hidden variable x, as well as the single-wing versions $P(F = \lambda | A = a)$ and $P(G = \gamma | B = b, x)$. Here F, G, A, B are just notational devices to record such outcomes, *which are no longer (necessarily) represented as random variables*. On this new understanding of the notation, the ***Nature*** assumption is formulated just as before, cf. (6.109). We do assume the existence of a probability space (X,Σ,μ) and of conditional probabilities

$$P(F = \lambda, G = \gamma | A = a, B = b, x), \;\; P(F = \lambda | A = a, x), \;\; P(G = \gamma | B = b, x),$$

defined μ-a.e. in x, in which the state of the world is specified as being $x \in X$. In terms of this space, the ***Freedom*** assumption means that

$$P(F = \lambda, G = \gamma | A = a, B = b) = \int_X d\mu(x)\, P(F = \lambda, G = \gamma | A = a, B = b, x), \tag{6.120}$$

for any settings (a,b), of which μ is independent (as the notation already indicated).

The crucial assumption replacing *Determinism* is ***Bell locality***, which reads

$$P(F = \lambda, G = \gamma | A = a, B = b, x) = P(F = \lambda | A = a, x) \cdot P(G = \gamma | B = b, x). \tag{6.121}$$

Bell's second theorem for stochastic hidden variable theories reads as follows.

Theorem 6.18. *Nature, Freedom, and Bell locality are contradictory.*

Proof. The idea of the proof is to introduce an artificial probability space in order to recover the framework of Theorem 6.15. To this end, we take

$$\tilde{X} = [0,1] \times [0,1] \times X; \tag{6.122}$$

$$d\tilde{\mu}(s,t,x) = ds \cdot dt \cdot d\mu(x). \tag{6.123}$$

where we denoted the elements of $\tilde{X}$ by (s,t,x). On $\tilde{X}$, define random variables

$$\tilde{F}_a(s,t,x) = 1_{[0,P(F=1|A=a,x)]}(s); \tag{6.124}$$

$$\tilde{G}_b(s,t,x) = 1_{[0,P(G=1|B=b,x)]}(t), \tag{6.125}$$

where 1_Δ is the indicator function for $\Delta \subseteq [0,1]$. Writing, as usual,

$$\tilde{P}(\tilde{F}_a = \lambda, \tilde{G}_b = \gamma) = \int_{\tilde{X}} d\tilde{\mu}(s,t,x)\,\{(s,t,x) \in \tilde{X} \mid \tilde{F}_a(s,t,x) = \lambda, \tilde{G}_b(s,t,x) = \gamma\},$$

we obtain (first for $\lambda = \gamma = 1$, from which the other cases follow):

$$\tilde{P}(\tilde{F}_a = \lambda, \tilde{G}_b = \gamma) = \int_X d\mu(x)\, P(F = \lambda | A = a, x) \cdot P(G = \gamma | B = b, x). \tag{6.126}$$

With *Freedom* and *Bell locality*, this yields

$$P(F = \lambda, G = \gamma | A = a, B = b) = \tilde{P}(\tilde{F}_a = \lambda, \tilde{G}_b = \gamma), \tag{6.127}$$

as in (6.114), so that the proof may be completed as for Theorem 6.15. □

Let us note that since in Bell's second theorem the settings (a,b) are treated as free parameters to begin with, the difference between X and Z evaporates, so that in the above formulae one might as well have replaced (X,μ) by the space (X_Z, μ_Z) that describes all relevant degrees of freedom *except the settings* (i.e., the experimentalist, in either human or machine form). Either way, Bell's locality condition may be disentangled into the following conditions (introduced by Jarrett and Shimony):

1. ***Parameter Independence*** (PI):

$$P(\lambda | a,b,x) = P(\lambda | a,x); \tag{6.128}$$

$$P(\gamma | a,b,x) = P(\gamma | b,x); \tag{6.129}$$

2. ***Outcome Independence*** (OI):

$$P(\lambda | a,b,\gamma,x) = P(\lambda | a,b,x); \tag{6.130}$$

$$P(\gamma | a,b,\lambda,x) = P(\gamma | a,b,x), \tag{6.131}$$

where we have abbreviated $P(F = \lambda | A = a, B = b, x)$ by $P(\lambda | a,b,x)$, etc., and have used the following notation (which states identities in case one has (6.91) - (6.93)):

$$P(\lambda|a,b,x) \equiv \sum_{\gamma} P(\lambda,\gamma|a,b,x); \tag{6.132}$$

$$P(\gamma|a,b,x) \equiv \sum_{\lambda} P(\lambda,\gamma|a,b,x); \tag{6.133}$$

$$P(\lambda|a,b,\gamma,x) \equiv \frac{P(\lambda,\gamma|a,b,x)}{P(\gamma|a,b,x)}; \tag{6.134}$$

$$P(\gamma|a,b,\lambda,x) \equiv \frac{P(\lambda,\gamma|a,b,x)}{P(\lambda|a,b,x)}, \tag{6.135}$$

It is easy to see that *Bell locality is equivalent to the conjunction of* PI *and* OI.

Note that the former (PI), akin to *Locality*, is a hidden or 'subsurface' version of the ***no signaling*** property of the 'surface' probabilities, which states that

$$P(\lambda|a,b) \equiv \sum_{\gamma} P(\lambda,\gamma|a,b)$$

is independent of b (and *vice versa*). But a violation of PI only leads to signaling if x can be operationally controlled, similar to the way in which experimental physicists prepare quantum states ψ. Hence it is reassuring that quantum mechanics satisfies PI if we see the quantum state ψ as a hidden variable: assuming

$$P(\lambda,\gamma|a,b,x) = P_{\psi_0}(F=\lambda, G=\gamma|A=a, B=b), \tag{6.136}$$

as computed in (6.102) - (6.105), PI is valid but OI is not. First, for $\lambda = 0$ or $\lambda = 1$,

$$P(\lambda|a,b,x) = \sum_{\gamma=0,1} P_{\psi_0}(F=\lambda, G=\gamma|a,b) = \tfrac{1}{2}\cos^2(a-b) + \tfrac{1}{2}\sin^2(a-b) = \tfrac{1}{2}, \tag{6.137}$$

which is independent of b, and likewise $P(\gamma|a,b,x) = \frac{1}{2}$, independently of a. This yields PI, which a similar computation shows to be true for any quantum state. On the other hand, given this result, OI would require

$$P_{\psi_0}(F=\lambda, G=\gamma|A=a, B=b) = P_{\psi_0}(F=\lambda|A=a)\cdot P_{\psi_0}(G=\gamma|B=b),$$

which is false, since by (6.102) - (6.105), Alice's and Bob's outcomes are correlated.

Hence Bell locality is violated by quantum mechanics, but this does not imply that "quantum mechanics is nonlocal" (as some say). Bell's is a very specific locality condition invented as a constraint on hidden variable theories. In another important sense, viz. *Einstein locality*, quantum mechanics *is* local, in that observables with spacelike separated localization regions commute (this is the case in quantum field theory, but also in any bipartite experiment of the type considered here, where Alice's operators commute with Bob's just by definition of the tensor product).

On the other hand, deterministic theories, which in the present context are defined as those for which all conditional probabilities like $P(\lambda,\gamma|a,b,x)$ are either zero or one (in which case one may introduce random variables reproducing these probabilities), violate PI but satisfy OI, at least if they reproduce the Born probabilities (such as Bohmian mechanics). Hence such theories violate Bell locality.

Finally, Bell-type inequalities like (6.117) also give information about quantum mechanics itself, particularly about the degree of entanglement of states. Let H_1 and H_2 be Hilbert spaces, with tensor product $H_1 \otimes H_2$. A unit vector $\psi \in H_1 \otimes H_2$ is called ***uncorrelated*** if it is of the form $\psi = \varphi_1 \otimes \varphi_2$, where $\varphi_k \in H_k$ are unit vectors, $k = 1,2$, and ***correlated*** otherwise. Clearly, the vectors (6.34) and (6.101) used in the experiments so far are correlated. The simplest result is then as follows.

Theorem 6.19. *Let a_1 and a_2 be self-adjoint operators on H_1, and let b_1 and b_2 be self-adjoint operators on H_2, each with spectrum contained in $[-1,1]$ (equivalently $\|X_a\| \leq 1$, etc.). Let ψ be a unit vector in $H_1 \otimes H_2$, and define two-point functions*

$$\langle F_i G_j \rangle = \langle \psi, a_i \otimes b_j \psi \rangle. \tag{6.138}$$

If ψ is uncorrelated, then the Bell inequality (6.117) *holds.*

Proof. This follows from the factorization property

$$\langle F_i G_j \rangle = \langle \varphi_1 \otimes \varphi_2, a_i \otimes b_j \varphi_1 \otimes \varphi_2 \rangle = \langle \varphi_1, a_i \varphi_1 \rangle \cdot \langle \varphi_2, b_j \varphi_2 \rangle = \langle F_i \rangle \cdot \langle G_j \rangle, \tag{6.139}$$

where $\langle F_i \rangle = \langle \varphi_1, a_i \varphi_1 \rangle$ and $\langle G_j \rangle = \langle \varphi_2, b_j \varphi_2 \rangle$. For either sign, this property yields

$$\langle F_2(G_1 - G_2) \rangle = \langle F_2 \rangle \langle G_1 \rangle (1 \pm \langle F_1 \rangle \langle G_2 \rangle) - \langle F_2 \rangle \langle G_2 \rangle (1 \pm \langle F_1 \rangle \langle G_1 \rangle). \tag{6.140}$$

The spectral assumption implies that $|\langle F_i \rangle| \leq 1$ and $|\langle G_j \rangle| \leq 1$, which will be used directly below, as well as its consequence $|1 \pm \langle F_1 \rangle \langle G_2 \rangle| = 1 \pm \langle F_i \rangle \langle G_j \rangle$. Hence

$$\begin{aligned} |\langle F_2(G_1 - G_2) \rangle| &\leq |1 \pm \langle F_1 \rangle \langle G_2 \rangle| + |1 \pm \langle F_1 \rangle \langle G_1 \rangle| \\ &= 1 \pm \langle F_1 \rangle \langle G_2 \rangle + 1 \pm \langle F_1 \rangle \langle G_1 \rangle \\ &= 2 \pm \langle F_1(G_1 + G_2) \rangle. \end{aligned} \tag{6.141}$$

Similarly,

$$|\langle F_1(G_1 + G_2) \rangle| \leq 2 \pm \langle F_2(G_1 - G_2) \rangle, \tag{6.142}$$

so that, writing $\Phi = \langle F_1 G_1 \rangle + \langle F_1 G_2 \rangle + \langle F_2 G_1 \rangle - \langle F_2 G_2 \rangle$, for either sign $\pm$ we have

$$|\Phi| \leq |\langle F_1(G_1 + G_2) \rangle| + |\langle F_2(G_1 - G_2) \rangle| \leq 4 \pm \Phi \tag{6.143}$$

If $\Phi \geq 0$ we choose the minus sign, whereas for $\Phi < 0$ we take the plus sign. Either way, we obtain $|\Phi| \leq 2$, which is the inequality (6.117). □

This result is actually much more general (as hinted at by the way that the proof only uses the uncorrelated vector state $\psi = \varphi_1 \otimes \varphi_2$). The simplest generalization is to replace pure states by mixed states, where we say that a density operator ρ on $H_1 \otimes H_2$ is *uncorrelated* if it is of the form $\rho = \sum_i p_i \rho_1 \otimes \rho_2$, where the p_i are probabilities and ρ_k is a density matrix on H_k, $k = 1,2$. Then all uncorrelated density matrices satisfy the inequality (6.117). Even more generally, uncorrelated states on C*-algebras or von Neumann algebras $A \otimes B$ satisfy (6.117), see Notes.

6.6 The Colbeck–Renner Theorem

One may try to strengthen Bell's second theorem by weakening its assumptions. A remarkable result in this direction states that, roughly speaking, any probabilistic hidden variable theory that satisfies *Freedom* and *Parameter Independence* and is compatible with quantum mechanics adds nothing to quantum mechanics. In other words, it appears that quantum mechanics "cannot be extended", or "is complete".

In fact, the result turns out to be more modest than this summary suggests, since the reasoning required to prove the claim hinges on certain assumptions which are satisfied by quantum mechanics itself, but might seem unnatural for a hidden variable theory. In any case, we have to state our notation and assumptions very clearly.

