Chapter 2
Quantum mechanics on a finite-dimensional Hilbert space

The quantum analogue of a finite set X (in its role as a configuration space in classical mechanics) is the finite-dimensional Hilbert space $\ell^2(X)$, by which we mean the vector space of functions $\psi : X \to \mathbb{C}$, equipped with the inner product

$$\langle \psi, \varphi \rangle = \sum_{x \in X} \overline{\psi(x)} \phi(x). \tag{2.1}$$

There is no issue of convergence here, but later on we will use the same notation for infinite sets X, where $\ell^2(X)$ is restricted to those functions (i.e. sequences) for which $\sum_{x \in X} |\psi(x)|^2 < \infty$ (which also guarantees convergence of the sum in (2.1)).

If $X \cong \underline{n}$ as sets (i.e., $|X| = n$), we have a unitary isomorphism of Hilbert spaces

$$\ell^2(\underline{n}) \cong \mathbb{C}^n, \tag{2.2}$$

through the map $\psi \mapsto (\psi(1), \ldots, \psi(n))$, where $\mathbb{C}^n$ has the standard inner product. $\langle w, z \rangle = \sum_i \overline{w}_i z_i$. In particular, the function $\delta_k \in \ell^2(\underline{n})$, defined by $\delta_k(l) = \delta_{kl}$, is mapped to the k'th standard basis vector $u_k \equiv |k\rangle$ of $\mathbb{C}^n$, i.e., $u_1 = (1, 0, \ldots, 0)$, etc. In the special case $X = \underline{N}^{\Lambda}$ considered in Chapter 1, we have $|X| = N^{|\Lambda|}$ and hence

$$\ell^2(\underline{N}^{\Lambda}) \cong \mathbb{C}^{(N^{|\Lambda|})} \cong \left(\mathbb{C}^N\right)^{\otimes |\Lambda|} = \bigotimes_{\mathbf{n} \in \Lambda} \mathbb{C}^N_{\mathbf{n}} \equiv \bigotimes_{\Lambda} \mathbb{C}^N, \tag{2.3}$$

where $\mathbb{C}^N_{\mathbf{n}} = \mathbb{C}^N$ for each $\mathbf{n} \in \Lambda$, so that the suffix $\mathbf{n}$ merely labels which copy of $\mathbb{C}^N$ is meant (see §C.13 for tensor products of Hilbert spaces). Explicitly, a canonical unitary isomorphism $\ell^2(\underline{N}^{\Lambda}) \to \bigotimes_{\Lambda} \mathbb{C}^N$ is given by linear extension of the map

$$\delta_x \mapsto \otimes_{\mathbf{n} \in \Lambda} u_{x(\mathbf{n})}, \tag{2.4}$$

where $x : \Lambda \to \underline{N}$ and hence $u_{x(\mathbf{n})} \in \mathbb{C}^N$. Thus elements of the tensor product $\bigotimes_{\Lambda} \mathbb{C}^N$ may be seen as wave-functions on spin configuration space (and *vice versa*). In particular, elementary tensor products of basis vectors in $\bigotimes_{\Lambda} \mathbb{C}^M$ correspond to wave-functions in $\ell^2(\underline{M}^{\Lambda})$ that are δ-peaked at some 'classical' spin configuration.

K. Landsman, *Foundations of Quantum Theory*,
Fundamental Theories of Physics 188, DOI 10.1007/978-3-319-51777-3_2

2.1 Quantum probability theory and the Born rule

In preparation for this chapter, the reader would do well to review Appendix A.

The probabilistic setting of quantum mechanics is given by the following counterpart of Definition 1.1 (from which conditional probabilities are lacking, though).

Definition 2.1. *Let H be a finite-dimensional Hilbert space.*

1. *A* **(quantum) event** *is a linear subspace L of H (which is automatically closed).*
2. *A* **(quantum) probability distribution** *is a* **density operator***, i.e., a positive operator ρ on H (in that $\langle\psi,\rho\psi\rangle \geq 0$ for all $\psi \in H$) such that*

$$\mathrm{Tr}(\rho) = 1. \tag{2.5}$$

   We denote the set of all density operators on H by $\mathscr{D}(H)$.
3. *A* **(quantum) random variable** *is a self-adjoint operator a on H (i.e., $a^* = a$).*
4. *The* **spectrum** *of a self-adjoint operator a is the set $\sigma(a) \subset \mathbb{R}$ of its eigenvalues.*

Being positive, a density matrix ρ is self-adjoint, so by Theorem A.10, notably (A.40), and Definition 2.1.2 we have

$$\rho = \sum_i p_i|\upsilon_i\rangle\langle\upsilon_i|, \; p_i > 0, \; \sum_i p_i = 1, \tag{2.6}$$

where the (υ_i) form an orthonormal set in H and $|\upsilon_i\rangle\langle\upsilon_i|$ is the (orthogonal) projection on the one-dimensional subspace $\mathbb{C}\cdot\upsilon_i$. As in the classical case, one special class of density operators and one special class of random variables stand out:

- Each *unit vector* $\psi \in H$ defines a density operator

$$\rho_\psi \equiv e_\psi = |\psi\rangle\langle\psi|, \tag{2.7}$$

  i.e., the (orthogonal) projection e_ψ on the one-dimensional subspace $\mathbb{C}\cdot\psi$. A basis (which by convention always means an *orthonormal* basis) of eigenvectors of ρ_ψ consists of $\upsilon_1 = \psi$ itself, supplemented by any basis $(\upsilon_2,\ldots,\upsilon_{\dim(H)})$ of the orthogonal complement of $\mathbb{C}\cdot\psi$. The corresponding probabilities in (2.6) are evidently $p_1 = 1$ and $p_i = 0$ for all $i > 1$.
- Each *quantum event* $L \subset H$ defines the corresponding projection e_L (which is self-adjoint, i.e. a random variable): If (υ_j) is a basis of L, then $e_L = \sum_j |\upsilon_j\rangle\langle\upsilon_j|$. If $L = H$ then $e_L = 1$ with $\sigma(e_L) = \{1\}$. If $L = \{0\}$ then $e_L = 0$ with $\sigma(e_L) = \{0\}$. In all other cases, i.e. for proper subspaces L, one has $\sigma(e_L) = \{0,1\}$.
  Conversely, any self-adjoint operator a with spectrum $\sigma(a) \subseteq \{0,1\}$ is given by $a = e_L$ for some subspace $L \subseteq H$; just take $L = \{\psi \in H \mid a\psi = 1\}$. Such operators correspond to yes-no questions to the system and lie at the basis of the logical interpretation of quantum theory due to Birkhoff and von Neumann; see §2.10.

The following quantum analogue of Theorem 1.2 is based on Theorem A.10.

Theorem 2.2. *A density operator* ρ *on* H *and a self-adjoint operator* $a: H \to H$ *jointly yield a probability distribution* p_a *on the spectrum* $\sigma(a)$ *by the* **Born rule**

$$p_a(\lambda) = \mathrm{Tr}\,(\rho e_\lambda). \tag{2.8}$$

The associated probability measure P_a *is given at* $\Delta \subseteq \sigma(a)$ *by (cf.* (A.42)*)*

$$P_a(\Delta) = \mathrm{Tr}\,(\rho e_\Delta). \tag{2.9}$$

Proof. Positivity of the numbers $p_a(\lambda)$ follows by taking the trace over a basis of eigenvectors υ_i of ρ, with corresponding eigenvalues $p_i \geq 0$. This yields

$$\mathrm{Tr}\,(\rho e_\lambda) = \sum_i p_i \, \|e_\lambda \upsilon_i\|^2 \geq 0.$$

Eqs. (A.38) and (2.5) then give $\sum_\lambda p_a(\lambda) = 1$. Eq. (2.8) follows from the equality $P_a(\Delta) = \sum_{\lambda \in \Delta} p_a(\Delta)$, cf. (1.2), and (A.42). □

In particular, if $\rho = \rho_\psi$, writing p_a^ψ for the associated probability, (2.8) yields

$$p_a^\psi(\lambda) = \langle \psi, e_\lambda \psi \rangle = \|e_\lambda \psi\|^2. \tag{2.10}$$

If in addition $\lambda \in \sigma(a)$ is non-degenerate, so that $e_\lambda = |\upsilon_\lambda\rangle\langle\upsilon_\lambda|$ for some unit vector υ_λ with $a\upsilon_\lambda = \lambda\upsilon_\lambda$, then the Born rule (2.9) assumes its original form

$$p_a^\psi(\lambda) = |\langle \psi, \upsilon_\lambda \rangle|^2. \tag{2.11}$$

Specializing (2.10) to the random variable $a = e_L$ defined by an event $L \subset H$ yields

$$p_{e_L}^\psi(1) = \|e_L \psi\|^2. \tag{2.12}$$

If $L = \mathbb{C} \cdot \varphi$ is one-dimensional, too, in which case we write $p_{e_\varphi}^\psi \equiv p_\varphi^\psi$, we have

$$p_\varphi^\psi(1) = |\langle \psi, \varphi \rangle|^2; \tag{2.13}$$

note the following equality of probability distributions on $\sigma(e_\varphi) = \sigma(e_\psi) = \{0,1\}$:

$$p_\varphi^\psi(1) = p_\psi^\varphi(1). \tag{2.14}$$

Expectation values and variances may be defined as in the classical case, viz.

$$E_\rho(a) = \mathrm{Tr}\,(\rho a); \tag{2.15}$$

$$\Delta_\rho(a) = E_\rho(a^2) - E_\rho(a)^2. \tag{2.16}$$

Similar to (1.11), we may also write the expectation value as

$$E_\rho(a) = \sum_{\lambda \in \sigma(a)} \lambda \cdot p_a(\lambda). \tag{2.17}$$

The special case $\rho = \rho_\psi$, for which we write $E_{\rho_\psi} \equiv E_\psi$, gives the usual formula

$$E_\psi(a) = \mathrm{Tr}\,(\rho_\psi a) = \langle \psi, a\psi \rangle. \tag{2.18}$$

As in the classical case one always has $\Delta_\rho(a) \geq 0$, but a major contrast between classical and quantum mechanics lies in the following result, cf. Proposition 1.3.