Definition 6.20. *A* **hidden variable theory** $\mathscr{T}$ *underlying quantum mechanics consists of a measurable space* (X,Σ) *whose points* x *label conditional probabilities*

$$P(a_1 = \lambda_1, \ldots, a_n = \lambda_n | x) \equiv P(\mathbf{a} = \lambda | x)$$

for the possible outcomes $\lambda = (\lambda_1, \ldots, \lambda_n)$ *of a measurement of any family* $\mathbf{a} = (a_1, \ldots, a_n)$ *of* n *commuting self-adjoint operators on any Hilbert space* H.

These formal conditional probabilities are a priori *only supposed to satisfy*

$$0 \leq P(\mathbf{a} = \lambda | x) \leq 1; \tag{6.144}$$

$$\sum_{\lambda} P(\mathbf{a} = \lambda | x) = 1. \tag{6.145}$$

Apart from these probabilities, for each Hilbert space H *and any pure state* $e \in \mathscr{P}_1(H)$*, the theory* $\mathscr{T}$ *yields a classical state* μ_e*, i.e., a probability measure on* X.

As the notation indicates, μ_e depends on e only and hence is independent of a and λ. From the point of view of $\mathscr{T}$, a quantum state *is* a probability measure on X! In what follows we assume for simplicity that H is finite-dimensional, so that $e = e_\psi$ for some unit vector $\psi \in H$. With slight abuse of notation we then write μ_ψ for μ_{e_ψ}.

An important special case will be the bipartite setting $H = H_1 \otimes H_2$, where Alice and Bob measure self-adjoint operators X and Y on H_1 and H_2, respectively, so that

$$n = 2, \;\; a_1 = X \otimes 1_{H_2}, \;\; a_2 = 1_{H_1} \otimes Y.$$

We then introduce settings $c = (a,b)$, as in the previous sections, so that we typically look at expressions like $P(X_a = \lambda_1, Y_b = \lambda_2 | x)$. The other case of interest will simply be $n = 1$ with $a_1 \equiv a$, $\lambda_1 \equiv \lambda$; indeed, this will be the case in the statement of the theorem (the bipartite case playing a role only in the proof, though a crucial one!).

The following notation will be quite important to the argument. An equality

$$P_\psi(\mathbf{a} = \lambda | x) = \alpha(x), \tag{6.146}$$

where $\alpha : X \to [0,1]$ is measurable (often even constant), abbreviates:

$$P(\mathbf{a} = \lambda | x) = \alpha(x) \text{ for almost every } x \text{ with respect to the measure } \mu_\psi.$$

That is, there is a subset $X' \subset X$ such that $\mu_\psi(X') = 0$ and $P_\psi(\mathbf{a} = \lambda|x) = \alpha(x)$ holds for any $x \in X \backslash X'$. If X is finite, this simply means that the equality holds for any x for which $\mu_\psi(\{x\}) > 0$. Since this notation may render equalities like

$$P_\psi(\mathbf{a} = \lambda|x) = P_\varphi(\mathbf{a}' = \lambda'|x), \tag{6.147}$$

ambiguous, we explicitly define (6.147) as the double implication

$$P_\psi(\mathbf{a} = \lambda|x) = \alpha(x) \Leftrightarrow P_\varphi(\mathbf{a}' = \lambda'|x) = \alpha(x).$$

Furthermore, for $\varepsilon \to 0$ we write

$$P_\psi(\mathbf{a} = \lambda|x) \overset{\varepsilon}{\approx} P_\varphi(\mathbf{a}' = \lambda'|x) \Leftrightarrow P_\psi(\mathbf{a} = \lambda|x) = P_\varphi(\mathbf{a}' = \lambda'|x) + O(\sqrt{\varepsilon}), \tag{6.148}$$

as well as

$$\psi \overset{\varepsilon}{\approx} \varphi \Leftrightarrow (1-\varepsilon) \leq |\langle \psi, \varphi \rangle| \leq 1. \tag{6.149}$$

We are now ready to state our assumptions for the Colbeck–Renner Theorem:

- ***Compatibility with Quantum Mechanics (CQ):*** for any unit vector $\psi \in H$,

$$\int_X d\mu_\psi(x)\, P(\mathbf{a} = \lambda|x) = p_\psi(\mathbf{a} = \lambda), \tag{6.150}$$

  where the quantum-mechanical prediction $p_\psi(\mathbf{a} = \lambda)$ is given by the Born rule

$$p_\psi(\mathbf{a} = \lambda) = \langle \psi, e^{(1)}_{\lambda_1} \cdots e^{(n)}_{\lambda_n} \psi \rangle, \tag{6.151}$$

  cf. (2.21), where $e^{(i)}_{\lambda_i}$ is the spectral projection on the eigenspace $H_{\lambda_i} \subset H$ of a_i.
- ***Unitary Invariance (UI):*** for any unit vector $\psi \in H$ and unitary u on H,

$$P_{u\psi}(\mathbf{a} = \lambda|x) = P_\psi(u^{-1}\mathbf{a}u = \lambda|x). \tag{6.152}$$

- ***Continuity of Probabilities (CP:*** If $\psi \overset{\varepsilon}{\approx} \varphi$, then $P_\psi(\mathbf{a} = \lambda|x) \overset{\varepsilon}{\approx} P_\varphi(\mathbf{a} = \lambda|x)$.

In the remaining axioms, $H = H_1 \otimes H_2$, and a and b are self-adjoint operators on H_1 and H_2, respectively (duly identified with operators $a \otimes 1_{H_2}$ and $1_{H_1} \otimes b$ on H).

- ***Parameter Independence (PI):***

$$\sum_{\gamma \in \sigma(b)} P(a = \lambda, b = \gamma|x) = P(a = \lambda|x); \tag{6.153}$$

$$\sum_{\lambda \in \sigma(a)} P(a = \lambda, b = \gamma|x) = P(b = \gamma|x). \tag{6.154}$$

- ***Product Extension (PE):*** for any pair of states $\psi_1 \in H_1$, $\psi_2 \in H_2$,

$$P_{\psi_1}(a = \lambda|x) = P_{\psi_1 \otimes \psi_2}(a = \lambda|x). \tag{6.155}$$

- ***Schmidt Extension (SE):*** if $\upsilon_i \in H_1$ $(i = 1, \ldots, \dim(H))$ are eigenstates of a, then for arbitrary orthogonal states $u_i \in H_2$ and coefficients $c_i > 0$ with $\sum_i c_i^2 = 1$,

$$P_{\sum_i c_i \cdot \upsilon_i}(a = x|x) = P_{\sum_i c_i \cdot \upsilon_i \otimes u_i}(a = x|x). \tag{6.156}$$

Note that **PI** makes sense, because (6.151) and (6.150) imply that for $p_\psi(\mathbf{a} = \lambda)$ to be nonzero we must have $\lambda_i \in \sigma(a_i)$ for each i. All assumptions are satisfied by quantum mechanics itself (seen as a hidden variable theory with ψ as the "hidden" variable x). In the context of hidden variable theories, though, one might doubt the plausibility of ***UI***, ***CP***, and ***SE***. But we need all these assumptions to prove:

Theorem 6.21. *If $\mathscr{T}$ satisfies* **CQ**, **UI**, **CP**, **PI**, **PE**, *and* **SE**, *then for any (finite-dimensional) Hilbert space H, unit vector $\psi \in H$, and operator $a \in B(H)_{\mathrm{sa}}$,*

$$P_\psi(a = \lambda|x) = p_\psi(a = \lambda). \tag{6.157}$$

In other words, the hidden variable x is even more hidden than expected, since knowing its value has no effect on the probabilities for the outcomes of experiments.

Proof. We first assume (without loss of generality) that a is nondegenerate as a self-adjoint matrix, in that it has distinct eigenvalues $(\lambda_1, \ldots, \lambda_{\dim(H)})$; this assumption will be removed at the end of the proof. The proof consists of three steps.

1. The theorem holds for $H = \mathbb{C}^2$ and any pair (a, ψ) for which

$$p_\psi(a = \lambda_1) = p_\psi(a = \lambda_2) = 1/2, \tag{6.158}$$

   This only requires assumptions ***CQ***, ***PI***, and ***SE***.
2. The theorem holds for $H = \mathbb{C}^l$, $l < \infty$ arbitrary, and any pair (a, ψ) for which

$$p_\psi(a = \lambda_1) = \cdots = p_\psi(a = \lambda_l) = 1/l. \tag{6.159}$$

   This is just a slight extension of step 1 and uses the same three assumptions.
3. The theorem holds in general. This requires all assumptions (as well as step 2).

Proof of step 1. Let $H = \mathbb{C}^2$, with basis (υ_1, υ_2) of eigenvectors of a, so that $\psi \in \mathbb{C}^2$ may be written as

$$\psi = (\upsilon_1 + \upsilon_2)/\sqrt{2}. \tag{6.160}$$

Without loss of generality, we may assume that $\lambda_1 = 1$ and $\lambda_2 = -1$. We now relabel a as a_0 and extend it to a family of operators $(a_k)_{k=0,1,\ldots,2N-1}$ by fixing an integer $N > 1$, putting $\theta_k = k\pi/2N$, and defining

$$c_k = e_{\theta_{k+\pi}} - e_{\theta_k}, \tag{6.161}$$

where, for any angle $\theta \in [0, 2\pi]$, the operator $e_\theta = |\theta\rangle\langle\theta|$ is the orthogonal projection onto the one-dimensional subspace spanned by the unit vector

$$|\theta\rangle = \sin(\theta/2) \cdot \upsilon_1 + \cos(\theta/2) \cdot \upsilon_2. \tag{6.162}$$

In the bipartite setting, we have operators $a_k = c_k \otimes 1_2$ and $b_k = 1_2 \otimes c_k$ on $\mathbb{C}^2 \otimes \mathbb{C}^2$, as well as a maximally correlated (Bell) state $\psi_{AB} \in \mathbb{C}^2 \otimes \mathbb{C}^2$, given by

$$\psi_{AB} = \frac{1}{\sqrt{2}}(\upsilon_1 \otimes \upsilon_1 + \upsilon_2 \otimes \upsilon_2). \tag{6.163}$$

Using assumptions ***PI*** and ***SE***, we then have, for $i = 1,2$ $\lambda_1 = 1$, and $\lambda_2 = -1$,

$$P_\psi(a = \lambda_i|x) = P_{\psi_{AB}}(a_0 = \lambda_i|x). \tag{6.164}$$

The quantum-mechanical prediction is

$$p_{\psi_{AB}}(a_0 = 1) = p_{\psi_{AB}}(a_0 = -1) = \tfrac{1}{2}. \tag{6.165}$$

Our goal is to show that also for each $x \in X$, knowing x is irrelevant in that

$$P_{\psi_{AB}}(a_0 = 1|x) = P_{\psi_{AB}}(a_0 = -1|x) = \tfrac{1}{2}. \tag{6.166}$$

To this effect we introduce the combination of probabilities

$$I^{(N)}(x) = P(a_0 = b_{2N-1}|x) + \sum_{k \in K_N, l \in L_N, |k-l|=1} P(a_k \neq b_l|x), \tag{6.167}$$

where $K_N = \{0, 2, \ldots, 2N-2\}$ and $L_N = \{1, 3, \ldots, 2N-1\}$. Our first inequality is

$$\begin{aligned}
|P(a_k = \lambda_i|x) - P(b_l = \lambda_i|x)| &= |P(a_k = \lambda_i, b_l = \lambda_i|x) + P(a_k = \lambda_i, b_l \neq \lambda_i|x) \\
&\quad - P(a_k = \lambda_i, b_l = \lambda_i|x) + P(a_k \neq \lambda_i, b_l = \lambda_i|x)| \\
&= |P(a_k = \lambda_i, b_l \neq \lambda_i|x) - P(a_k \neq \lambda_i, b_l = \lambda_i|x)| \\
&\leq P(a_k = \lambda_i, b_l \neq \lambda_i|x) + P(a_k \neq \lambda_i, b_l = \lambda_i|x) \\
&= P(a_k \neq b_l|x),
\end{aligned} \tag{6.168}$$

where $i = 1,2$, and we used ***PI***. This implies a second inequality: since $a_{2N} = -a_0$,

$$\begin{aligned}
|P(a_0 = 1|x) - P(a_0 = -1|x)| &= |P(a_0 = 1|x) - P(a_{2N} = 1|x)| \\
&\leq \sum_{k,l,|k-l|=1} |P(a_k = 1|x) - P(b_l = 1|x)| \\
&\leq \sum_{k,l,|k-l|=1} P(a_k \neq b_l|x) \leq I^{(N)}(x).
\end{aligned}$$

Integrating this with respect to the measure $\mu_{\psi_{AB}}$ and using ***CQ*** gives

$$\int_X d\mu_{\psi_{AB}}(x)\, |P(a_0 = 1|x) - P(a_0 = -1|x)| \leq \int_X d\mu_{\psi_{AB}}(x)\, I^{(N)}(x) = I^{(N)}_{\psi_{AB}}. \tag{6.169}$$

We wish to invoke the corresponding quantum-mechanical expression, defined by

$$I^{(N)}_{\psi_{AB}} = p_{\psi_{AB}}(a_0 = b_{2N-1}) + \sum_{k\in K_N, l\in L_N, |k-l|=1} p_{\psi_{AB}}(a_k \neq b_l). \tag{6.170}$$

A straightforward calculation shows that this expression is equal to

$$I^{(N)}_{\psi_{AB}} = 2N\sin^2(\pi/4N). \tag{6.171}$$

Since $\lim_{N\to\infty} I^{(N)}_{\psi_{AB}} = 0$, letting $N \to \infty$ in (6.169) therefore yields (6.166). From (6.164) we then obtain (6.158).

Proof of step 2. Let $H = \mathbb{C}^l$ and let $(\upsilon_i)_{i=1}^l$ be an orthonormal basis of eigenvectors of a, with corresponding eigenvalues λ_i, and phase factors for the eigenvectors υ_i such that $c_i > 0$ (and of course, $\sum_i c_i^2 = 1$) in the expansion

$$\psi = \sum_i c_i \upsilon_i. \tag{6.172}$$

The case of interest will be $c_1 = \cdots = c_l = 1/l$, but first we merely assume that $c_1 = c_2$ (the same reasoning applies to any other pair), with $\lambda_1 = 1$ and $\lambda_2 = -1$ (which involves no loss of generality either and just simplifies the notation). The other positive coefficients c_i are arbitrary. Generalizing (6.166), we will show that

$$P_\psi(a = 1|x) = P_\psi(a = -1|x). \tag{6.173}$$

This shows that if two Born probabilities defined by some quantum state e_ψ are equal, then the underlying hidden variable probabilities must be equal μ_ψ-a.e., too. Eq. (6.159) immediately follows from this result by taking all c_i to be equal.

As in step 1, we pass to the bipartite setting, introducing two copies of $H = \mathbb{C}^l$ denoted by $H_A = H_B = \mathbb{C}^l$, and define the correlated state

$$\psi_{AB} = \sum_i c_i \cdot \upsilon_i \otimes \upsilon_i \tag{6.174}$$

in $H_A \otimes H_B$. Eq. (6.164) again follows from assumptions ***PI*** and ***SE***. Throughout the argument of step 1, we now replace each probability $P(a_k = \lambda_i, b_l = \gamma_j|x)$ by an adapted probability $P^{(1)}(a_k = \lambda_i, b_l = \gamma_j|x)$, defined as the conditional probability

$$\begin{aligned} P^{(1)}(a_k = \lambda_i, b_l = \gamma_2|x) &= P(a_k = \lambda_i, b_l = \gamma_2||\lambda_i| = |\gamma_2| = 1, x) \\ &= \frac{P(a_k = \lambda_i, b_l = \gamma_2, |\lambda_i| = |\gamma_2| = 1|x)}{P(|\lambda_i| = |\gamma_2| = 1|x)}, \end{aligned} \tag{6.175}$$

for all x for which $P(|\lambda_i| = |\gamma_2| = 1|x) > 0$, whereas

$$P^{(1)}(a_k = \lambda_i, b_l = \gamma_2|x) = 0 \tag{6.176}$$

whenever $P(|\lambda_i| = |\gamma_2| = 1|x) = 0$. The same argument then yields (6.169), with P replaced by $P^{(1)}$ but with the same right-hand side. As in step 1,

$$P^{(1)}_{\psi_{AB}}(a_0 = 1|x) = P^{(1)}_{\psi_{AB}}(a_0 = -1|x), \tag{6.177}$$

which implies that

$$P_{\psi_{AB}}(a_0 = 1|x) = P_{\psi_{AB}}(a_0 = -1|x), \tag{6.178}$$

either because both sides vanish (if $P(|\lambda_i| = |\gamma_2| = 1|x) = 0$), or because (in the opposite case) the denominator $P(|\lambda_i| = |\gamma_2| = 1|x)$ cancels from both sides of (6.177).

Combined with (6.164), eq. (6.178) proves (6.173) and hence establishes step 2.

Proof of step 3. This is the most difficult step in the proof, relying on a technique wittily called ***embezzlement*** (which we only need for maximally entangled states). We will deal with three Hilbert spaces, namely $H = \mathbb{C}^l$, $H' = \mathbb{C}^m$, and $H'' = \mathbb{C}^n$ (where $n = m^N$ for some large N, see below), each with some fixed orthonormal basis $(\upsilon_i)_{i=1}^l$, $(\upsilon'_j)_{j=1}^m$, and $(\upsilon''_k)_{k=1}^n$, respectively. Given a further number $m_i \leq m$, we now list the nm basis vectors $\upsilon''_k \otimes \upsilon'_j$ of $H'' \otimes H'$ in two different orders:

1. $\upsilon''_1 \otimes \upsilon'_1, \ldots, \upsilon''_n \otimes \upsilon'_1, \upsilon''_1 \otimes \upsilon'_2, \ldots, \upsilon''_n \otimes \upsilon'_2, \ldots, \upsilon''_1 \otimes \upsilon'_m, \ldots, \upsilon''_n \otimes \upsilon'_m$;
2. $\upsilon''_1 \otimes \upsilon'_1, \ldots, \upsilon''_1 \otimes \upsilon'_{m_i}, \upsilon''_2 \otimes \upsilon'_1, \ldots, \upsilon''_2 \otimes \upsilon'_{m_i}, \ldots, \upsilon''_n \otimes \upsilon'_1, \ldots, \upsilon''_n \otimes \upsilon'_{m_i}, \ldots,$

where the remaining vectors (i.e., those of the form $\upsilon''_k \otimes \upsilon'_j$ for $1 \leq k \leq n$ and $j > m_i$) are listed in some arbitrary order.

Define

$$u^{(m_i)} : H'' \otimes H' \to H'' \otimes H' \tag{6.179}$$

as the unitary operator that maps the first list on the second. We will need the explicit expression

$$u^{(m_i)}(\upsilon''_k \otimes \upsilon'_1) = \upsilon''_{s^i_k} \otimes \upsilon'_{j^i_k}, \tag{6.180}$$

where for given $k = 1, \ldots, n$ the numbers $s^i_k = 1, \ldots, n_i$ (where n_i is the smallest integer such that $n_i m_i \geq n$) and $j^i_k = 1, \ldots, n_i$ are uniquely determined by

$$k = (s^i_k - 1)m_i + j^i_k. \tag{6.181}$$

We will actually work with two copies of $H'' \otimes H'$, called $H''_A \otimes H'_A$ and $H''_B \otimes H'_B$, with ensuing copies of $u^{(m_i)}_A$ and $u^{(m_i)}_B$ of $u^{(m_i)}$, and hence, leaving the isomorphism

$$H''_A \otimes H'_A \otimes H''_B \otimes H'_B \cong H''_A \otimes H''_B \otimes H'_A \otimes H'_B \tag{6.182}$$

implicit, we obtain a unitary operator

$$u^{(m_i)}_A \otimes u^{(m_i)}_B : H''_A \otimes H''_B \otimes H'_A \otimes H'_B \to H''_A \otimes H''_B \otimes H'_A \otimes H'_B. \tag{6.183}$$

The point of all this is that the unit vector

$$\kappa_n \in H''_A \otimes H''_A; \tag{6.184}$$

$$\kappa_n = \frac{1}{\sqrt{C(n)}} \sum_{k=1}^{n} \upsilon''_k \otimes \upsilon''_k, \tag{6.185}$$

where $C(n) = \sum_{k=1}^{n} 1/k$, acts as a "catalyst" in producing the maximally entangled state

$$\varphi \in H'_A \otimes H'_B; \tag{6.186}$$

$$\varphi = \frac{1}{\sqrt{m_i}} \sum_{j=1}^{m_i} \upsilon'_j \otimes \upsilon'_j, \tag{6.187}$$

from the uncorrelated state $\upsilon'_1 \otimes \upsilon'_1 \in H'_A \otimes H'_B$, in that for any $m_i \leq m$,

$$u_A^{(m_i)} \otimes u_B^{(m_i)} (\kappa_n \otimes \upsilon'_1 \otimes \upsilon'_1) \overset{\varepsilon/2}{\approx} \kappa_n \otimes \varphi. \tag{6.188}$$

Here $\varepsilon = 1/N$ if $n = m^{2N}$. This follows straightforwardly from (6.183) - (6.187).

After this preparation we are ready for the proof of step 3, continuing to use the notation established at the beginning of step 2, especially (6.172). As in step 1, we introduce two copies $H_A = H_B = \mathbb{C}^l$ of H, as well as two states

$$\psi_{AB} = \sum_i c_i \cdot \upsilon_i \otimes \upsilon_i \in H_A \otimes H_B; \tag{6.189}$$

$$\psi'''_{AB} = \kappa_n \otimes \upsilon'_1 \otimes \upsilon'_1 \otimes \psi_{AB} \in H'''_A \otimes H'''_B, \tag{6.190}$$

where κ_n is given by (6.185), we put

$$H''' = H'' \otimes H' \otimes H, \tag{6.191}$$

and in our notation we have ignored the obvious permutations of factors in the tensor product. For any $\varepsilon > 0$, pick $c'_i \in \mathbb{R}^+$ such that $(c'_i)^2 \in \mathbb{Q}^+$ and

$$|c'_i - c_i| < \varepsilon / \dim(H), \tag{6.192}$$

which implies that, in the sense of (6.149), we have

$$\sum_i c'_i \upsilon_i \overset{\varepsilon/2}{\approx} \sum_i c_i \upsilon_i. \tag{6.193}$$

Suppose

$$c'_i = \sqrt{p_i/q_i}, \tag{6.194}$$

with $p_i, q_i \in \mathbb{N}$ and $\gcd(p_i, q_i) = 1$, and define

$$m_i = p_i \prod_{i' \neq i} q_{i'}. \tag{6.195}$$

Consequently, writing

$$q = 1/\sqrt{\sum_{i'} m_{i'}}, \tag{6.196}$$

the following quotient is independent of i:

$$\frac{c_i'}{\sqrt{m_i}} = q. \tag{6.197}$$

Given the integers m_i thus obtained, we define a unitary operator

$$u : H''' \to H''' ; \tag{6.198}$$

$$u = \sum_{i=1}^{l} u^{(m_i)} \otimes |\upsilon_i\rangle\langle\upsilon_i|, \tag{6.199}$$

where $u^{(m_i)}$ is defined in (6.180). From this definition, with additional labels to denote the copies $u_A : H_A''' \to H_A'''$ and $u_B : H_B''' \to H_B'''$, and (6.188), and writing

$$\xi^{ij} = \upsilon_i \otimes \upsilon_j' \in H \otimes H', \tag{6.200}$$

with corresponding copies

$$\xi_{AA'}^{ij_i} \in H_A \otimes H_A'; \tag{6.201}$$

$$\xi_{BB'}^{ij_i} \in H_B \otimes H_B', \tag{6.202}$$

we then obtain the important relations

$$1_{H_A'''} \otimes 1_{H_B'''}(\psi_{AB}''') = \kappa_n \otimes \sum_{i=1}^{l} c_i \cdot \xi_{AA'}^{i1} \otimes \xi_{BB'}^{i1}; \tag{6.203}$$

$$u_A \otimes 1_{H_B'''}(\psi_{AB}''') = \frac{1}{\sqrt{C(n)}} \sum_{i=1}^{l} \sum_{k=1}^{n} \frac{c_i}{\sqrt{k}} \cdot \upsilon_{s_k}'' \otimes \upsilon_k'' \otimes \xi_{AA'}^{ij_k^i} \otimes \xi_{BB'}^{i1}; \tag{6.204}$$

$$1_{H_A'''} \otimes u_B(\psi_{AB}''') = \frac{1}{\sqrt{C(n)}} \sum_{i=1}^{l} \sum_{k=1}^{n} \frac{c_i}{\sqrt{k}} \cdot \upsilon_k'' \otimes \upsilon_{s_k}'' \otimes \xi_{AA'}^{i1} \otimes \xi_{BB'}^{ij_k^i}; \tag{6.205}$$

$$u_A \otimes u_B(\psi_{AB}''') \overset{\varepsilon}{\approx} q \cdot \kappa_n \otimes \sum_{i=1}^{l} \sum_{j_i=1}^{m_i} \xi_{AA'}^{ij_i} \otimes \xi_{BB'}^{ij_i}. \tag{6.206}$$

Here the right-hand sides of (6.203) - (6.206) have been arranged so as to obtain vectors in the six-fold tensor product

$$H_A'' \otimes H_B'' \otimes H_A \otimes H_A' \otimes H_B \otimes H_B'.$$

We will repeatedly invoke the following lemma, whose proof just unfolds the notation (on the appropriate identification of a with $a \otimes 1_{H_2}$ and of b with $1_{H_1} \otimes b$).