Proposition 2.3. *For each density operator ρ there exists a self-adjoint operator b such that $\Delta_\rho(b) > 0$. On the other hand, if $a^* = a$, then $\Delta_\rho(a) = 0$ iff the image of ρ lies in some fixed eigenspace of a, i.e., in terms of the spectral decomposition* (2.6) *we have $a\upsilon_i = \lambda\upsilon_i$ where λ is independent of i.*

Proof. We first prove the first claim for $H = \mathbb{C}^2$. By an appropriate choice of basis, we may assume that ρ is diagonal, i.e., $\rho = \mathrm{diag}(p_1, p_2)$, with $p_1, p_2 \in [0,1]$ and $p_1 + p_2 = 1$. Now take $b = \sigma_x$ (i.e., the first Pauli matrix), so that $\mathrm{Tr}\,(\rho b) = 0$ and $\mathrm{Tr}\,(\rho b^2) = 1$. Hence $\Delta_\rho(b) = 1$. Secondly, for general $H \cong \mathbb{C}^n$, diagonalize ρ and order the eigenvectors such that the above 2×2 case forms the upper left block, with at least one of the eigenvalues p_1, p_2 strictly positive. Take b to be σ_x in the upper left corner, and zero elsewhere. This once again yields $\Delta_\rho(b) = 1$.

For the second claim we use (2.6), and write $\rho_i \equiv \rho_{\upsilon_i}$. We note the inequality

$$\Delta_\rho(a) \geq \sum_i p_i \Delta_{\rho_i}(a), \tag{2.19}$$

with equality iff $\rho_i(a) = \rho_j(a)$ for all i, j; this follows from convexity of the function $x \mapsto x^2$. We now show that for any unit vector ψ we have $\Delta_{\rho_\psi} = 0$ iff $a\psi = \lambda\psi$. Assuming the latter gives $E_\psi(a) = \langle\psi, a\psi\rangle = \lambda$ and likewise $E_\psi(a^2) = \lambda^2$, hence $\Delta_{\rho_\psi}(a) = 0$. In the opposite direction, using $a^* = a$, elementary manipulations yield

$$\Delta_{\rho_\psi}(a) = \|(a - \langle\psi, a\psi\rangle)\psi)\|^2. \tag{2.20}$$

This clearly vanishes iff $a\psi = \langle\psi, a\psi\rangle\psi$, so $a\psi = \lambda\psi$, with $\lambda = \langle\psi, a\psi\rangle$.

Putting $\psi = \upsilon_i$ gives $\Delta_{\rho_i} = 0$ iff $a\upsilon_i = \lambda_i\upsilon_i$, and then $\Delta_{\sum_i p_i\rho_i}(a) = 0$ iff in addition $\rho_i(a) = \rho_j(a)$ for all i, j. Since $\rho_i(a) = \langle\upsilon_i, a\upsilon_i\rangle = \lambda_i$, we obtain $\lambda_i = \lambda_j$. □

As first recognized by von Neumann, Theorem 2.2 may be generalized to a family of self-adjoint operators *as long as they commute*. Thus we obtain the following counterpart of (1.12) - (1.13): a collection $a_1, \ldots, a_n$ of n *commuting* self-adjoint operators and a (single) density operator ρ on H jointly define a probability distribution $p_{a_1,\ldots,a_n}$ on the product $\sigma(a_1) \times \cdots \times \sigma(a_n)$ of the individual spectra by

$$p_{a_1,\ldots,a_n}(\lambda_1, \ldots, \lambda_n) = \mathrm{Tr}\,(\rho e^{(1)}_{\lambda_1} \cdots e^{(n)}_{\lambda_n}). \tag{2.21}$$

The proof of positivity of these numbers requires the spectral projections $e^{(i)}_{\lambda_i}$ to commute, which they do provided the a_i commute (if the a_i fail to commute, positivity of (2.21) is not guaranteed, although they do still sum op to unity; the possibility of defining joint probabilities is strictly limited to commuting random variables).

2.2 Quantum observables and states

Given a finite-dimensional Hilbert space H, the set $B(H)$ of all linear operators on H (which for $H = \mathbb{C}^n$ may be identified with the set $M_n(\mathbb{C})$ of complex $n \times n$ matrices) forms an involutive algebra under the natural (pointwise) operations

$$(\lambda \cdot a)\psi = \lambda(a\psi); \tag{2.22}$$
$$(a+b)\psi = a\psi + b\psi; \tag{2.23}$$
$$(ab)\psi = a(b\psi), \tag{2.24}$$

and finally with a^* given by the usual operator adjoint (A.15). Compare the corresponding classical expressions (1.18) - (1.20) and (1.22). Analogous to (1.24), we also have a norm on $B(H)$, defined by (A.18). It follows that *like* its classical counterpart $C(X)$, the involutive algebra $B(H)$ (or, in this case, $M_n(\mathbb{C})$) is a C*-algebra, cf. Definition C.1 in Appendix C. It crucially *differs* from $C(X)$ in that $B(H)$ is *non-commutative*. For this reason, the Gelfand spectrum, which in the classical case allowed us to reconstruct X from $C(X)$, turns out to be empty, cf. Proposition 2.10 below. Nonetheless, it makes good sense to copy Definition 1.14, *mutatis mutandis*:

Definition 2.4. *A* **state** *on $B(H)$ is a complex-linear map $\omega : B(H) \to \mathbb{C}$ satisfying:*

1. $\omega(a^*a) \geq 0$ *for each $a \in B(H)$ (***positivity***);*
2. $\omega(1_H) = 1$ *(***normalization***).*

The **state space** *$S(B(H))$ is the set of all states $\omega : B(H) \to \mathbb{C}$.*

Physicists may not like this definition, since it involves non-observable quantities. As in the classical case, we may introduce the self-adjoint (or ‘real’) part of $B(H)$:

$$B(H)_{\rm sa} = \{a \in B(H) \mid a^* = a\}, \tag{2.25}$$

which is a real vector space (though not a real algebra in the usual sense, cf. §C.25).

Definition 2.5. *A* **state** *on $B(H)_{\rm sa}$ is a real-linear map $\omega : B(H)_{\rm sa} \to \mathbb{R}$ satisfying:*

1. $\omega(a^2) \geq 0$ *for each $a \in B(H)$ with $a^* = a$ (***positivity***);*
2. $\omega(1) = 1$ *(***normalization***).*

The **state space** *$S(B(H)_{\rm sa})$ is the set of all states $\omega : B(H)_{\rm sa} \to \mathbb{R}$.*

Fortunately, there is no need for a fight over this point; the discussion is similar to the one below Definition 1.14 and is settled as follows.

Proposition 2.6. *The state spaces $S(B(H))$ and $S(B(H)_{\rm sa})$ may be identified: an element ω of the former defines an element $\omega_{\mathbb{R}}$ of the latter by restriction, whilst the unique decomposition $c = a + ib$ (where $a^* = a$ and $b^* = b$ are given by $a = \frac{1}{2}(c + c^*)$ and $b = -\frac{1}{2}i(c - c^*)$, respectively) gives $\omega(c) = \omega_{\mathbb{R}}(a) + i\omega_{\mathbb{R}}(b)$. Moreover,*

$$\|\omega\| = \|\omega_{\mathbb{R}}\| = 1. \tag{2.26}$$

Here the norm on the dual (Banach) space $B(H)^*_{\mathrm{sa}}$ of $B(H)_{\mathrm{sa}}$ is given by

$$\|\omega\| = \sup\{|\omega(a)|, a \in B(H)_{\mathrm{sa}}, \|a\| = 1\}. \tag{2.27}$$

This lemma holds for any Hilbert space H (cf. Theorem C.52), but it is instructive to restrict our proof to the finite-dimensional setting in which we currently work.

Proof. The first few claims are immediate from Proposition A.22. To prove (2.26), it suffices to prove that for any $a \in B(H)$ one has

$$|\omega(a)| \leq \|a\|, \tag{2.28}$$

since by normalization of states the bound is saturated by $a = 1_H$. Furthermore, even if ω is seen as an element of $B(H)^*$ rather than $B(H)^*_{\mathrm{sa}}$, eq. (2.28) needs to be shown only for self-adjoint a, for positivity of ω implies the Cauchy–Schwarz inequality

$$|\omega(a^*b)|^2 \leq \omega(a^*a)\omega(b^*b), \tag{2.29}$$

cf. (A.1), in which we may take $a = 1_H$ to find, assuming (2.28) for self-adjoint a,

$$|\omega(b)|^2 \leq \omega(b^*b) \leq \|b^*b\| = \|b\|^2, \tag{2.30}$$

where the last equality holds for any $b \in B(H)$ (turning the latter into a C*-algebra). Noting that b^*b is self-adjoint, this gives (2.28) for any a. To prove (2.28) for $a^* = a$, then, we firstly use (A.47), and secondly use Theorem 2.7 and eq. (2.6) to obtain

$$|\omega(a)| = |\mathrm{Tr}\,(\rho a)| = |\sum_i p_i \langle \upsilon_i, a\upsilon_i\rangle| \leq \sum_i p_i |\langle \upsilon_i, a\upsilon_i\rangle|. \tag{2.31}$$

Now let (ξ_j) be a basis of H consisting of eigenvectors of a, so that

$$\langle \upsilon_i, a\upsilon_i\rangle = \sum_j |\langle \upsilon_i, \xi_j\rangle|^2 \lambda_j, \; \sum_j |\langle \upsilon_i, \xi_j\rangle|^2 = 1.$$

Since $|\lambda_j| \leq \|a\|$ and $\sum_i p_i = 1$, the bound (2.28) follows from the estimate

$$\sum_i p_i |\langle \upsilon_i, a\upsilon_i\rangle| \leq \sum_i p_i \sum_j |\langle \upsilon_i, \xi_j\rangle|^2 |\lambda_j| \leq \sum_i p_i \sum_j |\langle \upsilon_i, \xi_j\rangle|^2 \|a\| = \|a\|. \tag{2.32}$$

Finally, combining (2.31) and (2.32) gives (2.28) for self-adjoint a. □

In view of this, we may work with either $S(B(H)_{\mathrm{sa}})$ or $S(B(H))$; denoting states simply by ω, the context will usually show if it is defined on $B(H)_{\mathrm{sa}}$ or on $B(H)$.

Despite its easy proof, the following result is of fundamental importance.