Lemma 6.22. *Assume* **PI** *and* **UI**. *For any pair of unitary operators u_1 on H_1 and u_2 on H_2, and any unit vector $\psi \in H_1 \otimes H_2$, one has*

$$P_{(u_1 \otimes 1_{H_2})\psi}(b = \gamma|x) = P_\psi(b = \gamma|x); \tag{6.207}$$

$$P_{(1_{H_1} \otimes u_2)\psi}(a = \lambda|x) = P_\psi(\lambda = x|x). \tag{6.208}$$

Since we assume that a is nondegenerate, there is a bijective correspondence between its eigenvalues $a = \lambda_i$ and its eigenvectors υ_i. Instead of $P(a = \lambda_i)$ dressed with whatever parameters x or ψ, we may then write $P(\upsilon_i)$, where a is understood, and analogously for the more complicated operators on tensor products of Hilbert space appearing below. Repeatedly using Lemma 6.22, we proceed as follows.

- From Step 2, using the notation explained below (6.172),

$$P_{q\cdot\sum_{i=1}^{l}\sum_{j_i=1}^{m_i}\xi^{ij_i}_{BB'}}(\xi^{ij}_{BB'}|x) = q^2. \tag{6.209}$$

- From (6.156) in ***PE*** and (6.209),

$$P_{q\cdot\sum_{i,j_i}\xi^{ij_i}_{AA'}\otimes\xi^{ij_i}_{BB'}}(\xi^{ij}_{BB'}|x) = q^2. \tag{6.210}$$

- From (6.155) in ***SE*** and (6.210),

$$P_{q\cdot\kappa_n\otimes\sum_{i,j_i}\xi^{ij_i}_{AA'}\otimes\xi^{ij_i}_{BB'}}(\xi^{ij}_{BB'}|x) = q^2. \tag{6.211}$$

- From (6.211), ***CP*** (whose notation we use), and (6.206),

$$P_{(u_A\otimes u_B)\psi'''_{AB}}(\xi^{ij}_{BB'}|x) \overset{\varepsilon}{\approx} q^2. \tag{6.212}$$

- Recall the number m (satisfying $m \geq m_i$ for all i). From (6.212) and Lemma 6.22,

$$\begin{aligned} &P_{(1_{H'''_A}\otimes u_B)\psi'''_{AB}}(\xi^{ij_i}_{BB'}|x) \overset{\varepsilon}{\approx} q^2 \ (j_i = 1,\ldots,m_i);\\ &P_{(1_{H'''_A}\otimes u_B)\psi'''_{AB}}(\xi^{ij_i}_{BB'}|x) \overset{\varepsilon}{\approx} 0 \ (j_i = m_i+1,\ldots,m). \end{aligned} \tag{6.213}$$

We now start a different line of argument, to be combined with (6.213) in due course.

- From ***PE***, ***SE***, and (6.172), with $\upsilon^i_A \in H_A$ denoting $\upsilon_i \in H$, we have

$$P_\psi(a = \lambda_i|x) \equiv P_\psi(\upsilon_i|x) = P_{\kappa_n\otimes\sum_i c_i\cdot\xi^{i1}_{AA'}\otimes\xi^{i1}_{BB'}}(\upsilon^i_A|x). \tag{6.214}$$

- Using Lemma 6.22, (6.203), and (6.204),

$$P_{\kappa_n\otimes\sum_i c_i\cdot\xi^{i1}_{AA'}\otimes\xi^{i1}_{BB'}}(\upsilon^i_A|x) = P_{(1_{H'''_A}\otimes u_B)\psi'''_{AB}}(\upsilon^i_A|x), \tag{6.215}$$

  and hence

$$P_\psi(a = \lambda_i|x) = P_{(1_{H'''_A}\otimes u_B)\psi'''_{AB}}(\upsilon^i_A|x). \tag{6.216}$$

- From quantum mechanics, notably (6.151), and (6.205), for any $i' \neq i$ we have

$$p_{(1_{H'''_A}\otimes u_B)\psi'''_{AB}}(\upsilon^{i'}_A \otimes \xi^{ij_i}_{BB'}) = 0. \tag{6.217}$$

- From ***CQ*** and (6.217), for any $i' \neq i$,

$$P_{(1_{H_A'''}\otimes u_B)\psi_{AB}'''}(\upsilon_A^{i'}, \xi_{BB'}^{ij_i}|x) = 0. \tag{6.218}$$

- From ***PI***,

$$P(\upsilon_A^{i'}|x) = \sum_{i,j_i} P(\upsilon_A^{i'}, \xi_{BB'}^{ij_i}|x); \tag{6.219}$$

$$P(\xi_{BB'}^{ij_i}|x) = \sum_{i'} P(\upsilon_A^{i'}, \xi_{BB'}^{ij_i}|x). \tag{6.220}$$

- From (6.218), (6.219), and (6.220),

$$P_{(1_{H_A'''}\otimes u_B)\psi_{AB}'''}(\upsilon_A^{i}|x) = \sum_{j_i} P_{(1_{H_A'''}\otimes u_B)\psi_{AB}'''}(\xi_{BB'}^{ij_i}|x). \tag{6.221}$$

Finally, from (6.214), (6.221), (6.213), and (6.197) we obtain

$$P_\psi(a = \lambda_i|x) \overset{\varepsilon}{\approx} \sum_{j_i}^{m_i} q^2 = m_i \cdot q^2 = c_i^2. \tag{6.222}$$

Since $c_i > 0$ we have $c_i^2 = |c_i|^2$; using (6.192) and letting $\varepsilon \to 0$ then proves step 3:

$$P_\psi(a = \lambda_i|x) = |c_i|^2 = p_\psi(a = \lambda_i). \tag{6.223}$$

Finally, we remove our standing assumption that the spectrum of a be nondegenerate. In the degenerate case one has

$$p_\psi(a = \lambda_i) = \sum_{j_i} p_\psi(\upsilon_{j_i}), \tag{6.224}$$

where the sum is over any orthonormal basis $(\upsilon_{j_i})_{j_i}$ of the eigenspace of λ_i. Similarly, since each vector υ_{j_i} gives $a = \lambda_i$, probability theory gives for all x,

$$P(a = \lambda_i|x) = \sum_{j_i} P(\upsilon_{j_i}|x). \tag{6.225}$$

The nondegenerate case of the theorem (which distinguishes the states υ_{j_i}) yields

$$P_\psi(\upsilon_{j_i}|x) = p_\psi(\upsilon_{j_i}), \tag{6.226}$$

from which (6.157) follows once again:

$$P_\psi(a = \lambda_i|x) = \sum_{j_i} P_\psi(\upsilon_{j_i}|x) = \sum_{j_i} p_\psi(\upsilon_{j_i}) = p_\psi(a = \lambda_i).$$

Our proof of the Colbeck–Renner Theorem is now complete. □

Under less stringent assumptions this theorem might have been regarded as the conclusion of von Neumann's program to disprove the possibility of completing quantum mechanics by adding hidden variables, but as yet this seems unwarranted.

Notes

§6.1. **From von Neumann to Kochen–Specker**

> 'For decades nobody spoke up against von Neumann's arguments, and his conclusions were quoted by some as the gospel'. (Belinfante, 1973, pp. 24)

Theorem 6.2 is due to non Neumann (1932, §IV.2); it was the first result to impose useful constraints on hidden variable theories, anticipating all later literature on the subject. Unfortunately (as part of their general anti-Copenhagen rhetoric), Bell and his followers left the realm of decent academic discourse by calling von Neumann's arguments against hidden variables 'silly' and 'foolish', through which they merely displayed the depth of their own misunderstanding of von Neumann's reasoning; see Caruana (1995), Bub (2011a), and especially Dieks (2016b). In fact, von Neumann (1932, p. 172) carefully qualifies his Theorem 6.2 by stating that it follows '*im Rahmen unserer Bedingungen*' (i.e. '*given our assumptions*'), of which he earlier (on p. 164) admits that linearity is physically reasonable only for *commuting* operators, but nonetheless justifies this assumption through an ensemble argument (now outdated, but by no means 'silly'). Though couched in agreeable academic parlance, the earlier critique by Hermann (1935) was misguided, too (Dieks, 2016b).

The Kochen–Specker Theorem is due to Kochen & Specker (1967); the authors were originally logicians. A similar but less precise statement had appeared earlier in Bell (1966), who was not cited by Kochen and Specker; some authors refer to the ***Bell–Kochen–Specker Theorem***. The *Nature* assumption has been experimentally verified, cf. Huang et al (2003). The proof of the fundamental Lemma 6.7 we present is essentially due to Kochen and Specker, as simplified by Peres (1995). Our independent proof for $\mathbb{C}^4$ is taken from Cabello et al (1996). Surveys of various proofs are given by Brown (1992) and Gould (2009); see also Waegell & Aravind (2012) and references therein, as well as Bub (1997) for another proof. From the Netherlands, we cannot fail to mention the short proof by Gill & Keane (1996). For geometric aspects (and even a link with M.C. Escher) see Zimba & Penrose (1993).

One finds two opposite directions of research around the Kochen–Specker Theorem. A computational one, which seems hardly relevant to conceptual issues in physics (the goal rather being *The Guinness Book of Records*), consists of attempts to find a *minimal* set of vectors that *cannot* be coloured. See, for example, Pavicic et al (2005) for arbitrary dimension and Arends (2009) and Uijlen & Westerbaan (2015) for $\mathbb{R}^3$, the latter paper showing that at least 22 vectors are needed.

The other, which is of significant conceptual importance and hence is worth some more extensive discussion, consists of attempts to find a *maximal* set of vectors that *can* be coloured. That is, one looks for large (preferably dense and measurable) subsets S^2_c of S^2 for which there exists a function $\tilde{V} : S^2_c \to \{0,1\}$ that satisfies:

- $\tilde{V}(-\mathbf{u}) = \tilde{V}(\mathbf{u})$ for each $\mathbf{u} \in S^2_c$;
- $\tilde{V}(\mathbf{u}_1) + \tilde{V}(\mathbf{u}_2) + \tilde{V}(\mathbf{u}_3) = 2$, for each (orthonormal) basis $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ of $\mathbb{R}^3$ whose elements lie in S^2_c.

The first result in this direction was obtained by Meyer (1999) and Havlicek et al (2001), who showed that one may take $S_c^2 = S^2 \cap \mathbb{Q}^3$; this choice was motivated by invoking finite precision arguments to circumvent the Kochen–Specker Theorem, see below. To write down a suitable function $\tilde{V} : S^2 \cap \mathbb{Q}^3 \to \{0,1\}$, we first define an auxiliary function $S : S^2 \cap \mathbb{Q}^3 \to \mathbb{Z}$ by

$$S\left(\frac{n_1}{m_1}, \frac{n_2}{m_2}, \frac{n_3}{m_3}\right) = \frac{n_3}{m_3} \cdot \frac{\mathrm{lcm}(m_1, m_2, m_3)}{\gcd(n_1, n_2, n_3)}, \tag{6.227}$$

where lcm is the *least common multiple* and gcd is the *greatest common divisor* of the argument. This function is obviously well defined. Then the following works:

$$\tilde{V}(x,y,z) = 0 \text{ if } S(x,y,z) \text{ is odd;} \tag{6.228}$$

$$\tilde{V}(x,y,z) = 1 \text{ if } S(x,y,z) \text{ is even.} \tag{6.229}$$

More generally, for an arbitrary n-dimensional) Hilbert space H, with $n < \infty$, Clifton & Kent (2000) proved the existence of a countable dense colorable subset $\mathscr{P}_1(H)_c$ of $\mathscr{P}_1(H)$ (cf. Definition 6.9), with the additional property that different resolutions of the identity drawn from $\mathscr{P}_1(H)_c$ never share a projection (so that the key strategy proof of Lemma 6.7, which is based on the existence of overlapping bases, falls apart). Given some enumeration $(e_i^{(1)}), (e_i^{(2)}), \ldots$ of the countable set of all resolutions of the identity drawn from $\mathscr{P}_1(H)_c$, so that each $(e_1^{(k)}, \ldots, e_n^{(k)})$ is a basis of H, $k \in \mathbb{N}$, each possible coloring $W = W_f$ bijectively corresponds to some function $f : \mathbb{N} \to \{1, \ldots, n\}$ through

$$W_f(e) = 1 \text{ if } e = e_{f(k)}^{(k)}; \tag{6.230}$$

$$W_f(e) = 0 \text{ otherwise.} \tag{6.231}$$

Note that because of the total incompatibility of the projections, each $e \in \mathscr{P}_1(H)_c$ belongs to a unique resolution $(e_i^{(k)})$, so that W_f is well defined. The statistical predictions of quantum mechanics may then be recovered as follows. For each density operator $\rho \in \mathscr{D}(H)$ we may define a probability measure μ_ρ on the set $\underline{n}^{\mathbb{N}}$ of all functions $f : \mathbb{N} \to \{1, \ldots, n\}$ by imposing the conditions

$$\mu_\rho\left(\{f \in \underline{n}^{\mathbb{N}} \mid W_f(e_i^{(k)}) = \lambda_i^{(k)} \forall i = 1, \ldots, n, k \in K\}\right) = \prod_{k \in K} \mathrm{Tr}\left(\rho \prod_{i=1}^{n} [e_i^{(k)} = \lambda_i^{(k)}]\right), \tag{6.232}$$

where $\lambda_i^{(k)} \in \{0,1\}$, $K \subset \mathbb{N}$ is finite, and $[e_i^{(k)} = \lambda_i^{(k)}]$ is the projection onto the corresponding eigenspace $H_{\lambda_i^{(k)}}$ of the projection $e_i^{(k)}$ (more generally, for $a \in B(H)_{\mathrm{sa}}$ we write $[a = \lambda]$ for the spectral projection e_λ defined by a and $\lambda \in \sigma(a)$). The subset of $\underline{n}^{\mathbb{N}}$ in the argument of μ_ρ is hereby declared measurable; existence and uniqueness of the measure μ_ρ on a suitable σ-algebra follow from the Kolmogorov extension theorem of measure theory, which applies because the marginals (6.232) satisfy the appropriate consistency conditions, cf. Hermens (2009) for details.

This formula guarantees that the left-hand side vanishes if $\lambda_i^{(k)} = 0$ for each i, and also if $\lambda_i^{(k)} = 1$ for more than one value of i. If $K = \{k_0\}$ is a singleton and $\lambda = (\lambda_1, \ldots, \lambda_n)$, then the right-hand side (and hence the left-hand side) is the Born probability for the outcome $e_i^{(k_0)} = \lambda_i$ for each i, i.e.,

$$\mu_\rho \left(\{ f \in \underline{n}^{\mathbb{N}} \mid W_f(e_i^{(k_0)}) = \lambda_i \, \forall \, i = 1, \ldots, n \} \right) = \mathrm{Tr} \left(\rho \prod_{i=1}^{n} [e_i^{(k_0)} = \lambda_i] \right). \quad (6.233)$$

Consequently, it is true by construction that for any admissible measurement in quantum mechanics (in that all observables commute), i.e., for each $k_0 \in \mathbb{N}$, averaging over the 'hidden variable' $f \in \underline{n}^{\mathbb{N}}$ reproduces the statistical predictions of quantum mechanics. This success is achieved at a high cost, however:

- Two random variables $e_i^{(k)}$ and $e_{i'}^{(k')}$ are statistically independent (with respect to μ_ρ) whenever $k \neq k'$, even though $\|e_i^{(k)} - e_{i'}^{(k')}\|$ may be arbitrarily small.
- For each $f \in \underline{n}^{\mathbb{N}}$ the associated coloring W_f is maximally discontinuous, in that for each $\mathbf{u} \in \mathscr{P}_1(H)_c$ and each $\varepsilon > 0$ there is $\mathbf{u}' \in \mathscr{P}_1(H)_c$ such that although $\|e_{\mathbf{u}} - e_{\mathbf{u}'}\| < \varepsilon$ one has $W_f(e_{\mathbf{u}}) \neq W_f(e_{\mathbf{u}'})$, so that in fact $|W_f(e_{\mathbf{u}}) - W_f(e_{\mathbf{u}'})| = 1$.

These facts were noted by Clifton & Kent themselves, and Appleby (2005) proved that they are a necessary feature of all constructions that involve sufficiently large subsets of $\mathscr{P}_1(H)$ that can be colored.

Without challenging their mathematical significance, these discontinuities undermine any potential physical relevance such models might have, and this in turn challenges the reason such models were introduced in the first place (Meyer, 1999), namely the (alleged) *finite precision loophole* of the Kochen–Specker Theorem.

The thrust of this loophole is that it would be an illusion for an experimentalist like Alice to claim that she measures some observable a with infinite accuracy; in fact, given $\varepsilon > 0$ she might equally well measure some a' with $\|a - a'\| < \varepsilon$. Consequently, finding a dense colorable subset $\mathscr{P}_1(H)_c \subset \mathscr{P}_1(H)$ should suffice for a hidden variable interpretation of quantum mechanics, since if Alice believes she measures some projection e, the model assigns a value $W(e')$ to the projection $e' \in \mathscr{P}_1(H)_c$ she actually measures (where e' is selected by some algorithm that is part of the theory itself, cf. Clifton & Kent (2000)), and presents that value to Alice as the outcome of her measurement. However, owing to the discontinuities just mentioned, this value is as arbitrary as the identification of e'.

As emphasized by Barrett & Kent (2004), this arbitrariness, although perhaps undesirable, does not by itself affect the ability of the Clifton–Kent model to reproduce the statistical predictions of quantum mechanics. On the other hand, it would be pretty awkward to have a theory whose individual value attributions are completely arbitrary, especially since the finite precision argument is predicated on the idea that observables close to the one Alice believes herself to measure (i.e., e) should have approximately the same value as the one she actually does measure (namely, e'). If this is not the case, her measurements are pointless and the hidden variable W_f would be empirically inaccessible and hence truly "hidden" (Appleby, 2005).

See also Hermens (2009, 2016). This last point applies to Corollary 6.12, which would no longer be true if the set X_A of all bases of $\mathbb{R}^3$ in Definition 6.11 would be replaced by some subset $X_A^c \subset X_A$ drawn from a colorable subset S_c^2 of S^2. Each $z \in X_Z$ would then correspond to some coloring $\mathbf{u} \mapsto \tilde{F}(\mathbf{u}, z)$ of S_c^2, which, by the above discussion, would be maximally discontinuous and hence empirically inaccessible. Nonetheless, such a theory does exist in principle.

The aim of maximizing colorable sets was pursued in a different direction by Bub & Clifton (1996); see also Bub (1997). Given a "preferred" observable $a \in B(H)_{\mathrm{sa}}$ and a pure state $e \in \mathscr{P}_1(H)$, these authors look for a maximal sublattice $\mathscr{P}(e,a)$ of $\mathscr{P}(H)$ that contains all spectral projections of a (but, despite the notation $\mathscr{P}(e,a)$, does not necessarily contain e!), admits sufficiently many lattice homomorphism $h : \mathscr{P}(e,a) \to \{0,1\}$ (i.e., binary valuations) such that the Born measure μ_e on $\sigma(a)$, i.e., $\mu_e(\Delta) = \mathrm{Tr}(ee_\Delta)$, $\Delta \subseteq \sigma(a)$, can be reproduced by averaging over these homomorphisms, and finally is invariant under all unitary isomorphisms of $\mathscr{P}(H)$ that commute with both e and a. Equivalently, one wants a maximal C*-subalgebra $A(a,e)$ of $B(H)$ that contains a, admits sufficiently many dispersion-free states so as to reproduce the Born probabilities defined by a in the given state e, and is invariant in the said way (a fourth condition used by Bub and Clifton is superfluous; see Bub, 1997, p. 128). Asuming for simplicity that $n = \dim(H) < \infty$, the answer is

$$A(a,e) = C^*(e_\lambda e e_\lambda, \lambda \in \sigma(a))' \tag{6.234}$$

where, as always, e_λ is the projection into the eigenspace H_λ for $\lambda \in \sigma(a)$, and the prime denotes the commutant (one might as well take the commutant of the *set* of all $e_\lambda e e_\lambda$). Equivalently, putting $e = e_\psi = |\psi\rangle\langle\psi|$, eq. (6.234) is the C*-algebra generated by all projections f_λ onto the nonzero components $e_\lambda \psi$ of ψ in each H_λ and all one-dimensional projections that are orthogonal to all f_λ (given that $\dim(H) < \infty$, this is the same as the linear span of these projections). Thus $A(a,e)$ always contains $C^*(a)$, since it contains each e_λ, $\lambda \in \sigma(a)$), but note that $A(a,e)$ need not be commutative. In comparison, if the requirement had been the reproduction of all Born probabilities for *arbitrary* pure states e rather than for some *given* e, the answer would have been any maximal abelian C*-algebra in $B(H)$ that contains $C^*(a)$; if a has non-degenerate spectrum, this is just $C^*(a)$ itself. The simplest possibility is

$$A(1_H, e) = C^*(e)' = \{e\}', \tag{6.235}$$

which is the linear span of all projections $f \in \mathscr{P}(H)$ for which either $e \leq f$ or $e \leq 1_H - f$ (i.e., if $e = e_\psi$, then either $\psi \in fH$ or $\psi \in (fH)^\perp$). In other words, we have $a \in A(1_H, e)$ iff ψ is an eigenvector of a (i.e. the eigenvector-eigenvalue link).

Each dispersion-free state on $A(a,e)$, or, equivalently, each homomorphism $h_\lambda : \mathscr{P}(e,a) \to \{0,1\}$, corresponds to one of the projections f_λ through $h_\lambda(f_\lambda) = 1$ and $h_\lambda(f) = 0$ for all other one-dimensional projections f in $\mathscr{P}(e,a)$. The Born probabilities from e are then recovered by assigning (Born) measure $\mathrm{Tr}(ef_\lambda)$ to h_λ.

Though interesting, this result mainly supports so-called modal interpretations of quantum mechanics, which we reject, since they tell us nothing physical about the measurement process and address the measurement problem only philosophically.

§6.2. **The Free Will Theorem**

The Free Will Theorem was published in two versions by Conway & Kochen (2006, 2009). Analogous results had previously been published by Heywood & Redhead (1983), Stairs (1983), Brown & Svetlichny (1990), and Clifton (1993), of which only the first paper was cited by Conway and Kochen. Moreover, the close relationship to Bell's (1964) Theorem might well be insisted on as a topic that should have been discussed in the original papers. Other critical literature (making the points listed in the preamble to this chapter) includes Bassi & Ghirardi (2007), 't Hooft (2007), Goldstein et al (2010), Wüthrich (2011), Hemmick & Shakur (2012), Cator & Landsman (2014), Hermens (2014, 2015), and Walleczek (2016).

The original (Strong) Free Will Theorem (FWT) states that three assumptions, called SPIN, TWIN, and MIN, imply that the response of a spin-one particle to the bipartite experiment with spin-one particles described above 'is not a function of properties of that part of the universe that is earlier than this response (...).' Here SPIN and TWIN are the first and second half of our *Nature* axiom, whilst MIN expresses a form of context-locality as well as the loose assumption that Alice and Bob may 'freely choose' their settings a and b, respectively. Accordingly, in our notation, Conway and Kochen only use the parameter space Z, rather than the full space X we need in order to consistently axiomatize determinism. Their formulation contains an implicit assumption of determinism, whose precise nature only becomes clear from their proof, and which is akin to our formulation, except for the crucial difference that the function they allude to only acts on the particle variables and not on the settings of the experiment (of which, as already noted, Conway and Kochen just say that the experimenters can 'freely choose' them).