Theorem 2.7. *If H is finite-dimensional, there is a bijective correspondence between states ω on $B(H)$ or $B(H)_{\mathrm{sa}}$ and density operators ρ on H, given by*

$$\omega(a) = \mathrm{Tr}\,(\rho a). \tag{2.33}$$

Proof. First note that linear algebra already yields (2.33) as a bijective correspondence between complex-linear maps ω and operators ρ, for example, because

$$\langle a,b\rangle = \mathrm{Tr}\,(a^*b) \tag{2.34}$$

defines an inner product on $B(H)$. Positivity and normalization of ω then translate to the corresponding properties of ρ. $\square$

The quantum analogue of Theorem 1.15, then, is as follows.

Theorem 2.8. *The state space $S(B(H)_{\mathrm{sa}}) = S(B(H))$ forms a compact convex set in the (real) vector space $B(H)^*_{\mathrm{sa}}$ (in its w^*-topology) and, putting the corresponding topology on $\mathscr{D}(H)$, eq.* (2.33) *defines an affine homeomorphism*

$$S(B(H)) \cong \mathscr{D}(H). \tag{2.35}$$

Proof. Convexity of $S(B(H))$ holds by Definition 2.4. For compactness, by Proposition 2.6 the state space $S(B(H))$ is contained in the closed unit ball B_1 of $B(H)^*_{\mathrm{sa}}$, which is compact in the w^*-topology (in the case at hand this is simply because $B(H)^*_{\mathrm{sa}}$ is finite-dimensional). It is easy to see that a convergent sequence of states actually converges to a state, since both conditions in Definition 2.4 are clearly preserved by w^* limits (in which $\omega_n \to \omega$ iff $\omega_n(a) \to \omega(a)$ for each $a \in B(H)$). $\square$

For infinite-dimensional Hilbert spaces eq. (2.35) is false; see §4.2. At the opposite end, the case $H = \mathbb{C}^2$ provides a beautiful illustration of this theorem (and more).

Proposition 2.9. *The state space $S(M_2(\mathbb{C}))$ of the 2×2 matrices is isomorphic (as a compact convex set) to the closed unit ball $B^3 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 \leq 1\}$.*

On this isomorphism, the extreme boundary (cf. Definition 1.10)

$$\partial_e B^3 = S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\} \tag{2.36}$$

corresponds to the set of all density matrices $\rho = \rho_\psi$, where $\psi \in \mathbb{C}^2$ with $\|\psi\| = 1$.

Proof. Any self-adjoint 2×2 matrix may be parametrized by $(t,x,y,z) \in \mathbb{R}^4$ as

$$\rho(t,x,y,z) = \tfrac{1}{2}\begin{pmatrix} t+z & x-iy \\ x+iy & t-z \end{pmatrix}. \tag{2.37}$$

The eigenvalues λ_i of $\rho(t,x,y,z)$, computed from its characteristic polynomial, are

$$\lambda_\pm = \tfrac{1}{2}(t \pm \sqrt{x^2+y^2+z^2}). \tag{2.38}$$

Condition (2.5) yields $t = 1$. Positivity of $\rho(1,x,y,z)$ is equivalent to positivity of its eigenvalues λ_i, which gives $x^2+y^2+z^2 \leq 1$. For the second claim, note that the ρ_ψ are just the one-dimensional projections, which in turn are the density matrices satisfying $\rho^2 = \rho$ (or require $\lambda_+ = 1$, $\lambda_- = 0$), so $x^2+y^2+z^2 = 1$. Finally, since convex sums $t\mathbf{v} + (1-t)\mathbf{w}$ in B^3 $(0 \leq t \leq 1)$ are given by straight line segments connecting $\mathbf{w}$ and $\mathbf{v}$ in $\mathbb{R}^3$, it immediately follows geometrically that $\partial_e B^3 = S^2$. $\square$

2.3 Pure states in quantum mechanics

In classical physics, the phase space X arose both as the Gelfand spectrum $\Sigma(C(X))$ of the C*-algebra of observables $C(X)$, cf. Definition 1.4 and Proposition 1.5, and as the pure state space $P(C(X))$ of $C(X)$, see Definition 1.10 and Theorem 1.16. In particular, $\Sigma(C(X)) \cong P(C(X))$ at least as sets. Because of this, any pure state $\omega \in P(C(X))$ is dispersion-free, since as an element of $\Sigma(C(X))$ it satisfies $\omega(f^2) = \omega(f)^2$ for any $f \in C(X)$. These two definitionally different (but classically coinciding) guises of X will fall apart in quantum mechanics; cf. Proposition 2.3.

Proposition 2.10. *If* $\dim(H) > 1$, *the Gelfand spectrum* $\Sigma(B(H))$ *of* $B(H)$ *is empty, i.e., there are no nonzero linear maps* $\omega : B(H) \to \mathbb{C}$ *that satisfy* $\omega(ab) = \omega(a)\omega(b)$.

In particular, there are no nonzero linear maps $\omega : B(H) \to \mathbb{C}$ *that are* **dispersion-free**, *i.e., satisfy* $\Delta_\omega(a) = 0$, *with* $\Delta_\omega(a) = \omega(a^2) - \omega(a)^2$.

Proof. Suppose $\omega \in \Sigma(B(H))$. Multiplicativity for $b = a = a^*$ implies that ω is positive, whereas for $b = 1_H$ it implies that ω is normalized. Hence ω must be a state. Now use Theorem 2.7 and use multiplicativity for $b = a = a^*$, implying that $\Delta_\rho(a) = 0$. This contradicts Proposition 2.3. □

On the other hand, the pure state space of $B(H)$ is by no means empty, and despite Proposition 2.10, we will see that the special density operators $\rho_\psi \equiv e_\psi$ in (2.7) to some extent do play the role of the points $x \in X$. Let us write

$$\mathscr{P}_1(H) = \{e \in B(H) \mid e^2 = e^* = e, \mathrm{Tr}\,(e) = 1\} \tag{2.39}$$

for the set of all one-dimensional projections on H; note that $\mathrm{Tr}\,(e) = \dim(eH)$ for $e \in \mathscr{P}(H)$. Each $e \in \mathscr{P}_1(H)$ takes the form $e = e_\psi$ for some unit vector ψ, see (2.7).

Lemma 2.11. *A density operator* ρ *is an extreme point of the convex set* $\mathscr{D}(H)$ *of all density operators on* H *iff* $\rho = \rho_\psi$ *for some unit vector* $\psi \in H$.

Proof. The argument is similar to the proof of Proposition 1.11. To show that $\rho_\psi \in \partial_e S(B(H))$, assume $\rho_\psi = t\rho_1 + (1-t)\rho_2$ for some $t \in (0,1)$ and $\rho_1, \rho_2 \in S(B(H))$. Evaluating this equality at $a = |\varphi\rangle\langle\varphi|$, where $\varphi \perp \psi$ yields $\langle\varphi, \rho_i\varphi\rangle = 0$ for $i = 1,2$, so that $\rho_1 = \rho_2 = \rho_\psi$. Conversely, the spectral decomposition (2.6) shows that $\rho \notin \partial_e S(B(H))$ whenever $\rho \neq \rho_\psi$ for some unit vector $\psi \in H$. □

Consequently, for the moment just as sets (and even as topological spaces), one has

$$P(\mathscr{D}(H)) = \mathscr{P}_1(H); \tag{2.40}$$

$$P(B(H)) \cong \mathscr{P}_1(H), \tag{2.41}$$

where the second isomorphism is given by (2.33). Defining a state ω_ψ by

$$\omega_\psi(a) = \langle\psi, a\psi\rangle, \tag{2.42}$$

cf. (2.18), the isomorphism (2.41) is the correspondence $\omega_\psi \leftrightarrow e_\psi$, cf. (2.7).

This isomorphism becomes more interesting if we note that both spaces are naturally equipped with ***transition probabilities***. For $P(B(H))$ we canonically have

$$\tau^{B(H)}(\omega_\psi,\omega_\varphi)=\inf\{\omega_\psi(a)\mid a\in B(H), 0\leq a\leq 1_H, \omega_\varphi(a)=1\}, \tag{2.43}$$

as in (1.38) for $A=B(H)$. Furthermore, on $\mathscr{P}_1(H)$ we define (with some foresight)

$$\tau^{\mathscr{P}_1(H)}(e,f)=\mathrm{Tr}\,(ef). \tag{2.44}$$

Theorem 2.12. *The pairs $(P(B(H)),\tau^{B(H)})$ and $(\mathscr{P}_1(H),\tau^{\mathscr{P}_1(H)})$ are isomorphic as sets with a transition probability. In particular, we have, cf.* (2.13)*,*

$$\tau^{B(H)}(\omega_\psi,\omega_\varphi)=|\langle\psi,\varphi\rangle|^2=\mathrm{Tr}\,(e_\psi e_\varphi)=\tau^{\mathscr{P}_1(H)}(e_\psi,e_\varphi). \tag{2.45}$$

Proof. The last equality is a simple computation. The first follows if we can show that the infimum in (2.43) is reached at $a=e_\varphi$. To this end, we prove that for any $0\leq a\leq 1_H$ with $\omega_\varphi(a)=1$ we must have $\langle\psi,a\psi\rangle\geq|\langle\varphi,\psi\rangle|^2$. Indeed, the condition $\omega_\varphi(a)=\langle\varphi,a\varphi\rangle=1$ with $\|a\|\leq 1$ (which follows from $0\leq a\leq 1_H$) and $\|\varphi\|=1$ imply, by Cauchy–Schwarz, that $a\varphi=\varphi$. Since $a^*=a$ (by positivity of a), we also have $a:(\mathbb{C}\cdot\varphi)^\perp\to(\mathbb{C}\cdot\varphi)^\perp$, so we may write $a=e_\varphi+a'$, with $a'\varphi=0$ and a' mapping $(\mathbb{C}\cdot\varphi)^\perp$ to itself. Then $a\geq 0$ implies $a'\geq 0$. If $\langle\psi,a\psi\rangle<|\langle\varphi,\psi\rangle|^2$, then $\langle\psi,a'\psi\rangle<0$, which contradicts positivity of a' (and hence of a). □

The theory of observables and spectral resolutions of the kind (1.45) may be worked out completely for the "quantum" transition probabilities in this theorem:

Proposition 2.13. *1. There is a bijective correspondence between self-adjoint operators $a\in B(H)$ and observables f on $(\mathscr{P}_1(H),\tau^{\mathscr{P}_1(H)})$ à la Definition 1.18.6:*

- *Given a self-adjoint operator a, define an observable f_a at $e_\psi\in\mathscr{P}_1(H)$ by*

$$f_a(e_\psi)=\mathrm{Tr}\,(e_\psi a)=\langle\psi,a\psi\rangle; \tag{2.46}$$

- *Given an observable $f=\sum_i c_i\tau_{e_i}^{\mathscr{P}_1(H)}$, define an operator a_f by*

$$a_f=\sum_i c_i e_i. \tag{2.47}$$

2. Each such observable $f=f_a$ has a unique spectral resolution as in (1.45)*, i.e.,*

$$f_a=\sum_{\lambda\in\sigma(a)}\lambda\cdot\tau_{S_\lambda}, \tag{2.48}$$

where S_λ is the (automatically orthoclosed) subset of $\mathscr{P}_1(H)$ whose elements e satisfy $eH\subseteq H_\lambda$, where $H_\lambda\subseteq H$ is the eigenspace for the eigenvalue $\lambda\in\sigma(a)$.