Conway and Kochen paraphrase their theorem as follows:

> 'if indeed we humans have free will, then elementary particles already have their own small share of this valuable commodity. More precisely, if the experimenter can freely choose the directions in which to orient his apparatus in a certain measurement, then the particles response (to be pedantic—the universe's response near the particle) is not determined by the entire previous history of the universe. (...) our theorem asserts that if experimenters have a certain freedom, then particles have exactly the same kind of freedom. Indeed, it is natural to suppose that this latter freedom is the ultimate explanation of our own. (...) Granted our three axioms [i.e., the physical ones and freedom of choice], the Free Will Theorem shows that nature itself is nondeterministic.'

However, such far-reaching conclusions seem unwarranted by the actual technical content of the theorem. Indeed, though it is also assumed in Bell's first theorem (see §6.5 below), the conjunction of *Determinism* and *Freedom* is *a priori* is uncomfortable, especially since the main novelty of the FWT lies in the emphasis Conway and Kochen (unlike Bell) put on free will. The authors acknowledge at least this point already on the first page of their first paper (Conway & Kochen, 2006), in which they anticipate criticism of the kind:

> "'I saw you put the fish in!" said a simpleton to an angler who had used a minnow to catch a bass.'

Indeed, also after more serious philosophical analysis, it has been concluded that:

> 'Their [Conway & Kochen's] case against determinism thus has all the virtues of theft over honest toil. It is truly indeterminism in, indeterminism out.' (Wüthrich, 2011)

Our formulation of the FWT, in which the original allusion to undefined free will in allowing arbitrary settings of the experiment has been replaced by complete determinism including the settings, avoids this criticism.

To derive (6.35) - (6.38), we use (6.21) to write down the formulae

$$\begin{aligned}
P_{\psi_0}(F_i = 1, G_j = 1 | A = a, B = b) &= \langle \psi_0, (1_3 - |\mathbf{u}_i\rangle\langle\mathbf{u}_i|) \otimes (1_3 - |\mathbf{v}_j\rangle\langle\mathbf{v}_j|)\psi_0\rangle; \\
P_{\psi_0}(F_i = 0, G_j = 0 | A = a, B = b) &= \langle \psi_0, |\mathbf{u}_i\rangle\langle\mathbf{u}_i| \otimes |\mathbf{v}_j\rangle\langle\mathbf{v}_j|\psi_0\rangle; \\
P_{\psi_0}(F_i = 1, G_j = 0 | A = a, B = b) &= \langle \psi_0, (1_3 - |\mathbf{u}_i\rangle\langle\mathbf{u}_i|) \otimes |\mathbf{v}_j\rangle\langle\mathbf{v}_j|\psi_0\rangle; \\
P_{\psi_0}(F_i = 0, G_j = 1 | A = a, B = b) &= \langle \psi_0, |\mathbf{u}_i\rangle\langle\mathbf{u}_i| \otimes (1_3 - |\mathbf{v}_j\rangle\langle\mathbf{v}_j|)\psi_0\rangle.
\end{aligned}$$

For example, for any pair of unit vectors $\mathbf{u}, \mathbf{v}$ we have

$$\begin{aligned}
&\langle \psi_0, |\mathbf{u}\rangle\langle\mathbf{u}| \otimes |\mathbf{v}\rangle\langle\mathbf{v}|\psi_0\rangle = \\
&\quad \tfrac{1}{3}\langle \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3, \mathbf{u}| \otimes |\mathbf{v}\rangle\langle\mathbf{v}|(\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3)\rangle = \\
&\quad \tfrac{1}{3}\langle \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3, \langle\mathbf{u}, \mathbf{v}\rangle\mathbf{u} \otimes \mathbf{v}\rangle \\
&= \tfrac{1}{3}\langle \mathbf{u}, \mathbf{v}\rangle^2,
\end{aligned}$$

which gives (6.36). The other cases are similar.

The implications of the finite precision loophole of the Kochen–Specker Theorem for the Free Will Theorem were analyzed by Hermens (2014), who concluded that this loophole does not apply. We give a more precise argument to this effect.

We have dense colorable subsets $X_A^c \subset X_A$ and $X_B^c \subset X_B = X_A$, where X_A^c may or may not coincide with X_B^c. If not, the perfect correlation condition (6.54) in the *Nature* assumption cannot even be stated, but even if $X_A^c = X_B^c$, since finite precision of experiment has been declared to be an issue it would be quite out of character to impose (6.54). Instead, one needs a probabilistic version of this condition, of which it will turn out that it cannot be satisfied. As in the notes to the previous section, for each density matrix ρ one needs a probability measure μ_ρ on Z that reproduces the statistical quantum-mechanical predictions for the associated quantum state. Compared to the notes to the previous section, the role of W is now played by z, in that for given F and G one might write

$$W(a,b) = (\hat{F}(a,z), \hat{G}(b,z). \tag{6.236}$$

This measure may be constructed analogously to (6.232), i.e., for any sequence $(a^{(k)})$ of bases drawn from X_A^c, any sequence $(b^{(k)})$ of bases drawn from X_B^c, and any sequences $(\lambda^{(k)})$ and $(\gamma^{(k)})$ in Λ, cf. (6.22), where $k \in K \subset \mathbb{N}$ is arbitrary, we define

$$\begin{aligned}
\mu_\rho(\{z \in Z \mid \hat{F}(a^{(k)}, z) = \lambda^{(k)}, \hat{G}(b^{(k)}, z) = \gamma^{(k)}, k \in K\} = \\
\prod_{k \in K} \mathrm{Tr}\left(\rho \prod_{i,j=1}^{3} [J^2_{\mathbf{u}_i} = \lambda_i^{(k)}] \cdot [J^2_{\mathbf{v}_j} = \gamma_j^{(k)}]\right),
\end{aligned} \tag{6.237}$$

where, as in the main text,

$$a = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3); \tag{6.238}$$

$$b = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3). \tag{6.239}$$

Note that $J^2_{\mathbf{u}_i}$ acts on Alice's Hilbert space $\mathbb{C}^3$ whilst $J^2_{\mathbf{v}_j}$ acts on Bob's. In particular, for fixed $k_0 \in K$ and $\lambda, \gamma \in \Lambda$, we have the special case of (6.237) for compatible measurements, viz.

$$\mu_\rho(\{z \in Z \mid \hat{F}(a^{(k_0)}, z) = \lambda, \hat{G}(b^{(k_0)}, z) = \gamma\} = \mathrm{Tr}\left(\rho \prod_{i,j=1}^{3} [J^2_{\mathbf{u}_i} = \lambda_i] \cdot [J^2_{\mathbf{v}_j} = \lambda_j]\right),$$

where in the main text we would have written $P_\rho(F = \lambda, G = \mu | A = a, B = b)$ for the right-hand side. Hence for the correlated state $\rho = |\psi_0\rangle\langle\psi_0|$ we obtain from (6.42):

$$\mu_{\psi_0}(\{z \in Z \mid \hat{F}_i(a,z) \neq \hat{G}_j(b,z)\}) = \tfrac{2}{3}(1 - \langle \mathbf{u}_i, \mathbf{v}_j\rangle^2), \tag{6.240}$$

which of course vanishes if $\mathbf{u}_i = \mathbf{v}_j$. If the expression $1 - \langle \mathbf{u}_i, \mathbf{v}_j\rangle^2$ appearing here is small, then the projections $e_{\mathbf{u}_i}$ and $e_{\mathbf{v}_j}$ are close (in norm), since

$$\|e_{\mathbf{u}_i} - e_{\mathbf{v}_j}\|^2 \leq 2(1 - \langle \mathbf{u}_i, \mathbf{v}_j\rangle^2). \tag{6.241}$$

Eq. (6.240) therefore allows us to make rigorous sense of Hermens' (2014) heuristic idea that the assumption (6.54) in the FWT should be modified as follows:

> 'if $\|e_{\mathbf{u}_i} - e_{\mathbf{v}_j}\|$ is small, then in most of the cases $\hat{F}_i(a,z) = \hat{G}_j(b,z)$.'

Namely, we replace (6.54) by the following approximate correlation condition:

- For every $\varepsilon > 0$ there is $\delta > 0$ such that if $1 - \langle \mathbf{u}_i, \mathbf{v}_j\rangle^2 < \delta$, then

$$\mu_{\psi_0}(\{z \in Z \mid \hat{F}_i(a,z) \neq \hat{G}_j(b,z)\}) < \varepsilon. \tag{6.242}$$

Indeed, if the theory existed, on could simply take $\delta = \varepsilon$. However, a theory satisfying (6.242) does not exist, as can be proved by contradiction: if $\hat{F}_i(a,z) = \hat{G}_j(b,z)$ for all pairs $(\mathbf{u}_i, \mathbf{v}_j)$ such that $1 - \langle \mathbf{u}_i, \mathbf{v}_j\rangle^2 < \varepsilon$, then the proof of Theorem 6.13 shows not only that (6.32) still holds on the modified *Nature* assumption (so that $\tilde{F}(\cdot, z)$ again defines a coloring of S^2), but that *in addition* we have

$$1 - \langle \mathbf{u}, \mathbf{u}'\rangle^2 < \delta \;\Rightarrow\; \tilde{F}(\mathbf{u}, z) = \tilde{F}(\mathbf{u}', z). \tag{6.243}$$

In particular, the apparently weaker correlation condition ending with (6.242) is actually *stronger* than its exact counterpart (6.54).

Thus Theorem 6.13 still holds on this revised *Nature* assumption, so that unlike the Kochen–Specker Theorem, the Free Will Theorem is immune to the finite precision loophole. The price for this immunity is that, quite against the spirit of the FWT, some probabilistic reasoning had to be invoked, so that the difference between the FWT and Bell's first theorem has blurred even further.

§6.3. **Philosophical intermezzo: Free will in the Free Will Theorem**

The literature on free will is immense. Introductory accounts include Walter (2001), which focuses on the connection with neuroscience, Doyle (2011), and Beebee (2013), the second of which remains largely philosophical, the third even completely. A very sophisticated recent defense of compatibilism is Ismael (2016). Lewis's 'local miracle compatibilism' was proposed in Lewis (1981). What's more:

> '[Lewis's paper is] the finest essay that has ever been written in defense of compatibilism—possibly the finest essay that has ever been written about any aspect of the free will problem.' (van Inwagen, 2008).

Saunders (1968) already made a point similar to Lewis's; see also Moore (1912, Ch. 6). For Lewis's theory of counterfactuals see Lewis (1973, 1979, 2000), as well as Menzies (2014). See also Fischer (1994), Beebee (2003, 2013), and Vihvelin (2013).

Although Lewis's position is called *local miracle compatbilism*, a miracle takes place neither in the actual world where Alice's hand is at rest nor in the possible world where she raises it, i.e., a law is broken neither in the former nor in the latter:

> 'This is what Lewis means by a 'miracle': an event *M* is a miracle if and only if *M* occurs at *possible world w*, and *M* is contrary to some *actual* law (or combination of laws) *L*. The point here is that while *M* is a miracle in Lewis's sense, it is not contrary to any of *w*'s laws of nature. At *w*, *L* simply isn't a law in the first place. So, as things *actually* happened—in the *actual* world—*L* is a law, and *m* does not occur, so there is no miracle in the usual sense of 'miracle'. *m* is only a 'miracle' in Lewis's special sense of 'miracle': something (*m*) happens in *w* that is contrary to the laws of nature in the *actual* world.'
> (Beebee, 2013, p. 62)

Unfortunately, confusion may arise if the quotation in the main text 'if I did it, a law would be broken' from Lewis (1981) is subjected to the following explanation:

> 'On Lewis's account of counterfactuals, the *truth conditions* for counterfactuals—what makes them true—are as follows. Suppose we have the counterfactual 'if *A* had been the case, *B* would have been the case' (so if *A* is 'I miss the bus' and *B* is 'I'm late', this counterfactual just says, 'if I'd missed the bus, I would have been late'). This counterfactual will be true if and only if, *at the closest possible world to the actual world* at which *A* is true, *B* is also true. So, our sample counterfactual, 'if I'd missed the bus, I would have been late', is true if and only if: *at the closest possible world to the actual world* at which I miss the bus, I'm late.' (Beebee, 2013, p. 60).