3. The product defined by (1.46) - (1.47) *is equal to*

$$f_a^2=f_{a^2}; \tag{2.49}$$

$$f_a\circ f_b=f_{(ab+ba)/2}. \tag{2.50}$$

Proof. Any spectral decomposition $a = \sum_i \lambda_i |\upsilon_i\rangle\langle\upsilon_i|$ puts f_a as defined in (2.46) in the general form (1.44), with $c_i = \lambda_i$ and $y_i = e_{\upsilon_i}$. The rest should be clear. □

We now turn to the quantum counterpart of Proposition 1.13. The main difference is that although extremal decompositions of mixed states into pure ones always exist, they are no longer unique. For example, for $H = \mathbb{C}^2$, we have

$$\rho \equiv \text{diag}(2/3, 1/3) = \tfrac{2}{3}\rho_{u_1} + \tfrac{1}{3}\rho_{u_2} = \tfrac{1}{2}(\rho_{\xi_1} + \rho_{\xi_2}),$$

where (u_1, u_2) is the standard basis of $\mathbb{C}^2$, and

$$\xi_1 = (\sqrt{2/3}, \sqrt{1/3}),\ \xi_2 = (\sqrt{2/3}, -\sqrt{1/3}).$$

More generally, take any basis (w_i) of $H \cong \mathbb{C}^n$, assume (2.6), and for each i for which $\sqrt{\rho}w_i \neq 0$ (where $\sqrt{\rho} = \sum_i \sqrt{p_i}|\upsilon_i\rangle\langle\upsilon_i|$), define $t_i = \|\sqrt{\rho}w_i\|^2$, as well as the unit vector $\xi_i = \sqrt{\rho}w_i/\|\sqrt{\rho}w_i\|$. Then $\rho = \sum_i t_i \rho_{\xi_i}$ is an extremal decomposition of ρ. The above example corresponds to the special case $t_1 = t_2 = 1/2$, with

$$n = 2,\ p_1 = 2/3,\ p_2 = 1/3,\ w_1 = (1/\sqrt{2}, 1/\sqrt{2}),\ w_2 = (1/\sqrt{2}, -1/\sqrt{2}).$$

One might require the ξ_i to be mutually orthogonal, but even that does not imply uniqueness of the extremal decomposition: take, for example, $\rho = (1/n) \cdot 1_n$, where 1_n is the $n \times n$ unit matrix on $H = \mathbb{C}^n$. Then any basis induces (2.6).

Nonetheless, under appropriate assumptions uniqueness does follow.

Proposition 2.14. *1. Any density operator ρ on H has an extremal decomposition*

$$\rho = \sum_{i=1}^{m} p_i \rho_{\psi_i}, \tag{2.51}$$

where $m \leq \dim(H)$, the p_i are probabilities, and the ψ_i are distinct unit vectors.

2. This decomposition can be chosen such that the ψ_i are mutually orthogonal, in which case it is unique iff each of the non-zero eigenvalues of ρ is simple.

Proof. The existence of the *extremal* decomposition (2.51) of ρ follows from its *spectral* decomposition (2.6), which also proves claim 2. If ρ has some degenerate non-zero eigenvalue, the example just given yields non-uniqueness of (2.51). For the converse direction, use uniqueness of the decomposition (2.6) under the condition that each of the non-zero eigenvalues of ρ is simple. □

In the light of Theorem 2.7, it would be interesting to reformulate Proposition 2.14 directly in terms of the states on $B(H)$; note our standing assumption $\dim(H) < \infty$!

Proposition 2.15. *1. Any state ω on $B(H)$ has an extremal decomposition*

$$\omega = \sum_{i=1}^{m} p_i \omega_i, \tag{2.52}$$

into distinct pure states $\omega_i \in P(B(H))$, where $m \leq \dim(H)$, $p_i > 0$, and $\sum_i p_i = 1$.

2. *The unit vectors* ψ_i *that correspond to the pure states* ω_i *in* (2.52) *via* (2.42) *are mutually orthogonal (and hence are part or all of a basis of H) iff*

$$\|\omega_i - \omega_j\| = 2 \ (i \neq j). \tag{2.53}$$

3. *Extremal decompositions* (2.52) *satisfying* (2.53) *exist and correspond bijectively to orthogonal families* (e_i) *of one-dimensional projections on H (i.e.,* $e_i e_j = \delta_{ij} e_i$ *and* $\mathrm{Tr}\,(e_i) = 1$, *respectively) for which* $\omega(e_i) > 0$, $\sum_i \omega(e_i) = 1$, *and*

$$\omega(ae_i) = \omega(e_i a), \ a \in B(H). \tag{2.54}$$

In terms of such a family, the decomposition (2.52) *is given by*

$$p_i = \omega(e_i); \tag{2.55}$$

$$\omega_i(a) = \frac{\omega(ae_i)}{\omega(e_i)}. \tag{2.56}$$

Hence an extremal decomposition (2.52) *with all* ω_i *mutually orthogonal in the sense of* (2.53) *is unique iff the family* (e_i) *with the above properties is.*

Proof. Claim 1 clearly follows from no. 3. To prove (2.53), assume (2.42), so that

$$\|\omega_i - \omega_j\| = \sup\{|\langle\psi_i, a\psi_i\rangle - \langle\psi_j, a\psi_j\rangle|, a \in B(H), \|a\| = 1\}. \tag{2.57}$$

Clearly, $|\langle\psi, a\psi\rangle| \leq 1$ when $\|a\| = \|\psi\| = 1$, hence $|\langle\psi_i, a\psi_i\rangle - \langle\psi_j, a\psi_j\rangle| \leq 2$, and the upper bound $\|\omega_i - \omega_j\| = 2$ in (2.57) is reached iff $|\langle\psi_1, a\psi_1\rangle| = 1$ and $\langle\psi_2, a\psi_2\rangle = -\langle\psi_1, a\psi_1\rangle$. By Cauchy–Schwarz, this holds iff $a\psi_1 = \lambda\psi_1$ as well as $a\psi_2 = -\lambda\psi_2$ for some $\lambda \in \mathbb{T}$. If $\psi_i \perp \psi_j$, then this is accomplished by the operator $a = |\psi_i\rangle\langle\psi_i| - |\psi_j\rangle\langle\psi_j|$; note that $\sigma(a) = \{-1, 1\}$ for $\dim(H) = 2$ and $\sigma(a) = \{-1, 0, 1\}$ for $\dim(H) > 2$, so indeed $\|a\| = 1$ by (A.47). If, on the other hand, $\langle\psi_i, \psi_j\rangle \neq 0$, then no a with $\|a\| = 1$ can meet these eigenvalue equations. One way to see this is to reduce to $H = \mathbb{C}^2$, since a in (2.57) can be replaced by eae, where e is the projection onto the linear span of ψ_i and ψ_j. Picking a basis of $\mathbb{C}^2$ (with say $\upsilon_1 = \psi_1$), the two eigenvalue equations for a yield a matrix representation of a, from which $\|a\|^2 = \|a^* a\|$ may be computed by calculating the eigenvalues of $a^* a$ and using (A.47). This gives $\|a\| > 1$ unless $\langle\psi_i, \psi_j\rangle = 0$.

One direction of the proof of the third claim easily follows from Theorem 2.7: any spectral decomposition (2.6) of ρ provides the projections

$$e_i = |\upsilon_i\rangle\langle\upsilon_i| \tag{2.58}$$

of the proposition. For example, eq. (2.54) comes down to $[\rho, e_i] = 0$, which is the case iff e_i commutes with all spectral projections of ρ, which clearly holds for (2.58). Uniqueness of the e_i then corresponds to uniqueness of (2.6) and hence to non-degeneracy of the non-zero eigenvalues p_i of ρ, as in Proposition 2.14.

The opposite direction, i.e., proving that (2.58) exhausts all possibilities for (2.53) - (2.54), is based on the GNS-construction and requires an entire subsection.

2.4 The GNS-construction for matrices

The proof of Proposition 2.15 may be completed on the basis of the GNS-construction began in §1.5, which in this subsection we develop for $A = B(H)$, where, as usual, $\dim(H) < \infty$. In that case, we may use Theorem 2.7 to simplify matters.

First, to prove (1.76) we use (2.33) and cyclicity of the trace, compute the trace by summing over a basis (υ_i) of eigenvectors of a^*a, say $a^*a\upsilon_i = \mu_i\upsilon_i$, where $\mu_i \geq 0$ by positivity of a^*a, and use (A.47) (for a^*a rather than a) to obtain:

$$\begin{aligned}\omega(b^*a^*ab) &= \mathrm{Tr}\,(\rho b^*a^*ab) = \sum_i \langle \upsilon_i, b\rho b^*a^*a\upsilon_i\rangle = \sum_i \mu_i \langle \upsilon_i, b\rho b^*\upsilon_i\rangle \\ &\leq \|a^*a\| \sum_i \langle \upsilon_i, b\rho b^*\upsilon_i\rangle = \|a\|^2 \mathrm{Tr}\,(\rho b^*b) = \|a\|^2\omega(b^*b),\end{aligned}$$

where we used $\langle \upsilon_i, b\rho b^*\upsilon_i\rangle = \langle b^*\upsilon_i, \rho b^*\upsilon_i\rangle \geq 0$ to justify the inequality.

We now explain all cases of interest, paying special attention to the ***commutant***

$$\pi_\omega(A)' = \{B \in B(H_\omega) \mid \pi_\omega(a)B = B\pi_\omega(a)\,\forall a \in A\}; \tag{2.59}$$

to distinguish operators on H from operators on H_ω, we write the latter in capitals. For simplicity we also put $H = \mathbb{C}^n$ (with the standard inner product), so that

$$B(H) = M_n(\mathbb{C}), \tag{2.60}$$

and all operators are matrices. Performing a suitable unitary transformation or change of basis if necessary, we also assume that the unit vectors υ_i in the spectral decomposition (2.6) of ρ form (all or part of) the standard basis $(\upsilon_1, \ldots, \upsilon_n)$ of $\mathbb{C}^n$. As in (1.74), we denote the null space by

$$N_\rho = \{a \in B(H) \mid \mathrm{Tr}\,(\rho a^*a) = 0\}. \tag{2.61}$$

- If $\rho = |\upsilon_j\rangle\langle\upsilon_j|$, the corresponding pure state (2.42) is $\omega(a) = \langle \upsilon_j, a\upsilon_j\rangle$, with

$$N_\rho = \{a \in A \mid a\upsilon_j = 0\}. \tag{2.62}$$

Hence $a \in N_\rho$ iff the j'th column $C_j(a)$ of a vanishes, so we have $a - b \in N_\rho$ iff $C_j(a) = C_j(b)$. Thus the equivalence class $a_\rho \in M_n(\mathbb{C})/N_\rho$ may be identified with $C_j(a)$. Consequently, we obtain

$$H_\rho = M_n(\mathbb{C})/N_\rho \cong \mathbb{C}^n, \tag{2.63}$$

under the unitary isomorphism $u : H_\rho \to \mathbb{C}^n$, $a_\rho \mapsto C_j(a)$, with inverse $u^{-1} : z \mapsto a_\rho$, $z \in \mathbb{C}^n$, where a is the matrix with $C_j(a) = z$ and zeros elsewhere (i.e., $a_{ij} = z_i$ and $a_{ik} = 0$ for all i and $k \neq j$). We likewise write $u^{-1}w = b_\rho$, with $b_{ij} = w_i$ and $b_{ik} = 0$ for all i and $k \neq j$. With $ua_\rho = z$ and $ub_\rho = w$, we obtain (beware: no sum over j!):

$$\langle a_\rho, b_\rho\rangle = \mathrm{Tr}\,(\rho a^*b) = \sum_i \overline{a_{ij}}b_{ij} = \sum_i \overline{z}_i w_i = \langle z, w\rangle_{\mathbb{C}^n} = \langle ua_\rho, ub_\rho\rangle_{\mathbb{C}^n}.$$

The GNS-representation π_ρ, originally given on H_ρ by (1.77), is accordingly transformed to $u\pi_\rho(a)u^{-1} \equiv \hat{\pi}_\rho$ on $\mathbb{C}^n$, which is given by

$$\hat{\pi}_\rho(a)w = u\pi_\rho(a)b_\rho = u(ab)_\rho = C_j(ab) = aw,$$

and the cyclic vector $u\Omega_\rho \in \mathbb{C}^n$ is just the basis vector υ_j from which we started. More generally, for a pure state (2.42) the GNS-representation $\pi_{\omega_\psi}(M_n(\mathbb{C}))$ is equivalent to the defining representation on $\mathbb{C}^n$, with canonical cyclic vector ψ. Finally, since only multiples of the unit matrix commute with all matrices, it follows that

$$\pi_{\omega_\psi}(M_n(\mathbb{C}))' \cong \mathbb{C}. \tag{2.64}$$

• The 'opposite' case occurs when ρ is *invertible*, in other words, when the sum over i in (2.6) has n nonzero terms. Hence

$$\mathrm{Tr}\,(\rho a^* a) = \sum_{i=1}^{n} p_i \|a\upsilon_i\|^2 \tag{2.65}$$

vanishes iff $a\upsilon_i = 0$ for each i, i.e., $a = 0$, so that $N_\rho = \{0\}$ and hence

$$H_\rho = M_n(\mathbb{C}). \tag{2.66}$$

The GNS-constructed inner product on $M_n(\mathbb{C})$, cf. (1.78), given by

$$\langle a_\rho, b_\rho \rangle = \mathrm{Tr}\,(\rho a^* b), \tag{2.67}$$

may be transformed into the usual one (2.34) by the following linear map:

$$u : M_n(\mathbb{C}) \to M_n(\mathbb{C}); \tag{2.68}$$

$$ua_\rho = a_\rho \rho^{1/2}. \tag{2.69}$$

This map is unitary from the Hilbert space $(M_n(\mathbb{C}), \langle\cdot,\cdot\rangle_\rho)$ to the Hilbert space $(M_n(\mathbb{C}), \langle\cdot,\cdot\rangle)$, for it is invertible, with inverse $u^{-1}a = a_\rho \rho^{-1/2}$, as well as isometric:

$$\langle u(a), u(b) \rangle = \mathrm{Tr}\,(\rho^{1/2} a^* b \rho^{1/2}) = \mathrm{Tr}\,(\rho a^* b) = \langle a_\rho, b_\rho \rangle.$$

The transformed representation $\hat{\pi}_\rho = u\pi_\rho(a)u^{-1}$ on $M_n(\mathbb{C})$ is simply given by

$$\hat{\pi}_\rho(a)b = ab, \tag{2.70}$$

and the cyclic vector $u\Omega_\rho$ in $M_n(\mathbb{C})$ becomes $\rho^{1/2}$, so that, as in (1.73),

$$\langle \rho^{1/2}, \hat{\pi}_\rho(a)\rho^{1/2} \rangle = \mathrm{Tr}\,(\rho a). \tag{2.71}$$

In this case, the commutant is easily computed to be

$$\hat{\pi}_\rho(M_n(\mathbb{C}))' \cong M_n(\mathbb{C}), \tag{2.72}$$

since any linear map $C : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ that satisfies $C(ab) = aC(b)$ for each $a,b \in M_n(\mathbb{C})$ is of the form $C(a) = ac \equiv R_c(a)$ for some $c \in M_n(\mathbb{C})$, namely $c = C(1)$; to see this, just take $b = 1$. Since this involves *right* multiplication R_c by c, which messes up the order in that $R_c R_d = R_{dc}$, one has a choice in implementing the isomorphism (2.72) either as a *linear anti-homomorphism* (of algebras) $C \mapsto R_c$, or as an *anti-linear homomorphism* $C \mapsto R_{c^*}$ (see also Theorem C.159).

Further insight into the structure of this representation comes from the realization

$$M_n(\mathbb{C}) \cong \mathbb{C}^n \otimes \mathbb{C}^n, \tag{2.73}$$

as Hilbert spaces under the unitary map $v : a \mapsto \sum_{ij} a_{ij} \upsilon_i \otimes \upsilon_j$. This yields

$$v\hat{\pi}_\rho(a)v^* = a \otimes 1_n, \tag{2.74}$$

as an operator on $\mathbb{C}^n \otimes \mathbb{C}^n$, and indeed for any Hilbert spaces H_1, H_2 one has

$$(B(H_1) \bigotimes \mathbb{C} \cdot 1_{H_2})' = \mathbb{C} \cdot 1_{H_1} \bigotimes B(H_2). \tag{2.75}$$

• Finally, in the 'intermediate' case the sum in the spectral decomposition (2.6) has $1 < m < n$ nonzero terms. Using the ensuing (partial) basis $(\upsilon_1, \dots, \upsilon_m)$ of $\mathbb{C}^m$ (viz. $\mathbb{C}^n$), analogously to (2.66) with (2.73) we obtain, up to unitary equivalence,

$$H_\rho \cong \mathbb{C}^n \otimes \mathbb{C}^m; \tag{2.76}$$

$$\pi_\rho(a) \cong a \otimes 1_m; \tag{2.77}$$

$$\Omega_\rho \cong \sum_{i=1}^{n} \sqrt{p_i}\, \upsilon_i \otimes \upsilon_i; \tag{2.78}$$

$$\pi_\rho(M_n(\mathbb{C}))' \cong M_m(\mathbb{C}). \tag{2.79}$$

The relevance of all this to the decomposition of states on $B(H)$ is as follows.

Proposition 2.16. *Let ω be a state on $B(H) \cong M_n(\mathbb{C})$. Then each decomposition*

$$\omega = \sum_i p_i \omega_i, \tag{2.80}$$

where the p_i are probabilities (but the states ω_i are not necessarily pure) is induced by a family (A_i) of nonzero operators in the commutant $\pi_\omega(B(H))'$ that satisfy:

$$0 \leq A_i \leq 1; \tag{2.81}$$

$$\sum_i A_i = 1. \tag{2.82}$$

Namely, given such a family of operators A_i, the decomposition (2.80) *is given by:*

$$p_i = \langle \Omega_\omega, A_i \Omega_\omega \rangle; \tag{2.83}$$

$$\omega_i(a) = \frac{\langle \Omega_\omega, \pi_\omega(a) A_i \Omega_\omega \rangle}{\langle \Omega_\omega, A_i \Omega_\omega \rangle}. \tag{2.84}$$

Proof. The claim that such a family yields (2.80) is trivial, except for the remark that automatically $p_i > 0$, since $\langle \Omega_\omega, A_i \Omega_\omega \rangle = 0$ would imply $\sqrt{A_i}\Omega_\omega = 0$ and hence

$$\sqrt{A_i} a_\omega = \sqrt{A_i}\pi_\omega(a)\Omega_\omega = \pi_\omega(a)\sqrt{A_i}\Omega_\omega = 0$$

for any $a \in B(H)$; by (1.72) this gives $\sqrt{A_i} = 0$ and therefore $A_i = \sqrt{A_i}^2 = 0$.