Removing any possible remaining doubt, on p. 62 she mentions that the closest possible world where I miss the bus is the world *w*. According to this explanation, then, Lewis's sentence 'if I did it, a law would be broken', would mean that *at the closest possible world to the actual world* in which I did it, a law *is* broken, i.e., in *w*. But according to Beebee's definition quoted in the main text of what Lewis means by a miracle, apparently this is not the right reading (and indeed it would, in our view, be nonsensical). Moreover, Lewis (1981) emphasizes that in the first bullet point in the main text above—which he defends—it is not the agent who would break a law, whereas in the second bullet point —rejected by Lewis—it is; in the first it is the breaking of some law at an earlier time that enables the agent to do what she, in our actual world, did not do. Thus Lewis's phrasing seems awkward.

Our development of Lewis's argument is indebted to Vihvelin (2013, pp. 164–165), who (re)states Lewis's first bullet point as the following conjunction:

1. **Slightly Different Past**: If I had raised my hand, the past would still have been exactly the same until shortly before the time of my decision.
2. **Slightly Different Laws:** If I had raised my hand, the laws would have been ever so slightly different in a way that permitted a divergence from the lawful course of actual history shortly before the time of my decision.

A second way in which Alice could (counterfactually) have raised here hand is through an instant (counterfactual) modification of the state of the world, as in Bennett (1984). This has been explicated by Vihvelin (2013, p. 165), too:

1. **Same Laws:** If I had raised my hand, the laws would still have been the same.
2. **Completely Different Past:** If I had raised my hand, past history would have been different all the way back to the Big Bang.

Here we prefer to write **Different Past**, since even though in this scenario the state indeed (by determinism) would have been different all the way back to the Big Bang, the entire trajectory of the world may or may not be close to the actual one. In this scenario, the two cases Lewis distinguishes take the form in the main text.

Since the main novelty of their papers lies in the emphasis on free will, the reader might wonder what Conway & Kochen themselves have to say about the subject. As we can read in the delightful biography of Conway by Roberts (2015), or watch in his video lectures on the Free Will Theorem (Conway, 2009), free will is indeed of great importance to at least the first author of the theorem. Unfortunately, his interest in free will seems unaccompanied by any philosophical sophistication, e.g.:

> 'Compatibilism in my view is silly. Sorry, I shouldn't just say straight off that it is silly. Compatibilism is an old viewpoint from previous centuries when philosophers were talking about free will. The were accustomed to physical theory being deterministic. And then there's the question: How can we have free will in this deterministic universe? Well, they sat and thought for ages and ages and ages and read books on philosophy and God knows what and they came up with compatibilism, which was a tremendous wrenching effect to reconcile 2 things which seemed incompatible. And they said they were compatible after all. But nobody would *ever* have come up with compatibilism if they thought, as turns out to be the case, that science wasn't deterministic. The whole business of compatibilism was to reconcile what science told you at the time, centuries ago down to 1 century ago: Science appeared to be totally deterministic, and how can we reconcile that with free will, which is not deterministic? So compatibilism, I see it as out of date, really. It's doing something that doesn't need to be done. However, compatibilism hasn't gone out of date, certainly, as far as the philosophers are concerned. Lots of them are still very keen on it. How can I say it? If you do anything that seems impossible, you're quite proud when you appear to have succeeded. And so really the philosophers don't want to give up this notion of compatibilism because it seems to damned clever. But my view is it's really nonsense. And it's not necessary. So whether it actually is nonsense or not doesn't matter.'
>
> (Conway, quoted in Roberts, 2015, pp. 361–362).

Finally, our version of van Inwagen's (1975) Consequence Argument is due to Beebee (2003), and the novel parts of this section are based on Landsman (2016c). For interesting philosophical criticism of this approach, see De Mola (2016).

§6.4. **Technical intermezzo: The GHZ-Theorem**

The GHZ Theorem appeared in Greenberger et al (1990) See also Clifton, Redhead, & Butterfield (1991) and Bub (1997). Innumerable variations on and generalizations of such arguments may be given, leading to equally many Free Will Theorems. All of these have their roots in algebraic properties of matrices, which hidden variable theories (in vain) try to reproduce.

§6.5. **Bell's theorems**

The original contributions to the theme of this section are Bell (1964, 1976), of which the first is one of the most famous papers of 20th century theoretical physics. Since there are more than 10,000 papers citing Bell (1964) alone, it is impossible to discuss all literature relevant to Bell's work. What we call his first theorem originates with Bell (1964), which incidentally was written after Bell (1966), but our treatment of the settings (taken from Cator & Landsman, 2014) is different. Though originally motivated as an attempt to make the Free Will Theorem look less of a *petitio principii*, it also addresses a problem Bell faced even according to some of his staunchest supporters (Norsen, 2009; Seevinck & Uffink, 2011), namely the tension between the idea that the hidden variables (in the pertinent causal past) should on the one hand include all ontological information relevant to the experiment, but on the other hand should leave Alice and Bob free to choose any settings they like.

His second theorem comes from Bell (1976), followed by Bell (1990a).
Apart from his own papers, which are reprinted in Bell, Gottfried & Veltman (2001), treatments of Bell's Theorems we regard as sound include Fine (1982), Jarrett (1984), Pitowsky (1989), van Fraassen (1991), Butterfield (1992a,b), Bub (1997), Werner, & Wolf (2001), Liang, Spekkens, & Wiseman (2011), Shimony (2013), Wiseman (2014), and Brown & Timpson (2015). Recent and mathematically innovative approaches include Abramsky & Brandenburger (2011), Acín et al (2015), and Fritz (2016). For history, see Gilder (2008) and Kaiser (2010).

Unfortunately, we have not been able to come to grips with (and hence do not cite) literature claiming that Bell's theorems are false, or have nothing to do with hidden variables, or prove that quantum mechanics (if not nature itself!) is nonlocal *per se*, or that he never changed his mind and only has one theorem saying it all.

The verification of (6.102) - (6.105) is analogous to the above computations deriving (6.35) - (6.38). In terms of the unit vector

$$v_a = \begin{pmatrix} \cos a \\ \sin a \end{pmatrix}, \tag{6.244}$$

the observable F Alice measures on setting $A = a$ is the projection $e_a = |v_a\rangle\langle v_a|$, and similarly for Bob. Hence the corresponding Born probabilities are given by

$$\begin{aligned}
P_{\psi_0}(F = 1, G = 1 | A = a, B = b) &= \langle \psi_0, e_a \otimes e_b \psi_0 \rangle; \\
P_{\psi_0}(F = 0, G = 0 | A = a, B = b) &= \langle \psi_0, (1_2 - e_a) \otimes (1_2 - e_b) \psi_0 \rangle; \\
P_{\psi_0}(F = 1, G = 0 | A = a, B = b) &= \langle \psi_0, e_a \otimes (1_2 - e_b) \psi_0 \rangle; \\
P_{\psi_0}((F = 0, G = 1 | A = a, B = b) &= \langle \psi_0, (1_2 - e_a) \otimes e_b \psi_0 \rangle.
\end{aligned}$$

For example, we have

$$\begin{aligned}\langle \psi_0, e_a \otimes e_b \psi_0\rangle &= \tfrac{1}{2}\langle \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2, |v_a\rangle\langle v_a| \otimes |v_b\rangle\langle v_b|(\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2)\rangle \\ &= \tfrac{1}{2}\langle \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2, (\cos a \cos b + \sin a \sin b) v_a \otimes v_b\rangle \\ &= \tfrac{1}{2}(\cos a \cos b + \sin a \sin b)^2 \\ &= \tfrac{1}{2}\cos^2(a-b).\end{aligned}$$

The CHSH-inequality (6.117) is due to Clauser, Horne, Shimony, & Holt (1969). The definitive (i.e., loophole-free) experimental verification of its violation in nature is Henson et al. (2015). A direct proof starts of (6.117) from the simpler inequality

$$P(F \neq H) \leq P(F \neq G) + P(G \neq H), \tag{6.245}$$

for three $\{0,1\}$-valued random variables F, G, H, which implies (6.117). To prove (6.245), one just writes

$$\begin{aligned}P(F \neq H) &= P(F=1, G=1, H=0) + P(F=1, G=0, H=0) \\ &\quad + P(F=0, G=1, H=1) + P(F=0, G=0, H=1),\end{aligned}$$

etc., and notes that each term on the left-hand side of (6.245) also occurs on the right-hand side. Since each term lies in $[0,1]$ and hence is positive, this implies (6.245). Our proof of Proposition 6.17 follows Werner & Wolf (2001), as does our proof of Theorem 6.18 (though not our formulation thereof, which once again derives from Cator & Landsman (2014). This proof shows that, as first noted by Fine (1982) and analyzed more deeply in Butterfield (1992b), there is no real distinction between the possibility of reproducing given (empirical) probabilities $P(F=\lambda, G=\gamma | A = a, B = b)$ *that satisfy Bell locality* by a *local deterministic hidden variable theory* or by a *local stochastic hidden variable theory*. Most current research in this direction, sparked by Popescu & Rohlich (1994), is therefore concerned with theories defined by formal joint conditional probabilities that satisfy a no signaling condition like OI instead of Bell locality, cf. Bub (2011b) and Brunner et al (2014) for reviews.

Formal conditional probabilities of the kind that Bell's second theorem uses have been axiomatized by e.g. Popper (1938) and Rényi (1955); the following axioms are theorems if conditional probabilities are defined à la Kolmogorov by (1.1). Let Σ be some σ-algebra and let $\mathscr{F} \subset \Sigma \backslash \{\emptyset\}$ be an ideal in Σ in the sense that if $B \in \Sigma$ and $C \in \mathscr{F}$, then $B \cap C \in \mathscr{F}$. A ***conditional probability*** on $(\Sigma, \mathscr{F})$ is a map

$$P : \Sigma \times \mathscr{F} \to [0,1]; \tag{6.246}$$

$$(A, C) \mapsto P(A|C), \tag{6.247}$$

such that:

1. For each $C \in \mathscr{F}$ the map $A \mapsto P(A|C)$ is a probability measure on Σ;
2. $P(A \cap B|C) = P(A|B \cap C) \cdot P(B|C)$, for each $A, B \in \Sigma$ and $C \in \mathscr{F}$.

Van Fraassen (1991) noted that if (6.121) holds, then the variable x is a ***common cause*** in the sense of Reichenbach for Alice's and Bob's outcomes (see Hofer-Szabó (2015) for a recent paper in this direction). To explain this observation, suppose two random processes F and G (like Alice's and Bob's measurements) are correlated, i.e., $P(F=\lambda, G=\gamma) \neq P(F=\lambda)P(G=\gamma)$. What might cause the correlation?

1. *Chance*. If Alice and Bob independently throw dice but always get the same result, there is a computable nonzero probability for this to happen without any reason. But this probability decreases as the number of occurrences grows.
2. *Causation*. One outcome influences or even determines the other. Maybe Bob, whose experiment is genuinely random, is able to manipulate Alice's experiment once he has seen his outcome. But according to relativity theory or other basic notions of causality in space-time, this should be impossible if Alice and Bob perform their measurements simultaneously and far from each other.
3. *Ur-determinism*. The initial conditions at the Big Bang plus deterministic Laws of Nature imply the correlation. However, physics becomes pointless if we endorse this option. The notion of *explanation* as the purpose of science is defeated and there is little difference between this argument and Divine Predestination.
4. *Identity*. The motions of my mirror image are strongly correlated with me, but that is because this image is really the same as me (at least in so far as motion is concerned, as opposed to e.g. thoughts). This example might also be explained using causation. Another example consists of Alice and Bob filming the same random process (which may also be explained using the following concept).
5. *Common Cause* A random process X is said to be a ***common cause*** for two correlated random processes if it precedes both and satisfies

$$P(F=\lambda, G=\gamma|X=x) = P(F=\lambda|X=x)P(G=\gamma|X=x). \tag{6.248}$$

   Another way to write this is $P(F=\lambda|G=\gamma, X=x) = P(F=\lambda|X=x)$, which shows that a common cause X screens off the dependence of F on G. Often the common cause is hidden and has to be inferred from the observed correlation (having excluded other explanations, like the ones above). A nice example of this is the inference of a manuscript called Q in New Testament studies. It is clear that the Gospels of Matthew and Luke both draw on Mark, but they also contain strikingly similar or even identical non-Markan passages. For various reasons it is unlikely that either one copied these from the other, so that the main hypothesis is that they both rely on Q, which is now lost. See e.g. Mack (1993).