Conversely, each state ω_i in (2.80) defines a sesquilinear form Q_i on H_ω by $Q_i(a_\omega, b_\omega) = \omega_i(a^*b)$, which is well defined by $\omega_i(a^*a) \leq \omega(a^*a)$ and (A.1), and is positive because ω_i is a state. Proposition A.23 then provides us with a positive operator A_i for which $Q_i(a_\omega, b_\omega) = \langle a_\omega, A_i b_\omega \rangle$, hence $\omega_i(a^*b) = \langle a_\omega, A_i b_\omega \rangle$. Next,

$$\langle a_\omega, A_i \pi_\omega(c) b_\omega \rangle = \langle a_\omega, A_i (cb)_\omega \rangle = \omega_i(a^*cb) = \langle (c^*a)_\omega, A_i b_\omega \rangle = \langle a_\omega, \pi_\omega(c) A_i b_\omega \rangle,$$

so $A_i \in \pi_\omega(B(H))'$. Finally, the bound (2.81) corresponds to $0 \leq p_i \leq 1$ in (2.80), whilst $\omega(1) = 1$, or equivalently $\sum_i p_i = 1$, yields (2.82). □

We now complete the proof of Proposition 2.15. We assume (2.33), where we initially take ρ to be invertible. We omit the hat in (2.70) as well as the suffix ω or ρ on vectors. As noted, we then have $\Omega_\rho = \rho^{1/2}$, and we also know that A_i is given by $A_i b = b a_i$ for some $a_i \in M_n(\mathbb{C})$, viz. $a_i = A_i 1_n$ (where $1_n = 1_H$ is to be distinguished from $\Omega_\rho = \rho^{1/2}$). In this case, (2.81) means $0 \leq \mathrm{Tr}\,(b^*ba_i) \leq 1$ for each b with $\mathrm{Tr}\,(b^*b) = 1$, which is true iff $0 \leq a_i \leq 1$, whereas (2.82) immediately yields $\sum_i a_i = 1$. In terms of such a family (a_i) in $M_n(\mathbb{C})$ itself, the decomposition (2.80) of $\omega = \mathrm{Tr}\,(\rho -)$ into *arbitrary* states ω_i follows from (2.83) - (2.84) as

$$p_i = \mathrm{Tr}\,(\rho a_i); \tag{2.85}$$

$$\omega_i(a) = \mathrm{Tr}\,(\rho_i a); \tag{2.86}$$

$$\rho_i = \frac{\rho^{1/2} a_i \rho^{1/2}}{\mathrm{Tr}\,(\rho a_i)}. \tag{2.87}$$

To obtain *pure and orthogonal* states ω_i, we subsequently ask when the new density matrices ρ_i are mutually orthogonal one-dimensional projections $\rho_i = |\upsilon_i\rangle\langle\upsilon_i|$.

To answer this, we use the spectral theorem (A.37) - (A.38) applied to ρ, which gives $\rho = \sum_j p_j e_j$ and hence $\rho^{1/2} = \sum_j \sqrt{p_j} e_j$, so that

$$\rho^{1/2} a_i \rho^{1/2} = \sum_{j,k} \sqrt{p_j p_k} e_j a_i e_k. \tag{2.88}$$

This can only be proportional to a one-dimensional projection if each a_i is a one-dimensional projection that commutes with all spectral projections e_j of ρ (and hence also commutes with ρ itself), and all further constraints on the a_i may then only be satisfied if $a_i = |\upsilon_i\rangle\langle\upsilon_i|$, for some basis (υ_i) of eigenvector υ_i of ρ.

A similar analysis applies to non-invertible ρ, the only new point being that projections e_i orthogonal to the range of ρ fall into the null space N_ρ, cf. (2.76) - (2.79), and hence do not contribute to (2.52), so that they may be ignored. □

2.5 The Born rule from Bohrification

The Bohrification approach to quantum mechanics studies noncommutative algebras of observables like $B(H)$ through their commutative subalgebras. In this section we show how the Born rule (2.8) emerges from that perspective. Our discussion is based on the interplay between the three kinds of (finite-dimensional) C*-algebras:

- $C(X)$ is a C*-algebra under the pointwise operations (1.18) - (1.20) and the supremum-norm (1.24); we still assume that X is finite.
- $B(H)$ is a C*-algebra under the pointwise operations (2.22) - (2.24) and the operator norm (A.18); our standing assumption remains $\dim(H) < \infty$.
- $C^*(a)$ is the C*-algebra generated by $a \in B(H)$ and 1_H (i.e., the intersection of all unital C*-algebras in $B(H)$ that contain a). If $a^* = a$, then $C^*(a)$ is commutative.

Each of these is *unital*, since $C(X)$ has a unit 1_X (i.e. the function $x \mapsto 1$), $B(H)$ has a unit 1_H (i.e. the operator $\psi \mapsto \psi$), and $C^*(a)$ shares the unit 1_H. The first two classes overlap just in case $\dim(H) = 1$ and X is a singleton (in which case $B(\mathbb{C}) = C(*) = \mathbb{C}$); otherwise, the fundamental difference between the two is that $C(X)$ is *commutative* in that $fg = gf$ for all f, g, whereas $B(H)$ is *non-commutative*. However, the system of C*-algebras $C^*(a)$ within $B(H)$, where $a \in B(H)_{\mathrm{sa}}$ varies, to some extent bridges the gap between the commutative and the non-commutative worlds. This relatively simple situation goes to the heart of exact Bohrification.

Theorem 2.17. *Let $a^* = a \in B(H)$, where H is a finite-dimensional Hilbert space.*

1. *The commutative C*-algebra $C^*(a)$ consists of all polynomials in a.*
2. *Any element of $C^*(a)$ is a linear combination of the spectral projections e_λ of a.*
3. *For functions $f : \sigma(a) \to \mathbb{C}$, the map $f \mapsto f(a)$ defined by*

$$f(a) = \sum_{\lambda \in \sigma(a)} f(\lambda) \cdot e_\lambda. \tag{2.89}$$

gives a (necessarily unital) isomorphism of commutative C-algebras*

$$C(\sigma(a)) \cong \mathbb{C}^{|\sigma(a)|} \cong C^*(a). \tag{2.90}$$

Proof. Noting that any function on the finite subset $\sigma(a)$ of $\mathbb{R}$ is continuous, this is a restatement of Theorem A.15 for finite-dimensional Hilbert spaces. □

We now come to the main point. States on unital C*-algebras A may be defined just as in Definitions 1.14 and 2.5, i.e. as positive linear functionals $\omega : A \to \mathbb{C}$ that satisfy $\omega(1_A) = 1$ (cf. Proposition C.5). Recall Theorem 1.15 and Theorem 2.7.

Theorem 2.18. *Let ω be a state on $B(H)$, represented by a density operator ρ via* (2.33)*, and let $a \in B(H)$ be a self-adjoint operator. Then the restriction of ω to $C^*(a) \subset B(H)$ is a state, which also induces a state $\omega_{|C(\sigma(a))}$ on $C(\sigma(a))$ through* (2.89) - (2.90)*, i.e., $\omega_{|C(\sigma(a))}(f) = \omega(f(a))$. The probability measure on $\sigma(a)$ that corresponds to the state $\omega_{|C(\sigma(a))}$ on $C(\sigma(a))$, then, is given by the Born rule* (2.9).

Proof. First, the restriction of a state on a given unital C*-algebra to a unital C*-subalgebra remains a state. Second, isomorphisms of unital C*-algebras pull back to state spaces in that, if $\varphi : A \to B$ is an isomorphism, and ω is a state on B, then $\varphi^*\omega : A \to \mathbb{C}$ is a state on A, where $\varphi^*(a) = \omega(\varphi(a))$. We now compute

$$\begin{aligned}\omega_{|C(\sigma(a))}(f) &= \omega(f(a)) = \mathrm{Tr}\,(\rho f(a)) \\ &= \sum_{\lambda\in\sigma(a)} \mathrm{Tr}\,(\rho e_\lambda) f(\lambda) = \sum_{\lambda\in\sigma(a)} p_a(\lambda) f(\lambda) \\ &= E_{P_a}(f), \end{aligned} \tag{2.91}$$

where, from left to right, the first equality is just the definition of $\omega_{|C(\sigma(a))}$, whereas the others in turn follow from (2.33), (2.89), (2.8), and (1.9), respectively. □

Note that Theorem 2.18 implies Theorem 2.2. The simplest nontrivial illustration is:

$$H = \mathbb{C}^n; \tag{2.92}$$

$$\omega = \omega_\psi; \tag{2.93}$$

$$\psi = \sum_{i=1}^{n} c_i u_i; \tag{2.94}$$

$$a = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^{n} \lambda_i |u_i\rangle\langle u_i|, \tag{2.95}$$

with respect to the standard basis (u_i) of $\mathbb{C}^n$, with all $\lambda_i \in \mathbb{R}$ different, cf. (2.42). The C*-algebra $C^*(a) \cong \mathbb{C}^n$ then consists of all diagonal matrices

$$b = \mathrm{diag}(b_1, \ldots, b_n). \tag{2.96}$$

Since obviously

$$\sigma(a) = \{\lambda_1, \ldots, \lambda_n\}, \tag{2.97}$$

the isomorphism (2.90) is given by

$$f \mapsto \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n)). \tag{2.98}$$

The computation (2.91) in the proof of Theorem 2.18 then becomes

$$\begin{aligned}\omega_{\psi|C(\sigma(a))}(f) &= \langle \psi, \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n))\psi\rangle = \sum_{i=1}^{n} |c_i|^2 f(\lambda_i) \\ &= \sum_{i=1}^{n} p_a(\lambda_i) f(\lambda_i), \end{aligned} \tag{2.99}$$

from which the Born probabilities p_a may be read off as the familiar expressions

$$p_a(\lambda_i) = |c_i|^2. \tag{2.100}$$

For an analogous treatment of the generalized Born rule (2.21), we first refer to Definition A.16 for the the pertinent definitions, especially of the joint spectrum

$$\sigma(\underline{a}) \subseteq \sigma(a_1) \times \cdots \times \sigma(a_n) \subset \mathbb{R}^n$$

of a family $\underline{a} = (a_1, \ldots, a_n)$ of commuting self-adjoint operators. As in the case of a single operator, we define $C^*(\underline{a})$ as the smallest unital C*-subalgebra of $B(H)$ that contains each a_i. Generalizing Theorem A.15, we have:

Theorem 2.19. *Let $\underline{a} = (a_1, \ldots, a_n)$ be commuting self-adjoint operators on H. Then $C^*(\underline{a})$ is commutative, and there is a unique isomorphism of C*-algebras*

$$C^*(\underline{a}) \cong C(\sigma(\underline{a})), \tag{2.101}$$

under which $1_H \in C^*(\underline{a})$ corresponds to the unit function $1_{\sigma(\underline{a})} : \underline{\lambda} \mapsto 1$ in $C(\sigma(a))$, and $a_i \in C^*(\underline{a})$ corresponds to the projection $\pi_i : \underline{\lambda} \mapsto \lambda_i$ in $C(\sigma(\underline{a}))$.

For further discussion, see Appendix A, Theorem A.17.

Theorem 2.18 may then be generalized in the following way, with similar proof.

Theorem 2.20. *Let ω be a state on $B(H)$, represented by a density operator ρ, and let $\underline{a} = (a_1, \ldots, a_n)$ be commuting self-adjoint operators on H. Then the restriction of ω to $C^*(\underline{a}) \subset B(H)$ is a state, which induces a state $\omega_{|C(\sigma(\underline{a}))}$ on $C(\sigma(\underline{a}))$ through the isomorphism* (2.101). *Then the probability measure on the joint spectrum $\sigma(\underline{a})$ that corresponds to $\omega_{|C(\sigma(\underline{a}))}$ is given by the generalized Born rule* (2.21), *i.e.,*

$$p_{\underline{a}}(\underline{\lambda}) = \mathrm{Tr}\,(\rho e_{\underline{\lambda}}). \tag{2.102}$$

Strictly speaking, in the present context one should restrict (2.21) to $\underline{\lambda} \in \sigma(\underline{a})$, but the claim is correct even if one does not, for the (Born) probability assigned to values $\underline{\lambda} \in \sigma(a_1) \times \cdots \times \sigma(a_n)$ that do not lie in $\sigma(\underline{a})$ is simply zero.

As shown in Proposition A.19 in Appendix A, the multi-operator case is a special case of the single-operator case, in that $C^*(\underline{a}) = C^*(a)$ for a suitable self-adjoint operator a. Since the converse is obvious, Theorems 2.18 and 2.20 are equivalent. Corollary A.20 in Appendix A even shows that *any* unital commutative C*-algebra C in $B(H)$ takes the form $C = C^*(a)$ for some self-adjoint operator $a \in B(H)$. Comparing the restrictions of a state ω on $B(H)$ to C as the latter varies therefore comes down to asking how the various Born probability distributions p_a on $C^*(a)$ are related to each other as a varies. It is clear from (2.8) that if p_a and p_b come from the same density operator ρ (as the notation indicates), then for $\lambda \in \sigma(a)$ and $\mu \in \sigma(b)$,

$$e_{\lambda}^{(a)} = e_{\mu}^{(b)} \Rightarrow p_a(\lambda) = p_b(\mu). \tag{2.103}$$

Indeed, this is the only compatibility condition between p_a and p_b, showing that $p_a(\lambda)$ only depends on a and λ through the associated spectral projection $e_{\lambda}^{(a)}$. Condition (2.103) is a version of a general property of quantum mechanics called ***non-contextuality***, which in this case means that, given its spectral projection $e_{\lambda}^{(a)}$, the 'context' operator a is otherwise irrelevant for the Born probability $p_a(\lambda)$.

2.6 The Kadison–Singer Problem

It should be clear from the example in the previous section that *pure* states ω_ψ on $B(H)$ may well give rise to *mixed* states on $C^*(a)$; referring to (2.94) and (2.100), this is the case whenever $c_i \neq 0$ for more than one value of the index i. If, on the other hand, $c_i \neq 0$ for just a single value $i = j$, then $\psi = u_j$ (up to a phase), or, equivalently, $\omega_\psi(a) = \langle u_j, au_j \rangle$. In that case, the given state ω_ψ is pure both on $B(H)$ and on $C^*(a)$, and the associated probability measure $\omega_{\psi|C(\sigma(a))}$ on the spectrum $\sigma(a)$ is supported by a single point, namely $\lambda_j \in \sigma(a)$.

This example suggests a general problem (first posed in the non-trivial case where H is infinite-dimensional by Kadison and Singer in 1959) that is of great relevance for the Bohrification program. Namely, let A be a maximal commutative unital C*-algebra in $B(H)$ and let ω_A be a pure state on A. We may then ask:

1. Does ω_A have an extension to a state ω on $B(H)$ at all (i.e., $\omega_{|A} = \omega_A$)?
2. If so, is ω uniquely determined by its restriction ω_A?
3. Either way, if ω exists, can it be chosen so as to be pure (assuming ω_A is)?

If $\dim(H) < \infty$, all these questions are easy to answer at one stroke:

Theorem 2.21. *Let* $\dim(H) < \infty$ *and let* ω_A *be a pure state on a maximal commutative unital C*-algebra A in* $B(H)$*. Then* ω_A *has a unique extension to a state* ω *on* $B(H)$*, which is necessarily pure.*

Proof. As explained after the proof of Corollary A.20 in Appendix A, we may simply assume that $H = \mathbb{C}^n$ and that A consists of all diagonal matrices; call this collection $D_n(\mathbb{C})$ (for every other case is unitarily equivalent to this one). Clearly,

$$D_n(\mathbb{C}) \cong \mathbb{C}^n, \tag{2.104}$$

from which we see that if ω_A is pure, then it must be given on $b \in D_n(\mathbb{C})$ by

$$\omega_A(b) = b_j, \tag{2.105}$$

for some j, cf. (2.96). If ω exists, it is given by (2.33). Using (2.6), condition (2.105) then enforces the following constraint on the p_i and υ_i (where (u_i) is the standard basis of $\mathbb{C}^n$ and (υ_i) is an orthonormal set diagonalizing the density operator ρ):

$$\sum_i p_i |\langle u_j, \upsilon_i \rangle|^2 = 1. \tag{2.106}$$

Since $\sum_i p_i = 1$ and $|\langle u_j, \upsilon_i \rangle| \leq 1$, eq. (2.106) can only hold, for given j, if

$$|\langle u_j, \upsilon_i \rangle| = 1 \tag{2.107}$$

for all i with $p_i > 0$. Since u_j is a unit vector whilst the (υ_i) are an orthonormal set, (2.107) can only be true if there is a single i for which $p_i > 0$, namely $i = j$ (and hence $p_j = 1$), in which case υ_j must equal u_j up to a phase. Hence $\rho = |u_j\rangle\langle u_j|$, which shows that ρ exists, is unique, and is pure. □

At least in operational interpretations of quantum mechanics, this theorem implies that a *pure* quantum state (i.e., on $B(H)$) is completely determined by the outcome of a measurement of some maximal observable a, whose outcome, after all, gives one of the eigenvalues λ_j in (2.95) and hence fixes the post-measurement state to be the one given by (2.105). This is, indeed, a typical way of preparing a state.

As one might expect, this is no longer true if $A = C^*(a)$ fails to be maximal (in which case a measurement of a would not provide enough information about the quantum state). Namely, suppose $a = \sum_{\lambda\in\sigma(a)} \lambda \cdot e_\lambda$, as in (A.37); the maximal case occurs iff $\mathrm{Tr}\,(e_\lambda) = \dim(H_\lambda) = 1$ for all $\lambda \in \sigma(a)$ (equivalently, all eigenvalues λ_i in (A.37) are different). If not, suppose $\dim(H_\lambda) > 1$ for some λ. Then any unit vector $\psi \in H_\lambda$ gives rise to a pure state ω_ψ on $B(H)$, which remains pure on A (it is given by $\omega_{\psi|A}(a) = \lambda$ and hence induces the Dirac probability measure δ_λ on $\sigma(a)$).

Dropping the purity condition on ω_A loses uniqueness of the extension ω, too, even if A is maximal: take $b = \mathrm{diag}(b_1,\ldots,b_n) \in A = D_n(\mathbb{C})$, and assume that

$$\omega_A(b) = \sum_i p_i b_i \tag{2.108}$$

has more than one term (with $p_i > 0$ and $\sum_i p_i = 1$ as always), cf. (2.105). Then:

- *any* pure state ω_ψ as in (2.94), such that $|c_i|^2 = p_i$ for all i, extends ω_A;
- the "decohered" *mixed* state $\omega = \sum_i p_i |\upsilon_i\rangle\langle\upsilon_i|$ extends ω_A, too.

Further insight in the state extension problem comes from the following result.

Proposition 2.22. *Let A be any unital C*-algebra in $B(H)$ (i.e., A is not necessarily commutative) and let ω_A be a pure state on A. Then the set*

$$S_A = \{\omega \in S(B(H)) \mid \omega_{|A} = \omega_A\} \tag{2.109}$$

of all states on $B(H)$ whose restriction $\omega_{|A}$ to A is the given state ω_A, is a compact convex subspace of the total state space $S(B(H))$ of $B(H)$, whose extreme boundary $\partial_e S_A$ consist of pure states on $B(H)$, i.e., $\partial_e S_A \subset P(B(H))$. Consequently, ω_A has a unique extension to a state on $B(H)$ iff it has a unique pure extension.

Proof. Convexity and (w^*) compactness are obvious. Let $\omega \in \partial_e S_A$ and suppose $\omega = t\omega_1 + (1-t)\omega_2$ for some $t \in (0,1)$ and $\omega_1, \omega_2 \in S(B(H))$. By assumption, $\omega_A = \omega_{|A} = t\omega_{1|A} + (1-t)\omega_{2|A}$ is pure on A, so $\omega_{1|A} = \omega_{2|A} = \omega_A$, hence $\omega_1, \omega_2 \in S_A$. Since $\omega \in \partial_e S_A$, this implies $\omega_1 = \omega_2 = \omega$. Hence ω is pure on $B(H)$.

Finally, S_A is a singleton iff its boundary $\partial_e S_A$ is (since any state in S_A has a convex decomposition in terms of states in its boundary), yielding the last claim. □

This proposition remains true for infinite-dimensional H (and even for arbitrary C*-algebras), but Theorem 2.21 becomes much more complicated. As we shall see, maximal commutative unital C*-subalgebra of $B(H)$ are no longer unique up to unitary equivalence, and the validity of the claim depends on which type of maximal subalgebra is considered. Also, the proof of what then is called the ***Kadison–Singer Conjecture*** becomes extremely difficult (with questionable relevance to physics).

2.7 Gleason's Theorem

Gleason's Theorem answers the following question in the positive: given probability distributions p_a on $\sigma(a)$, for each self-adjoint operator $a \in B(H)$, satisfying (2.103), is there a single state ω on $B(H)$ inducing these probabilities through the Born rule? This question is closely related to various others that involve equivalent structures, cf. Definition 1.1. We denote the unit sphere in H by $H_1 = \{\psi \in H, \|\psi\| = 1\}$, and write $\mathscr{P}(H) = \{e \in B(H) \mid e^2 = e^* = e\}$ for the set of all projections on H.