From this perspective, the amazing fact is that the correlations in the Alice and Bob experiment with either spin-1 particle or photons cannot be explained by a common cause, since its existence (in the form of x) would imply the Bell inequality. However, of the four other explanations described above, no. 1 is ridiculous given the statistics of the relevant experiments, no. 2 is at odds with relativity, and no. 4 seems inapplicable. This leaves no. 3, which seems only supported by 't Hooft (2016), who denies the independence assumptions (i.e. between the settings and the state of the pair of particles undergoing measurement) lying at the basis of both the Free Will Theorem and Bell's theorems. Every way you look at it you lose!

Generalizations of Theorem 6.19 to operator algebras were given e.g. by Baez (1987), Raggio (1988), Werner (1989), and Bacciagaluppi (1993), as follows. Let A and B be unital C*-algebras, with projective tensor product $A\hat{\otimes}B$ (i.e., the completion of the algebraic tensor product $A\otimes B$ in the maximal C*-cross-norm), cf. §C.13; the choice of the projective tensor product guarantees that each state on $A\otimes B$ extends to a state on $A\hat{\otimes}B$ by continuity; conversely, since $A\otimes B$ is dense in $A\hat{\otimes}B$, each state on the latter is uniquely determined by its values on the former. In particular, product states $\rho\otimes\sigma$ and mixtures $\omega=\sum_i p_i\rho_i\otimes\sigma_i$ thereof are well defined on $A\hat{\otimes}B$. If $A\subset B(H_1)$ and $B\subset B(H_2)$ are von Neumann algebras, and all states considered are normal, it is easier to work with the ***spatial*** tensor product $A\overline{\otimes}B$, defined as the double commutant (or weak completion) of $A\otimes B$ in $B(H_1\otimes H_2)$. Any *normal* state on $A\otimes B$ extends to a normal state on $A\overline{\otimes}B$ by continuity. Below we use $\hat{\otimes}$, but the results also work for $\overline{\otimes}$. In what follows, A and B are *unital* C*-algebras.

Definition 6.23. *Let ω be a state on $A\hat{\otimes}B$.*

1. *A* **product state** *is a state of the form $\omega=\rho\otimes\sigma$, i.e., ω is defined by linear (and continuous) extension of $\omega(a\otimes b)=\rho(a)\sigma(b)$.*
2. *A state ω is* **uncorrelated** *when it is in the w*-closure of the convex hull of the product states on $A\hat{\otimes}B$. In particular, states $\omega=\sum_i p_i\rho_i\otimes\sigma_i$, where $p_i>0$ and $\sum_i p_i=1$, are uncorrelated (w*-convergent infinite sums are allowed here).*
3. *A state is* **correlated** *when it is* not *uncorrelated.*

An uncorrelated state ω is pure precisely when it is a product of pure states. This has the important consequence that both its restrictions $\omega_{|A}$ and $\omega_{|B}$ to A and B, respectively, are pure as well (the restriction $\omega_{|A}$ of a state ω on $A\hat{\otimes}B$ to, say, A is given by $\omega_{|A}(a)=\omega(a\otimes 1_B)$, where 1_B is the unit element of B, etc.). A correlated *pure* state has the property that its restriction to A or B is *mixed*.

Proposition 6.24. *The following conditions are equivalent:*

- *Each state on $A\hat{\otimes}B$ is uncorrelated;*
- *Each pure state on $A\hat{\otimes}B$ is a product state;*
- *At least one of the C*-algebras A and B is commutative.*

For the proof see Takesaki (2002), Theorem 4.14.

Corollary 6.25. *Correlated states exist iff A and B are both noncommutative.*

As one might expect, this result is closely related to the Bell inequalities:

Proposition 6.26. *For any $\omega\in S(A\hat{\otimes}B)$, the following conditions are equivalent:*

- *ω is uncorrelated.*
- *For all self-adjoint operators $a_1,a_2\in A$ and $b_1,b_2\in B$ of norm ≤ 1 we have*

$$|\omega(a_1(b_1+b_2)+a_2(b_1-b_2))|\leq 2. \tag{6.249}$$

See Baez (1987), Raggio (1988), Bacciagaluppi (1993), and Landsman (2006a).

Corollary 6.27. *If A or B is commutative, then* (6.249) *holds for all states ω.*

An elegant geometric approach to the Bell inequalities was developed by Pitowsky (1989, 1994), which we now summarize (also cf. Werner & Wolf, 2001).

Suppose we have a bipartite experiment with m different settings $A = a_1, \ldots a_m$ and $B = b_1, \ldots, b_m$ on each wing, and binary outcomes, i.e., in $\{0,1\}$. We now denote the probability $P(F = 1|A = a_i)$ that $F(a_i)$ (i.e. the particular property measured by experiment F at setting a_i) is true by p_i ($i = 1, \ldots, m$), and likewise we write p_{j+m} for $P(G|B = b_j)$, i.e., the probability that $G(b_j)$ is true, once again for $j = 1, \ldots, m$. Furthermore, we abbreviate the probability that $F(a_i)$ and $G(b_j)$ are both true by

$$p_{i,j+m} \equiv P(F = 1, G = 1|A = a_i, B = b_j) \;\; (i, j = 1, \ldots, m). \tag{6.250}$$

The $2m+m^2$ "surface probabilities" $\mathbf{p} = (p_1, \ldots, p_{2m}, p_{1,m+1}, \ldots, p_{m,2m})$ form a vector in $\mathbb{R}^{2m+m^2}$, which we wish to constrain by the following assumption: there is a fact of the matter underlying each experiment according to which the pair $(F(a_i), G(b_j))$ already had a truth value for each possible setting (a_i, b_j), independently of any measurement being carried out or not ("***local realism***"). Thus the probabilities $\mathbf{p}$ (which now arguably have an ignorance interpretation) must lie in the convex polytope in $\mathbb{R}^{|2m+m^2|}$ defined as the convex hull C_m of the following set of (extreme) points: for each $2m$-tuple $\lambda = (\lambda_1, \ldots, \lambda_{2m})$, where $\lambda_i \in \{0,1\}$, define

$$\mathbf{x}_\lambda = (\lambda_1, \ldots, \lambda_{2m}, \lambda_1 \cdot \lambda_{m+1}, \ldots, \lambda_m \cdot \lambda_{2m}) \in \mathbb{R}^{2m+m^2}, \tag{6.251}$$

i.e., the entry at place k is λ_k ($k = 1, \ldots, 2m$) and the entry at place (i, j) is $\lambda_i \cdot \lambda_{m+j}$, where $i, j = 1, \ldots, m$. The interpretation of this is that $\mathbf{x}_\lambda$ represents the particular fact of the matter where $F(a_i)$ has truth value λ_i and $G(b_j)$ has truth value λ_{m+j}, so that their conjunction $(F(a_i), G(b_j))$ has truth value $\lambda_i \cdot \lambda_{m+j}$. In this state the probability of the said configuration is one and all other states have probability zero; arbitrary probability assignments then lie in C_m. The point, then, is to characterize the convex polytope $C_m \subset \mathbb{R}^{2m+m^2}$ through a finite set of inequalities, which turn out to be generalized Bell inequalities. Seeing this result requires some background.

Let V be a real topological vector space with (continuous) dual V^*; if $V = \mathbb{R}^n$ we may also put $V^* = \mathbb{R}^n$ and write $\varphi(v)$ as an inner product $\langle \varphi, v \rangle$ in what follows.

1. Any (not necessarily convex) subset $S \subset V$ has a ***polar*** $S^o \subset V^*$ defined by

$$S^o = \{\varphi \in V^* \mid \varphi(v) \leq 1 \, \forall v \in S\}, \tag{6.252}$$

which is a closed convex subset of V^*. If $S = K$ is a compact convex set, we have

$$K^o = \{\varphi \in V^* \mid \varphi(v) \leq 1 \, \forall v \in \partial_e K\}. \tag{6.253}$$

2. The ***bipolar theorem*** (cf. e.g. Simon (2011, Theorem 5.5) states that

$$S^{oo} = \mathrm{co}(S \cup \{0\}). \tag{6.254}$$

In particular, if K a closed convex set containing the origin, then

$$K^{oo} = K, \tag{6.255}$$

and hence, if K^o is a compact convex set, we may reconstruct K from K^o as

$$K = \{v \in V \mid \varphi(v) \leq 1 \, \forall \varphi \in \partial_e K^o\}. \tag{6.256}$$

3. In particular, if K is a convex polytope in a finite-dimensional vector space containing the origin, then so is K^o. In that case, $\partial_e K^o$ is a finite set and so points in K are characterized by a *finite* set of *linear* inequalities (6.256), which describe the faces of the polytope. In this case, the associated (dual) description of K is called the ***Minkowski–Weyl Theorem***, see e.g. Paffenholz (2010) for applications.

For example, among the five Platonic solids (i.e. in $\mathbb{R}^3$) the cube and the octahedron are dual to each other, as are the dodecahedron and the icosahedron, whereas the terahedron is self-dual. *A propos*, the latter arises as the convex polytope C_1 for $m = 1$ in the above story: clearly $2m + m^2 = 3$, and for the vertices of C_1 one takes the four points $\mathbf{x}_\lambda$ ensuing from the four possibilities $\lambda = (0,0), (1,0), (0,1), (1,1)$, i.e., $\mathbf{x}_\lambda = (0,0,0), (1,0,0), (0,1,0), (1,1,1)$. Then the inequalities in (6.256) are

$$p_{1,2} \geq 0, \;\; p_1 \geq p_{1,2}, \;\; p_2 \geq p_{1,2}, \;\; p_1 + p_2 - p_{1,2} \leq 1. \tag{6.257}$$

For $m = 2$ the ensuing convex polytope $C_2 \subseteq \mathbb{R}^8$ is the convex hull of 16 extreme points, whose inequalities may be found in Pitowsky (1989, p. 27); these imply the CHSH inequality, whose violation in quantum mechanics therefore shows that the probabilities in question have no local realistic model.

More generally, suppose we have n yes-no experiments $(E_1, \ldots, E_n)$ and some subset S_n of the set $\{(i,k) \mid 1 \leq i < k \leq n\}$ (above we had $n = 2m$, $E_i = F(a_i)$ for $i = 1, \ldots, m$, $E_{m+j} = G(b_j)$ for $j = 1, \ldots, m$, and $S_n = \{(i, m+j) \mid 1 \leq i, j \leq m\}$). This gives surface probabilities $(p_1, \ldots, p_n, p_{i,k})$, where $(i,k) \in S_n)$, which form a vector $\mathbf{p}$ in $\mathbb{R}^{n+|S_n|}$. As in (6.251), each truth assignment $\lambda = (\lambda_1, \ldots, \lambda_n)$, $\lambda \in \{0,1\}$, then defines a point $\mathbf{x}_\lambda \in \mathbb{R}^{n+|S_n|}$ with coordinates $(\lambda_1, \ldots, \lambda_n, \lambda_i \cdot \lambda_k)$, where once again $(i,k) \in S_n$. This set of 2^n points in turn spans a convex polytope C_{S_n} characterized by inequalities following from the dual characterization (6.256). Classical thinking would constrain the $\mathbf{p}$ so as to lie in C_{S_n}, and indeed we have $\mathbf{p} \in C_{S_n}$ iff there is a probability space (X, G, μ) such that $p_i = \mu(A_i)$ and $p_{i,k} = \mu(A_i \cap A_k)$ for certain events $A_i \in \Sigma$, cf. Theorem 2.3 in Pitowsky (1989), which is based on Fine (1982).

Some authors claim on this basis that Bell-type inequalities have nothing to do with physics, but surely the point is that some physical assumptions (notably local realism) have to be made in order to justify the "classical thinking" behind C_{S_n}.

§6.6. **The Colbeck–Renner Theorem**

This section is based on Colbeck & Renner (2011, 2012a, 2012b), where the main idea originates (alas with unclear assumptions and at best heuristic "proofs"), Braunstein & Caves (1990), who provided steps 1 and 2 of the proof, and Landsman (2015), whom we follow closely. See also Leegwater (2016) for a technically different approach (by a far more complicated argument, Leegwater seems to manage to do without our **CP** assumption, i.e., continuity of probabilities).