Definition 2.23. *Let H be a finite-dimensional Hilbert space, with unit sphere H_1.*

1. *A* **probability distribution** *on $\mathscr{P}(H)$ is a map $p : H_1 \to [0,1]$ that satisfies*

$$\sum_{i=1}^{\dim H} p(\upsilon_i) = 1, \text{ for any basis } (\upsilon_i) \text{ of } H. \tag{2.110}$$

2. *A* **probability measure** *on $\mathscr{P}(H)$ is a map $P : \mathscr{P}(H) \to [0,1]$ that satisfies:*

$$P(e+f) = P(e) + P(f) \text{ whenever } ef = 0 \Leftrightarrow eH \perp fH; \tag{2.111}$$

$$P(1_H) = 1. \tag{2.112}$$

Note that p is really defined on $\mathscr{P}_1(H)$, for we have $p(z\upsilon) = p(\upsilon)$ for all $z \in \mathbb{T}$ and $\upsilon \in H_1$; to see this, extend $z\upsilon$ and υ to a basis of H in the same way and use (2.110).

As in Definition 1.1, these notions of probability are equivalent, cf. (A.28):

- Given a probability measure P, one obtains a probability distribution p by

$$p(\upsilon) = P(e_\upsilon). \tag{2.113}$$

- Given a probability distribution p, Lemma 2.24 below guarantees that

$$P(e) = \sum_{i=1}^{\dim(eH)} p(\upsilon_i), \tag{2.114}$$

where (υ_i) is any basis of eH, defines a probability measure P.

Lemma 2.24. *If p is a probability distribution on $\mathscr{P}(H)$ and $L \subset H$ is a linear subspace, with basis (υ_i), then $\sum_{i=1}^{\dim(L)} p(\upsilon_i)$ is independent of this basis choice.*

Proof. Extend (υ_i) to a basis of H by adding a basis (υ'_j) of $L^\perp$. Take another basis (υ''_i) of L and complete it to a basis of H by using the same basis (υ'_j) of $L^\perp$. Then

$$\sum_i p(\upsilon_i) + \sum_j p(\upsilon'_j) = \sum_i p(\upsilon''_i) + \sum_j p(\upsilon'_j) = 1, \tag{2.115}$$

where we once again used (2.110). Hence $\sum_i p(\upsilon_i) = \sum_i p(\upsilon''_i)$. □

Clearly, a state ω on $B(H)$ induces a probability measure P on $\mathscr{P}(H)$ by

$$P(e) = \omega(e) = \mathrm{Tr}\,(\rho e), \tag{2.116}$$

where ρ is the density operator associated to ω, as in (2.33). Therefore, it is a natural question if any probability measure on $\mathscr{P}(H)$ is induced by some state on $B(H)$ by (2.116). This question is equivalent to the one above:

Proposition 2.25. • *A probability measure P on $\mathscr{P}(H)$ induces non-contextual probability distributions p_a on $\sigma(a)$ for each self-adjoint $a \in B(H)$ by*

$$p_a(\lambda) = P(e_\lambda^{(a)}); \tag{2.117}$$

• *Conversely, a family (p_a) of non-contextual probability distributions (i.e. satisfying* (2.103)*) gives rise to a probability measure P on $\mathscr{P}(H)$ by*

$$P(e) = p_e(1). \tag{2.118}$$

Proof. As defined by (2.117), p_a is a probability distribution on $\sigma(a)$: by (A.38),

$$\sum_{\lambda\in\sigma(a)} p_a(\lambda) = \sum_{\lambda\in\sigma(a)} P\left(e_\lambda^{(a)}\right) = P\left(\sum_{\lambda\in\sigma(a)} e_\lambda^{(a)}\right) = P(1_H) = 1. \tag{2.119}$$

Conversely, suppose $ef = 0$. Introduce $g = 1 - e - f$, and consider the self-adjoint operator $a = \lambda_1 e + \lambda_2 f + \lambda_3 g$, for three different real numbers $\lambda_1, \lambda_2, \lambda_3$. By (2.103),

$$P(e) = p_e(1) = p_a(\lambda_1),\, P(f) = p_f(1) = p_a(\lambda_2),\, P(g) = p_g(1) = p_a(\lambda_3).$$

Furthermore, since $\sigma(a) = \{\lambda_1, \lambda_2, \lambda_3\}$, we have $p_a(\lambda_1) + p_a(\lambda_2) + p_a(\lambda_3) = 1$ and hence $P(e) + P(f) + P(g) = 1$. Also, $P(e + f) + P(g) = P(e + f + g) = P(1_H) = 1$. The last two equations give $P(e + f) = P(e) + P(f)$. □

Suppose $(e_i)_{i=1}^N$ is a family of projections on H such that $\sum_i e_i = 1_H$ and $e_i e_j = \delta_{ij} e_i$. Such a family generates a commutative unital C*-algebra $C = C^*(e_1, \ldots, e_N)$ in $B(H)$, which coincides with $C^*(a)$ for $a = \sum_i \lambda_i e_i$, where all $\lambda_i \in \mathbb{R}$ are different, so that $\sigma(a) = \{\lambda_1, \ldots, \lambda_N\}$. All commutative unital C*-algebras in $B(H)$ arise in this way, and C is maximally abelian iff $N = \dim(H)$, i.e., iff each e_i is one-dimensional. The point is that a probability measure P on $\mathscr{P}(H)$ induces a state ω_C on each $C = C^*(e_1, \ldots, e_N)$ (or, for $C = C^*(a)$, a probability measure P_a on $\sigma(a)$):

1. if $a \in C$ is self-adjoint, then we have unique spectral resolutions (A.37), and put

$$\omega_C(a) = \sum_{\lambda\in\sigma(a)} \lambda P(e_\lambda). \tag{2.120}$$

2. if $c = a + ib \in C$ with a and b self-adjoint, we define $\omega_C(c) = \omega_C(a) + i\omega_C(b)$.

By Lemma 2.24, the map ω_C thus defined coincides with the linear extension of the map $e_i \mapsto P(e_i)$ to C, which also shows that ω_C in linear. Clearly, ω_C is a state on C.

Again by Lemma 2.24, the ensuing family of states ω_C on all commutative unital C*-algebras $C \subset B(H)$ is *non-contextual* (or, one might say *compatible*) in the sense that if $b \in C \cap C'$, then $\omega_C(b) = \omega_{C'}(b)$. In particular, if $C' \subset C$, then $\omega_{C|C'} = \omega_C$ (where $\omega_{C|C'}$ is the restriction of ω_C to C'). It is convenient to extend this non-contextual family (ω_C) of states to a well-defined map $\omega : B(H) \to \mathbb{C}$ by putting

$$\omega(a+ib) = \omega_{C^*(a)}(a) + i\omega_{C^*(b)}(b),\ a,b \in B(H), a^* = a, b^* = b. \tag{2.121}$$

Definition 2.26. *A* **quasi-state** *on $B(H)$ is a map $\omega : B(H) \to \mathbb{C}$ that is positive ($\omega(a^*a) \geq 0$) and normalized ($\omega(1_H) = 1$), cf. Definition 2.4, and otherwise:*

1. *satisfies $\omega(a) = \omega(a') + i\omega(a'')$, where $a' = \frac{1}{2}(a+a^*)$ and $a'' = -\frac{1}{2}i(a-a^*)$.*
2. *is linear on each commutative unital C*-algebra in $B(H)$.*

Note that a' and a'' are self-adjoint, so that ω is fixed by its values on $B(H)_{\mathrm{sa}}$. Hence we have $\omega(za) = z\omega(a)$, $z \in \mathbb{C}$, and $\omega(a+b) = \omega(a) + \omega(b)$ whenever $ab = ba$.

Proposition 2.27. *The map $\omega : B(H) \to \mathbb{C}$ defined by* (2.120) *and* (2.121) *is a quasi-state on $B(H)$. Any quasi-state on $B(H)$ arises in this way, giving a bijective correspondence between quasi-states on $B(H)$ and probability measures on $\mathscr{P}(H)$.*

Proof. The first claim holds by construction. Conversely, a quasi-state ω yields a probability measure P via $P(e) = \omega(e)$, cf. (2.116). □

Theorem 1.15 shows that each state on $C(X)$ is induced by a probability measure (and, trivially, also the other way round). Although Theorem 2.7 is already a quantum version of Theorem 1.15, an even better parallel would involve the probability measures of Definition 2.23. This is indeed what ***Gleason's Theorem*** achieves, *en passant* answering all versions of our lead question:

Theorem 2.28. *Let H be a finite-dimensional Hilbert space of dimension > 2. Then each probability measure P on $\mathscr{P}(H)$ is induced by a unique state ω on $B(H)$ via*

$$P(e) = \omega(e). \tag{2.122}$$

Equivalently, each probability distribution p on $\mathscr{P}(H)$ is given by

$$p(\upsilon) = \langle \upsilon, \rho\upsilon \rangle, \tag{2.123}$$

where ρ is a unique density operator on H. Hence every quasi-state is a state.

This completes the following list (of which 1–5 do not require Gleason's Theorem).

Corollary 2.29. *Let H be a finite-dimensional Hilbert space. The following notions are equivalent (i.e., there are natural bijective correspondence between):*

1. *Non-contextual families of states on commutative unital C*-algebras $C \subset B(H)$;*
2. *Non-contextual families of probability measures on spectra $\sigma(a)$, cf.* (2.103)*;*
3. *Probability distributions on $\mathscr{P}(H)$;*
4. *Probability measures on $\mathscr{P}(H)$;*
5. *Quasi-states on $B(H)$;*
6. *States on $B(H)$.*

2.8 Proof of Gleason's Theorem

The difficulty of Theorem 2.28 should already be clear from the fact that it is false if $\dim(H) = 2$: as we have seen in (2.37), a state on $M_2(\mathbb{C}) = B(\mathbb{C}^2)$ is given by three real parameters, whereas a probability measure P on $\mathscr{P}(\mathbb{C}^2)$ can assign arbitrary values $P(e)$ to one-dimensional projections e, as long as $P(1-e) = 1 - P(e)$. Equivalently, this time from the perspective of probability distributions p, each unit vector in $\mathbb{C}^2$ belongs to a unique basis (up to a phase), so that p can assign an arbitrary value to one of the two vectors in each basis and is unconstrained otherwise.

In higher dimensions, however, one-dimensional projections always belong to infinitely many orthogonal sets, whilst unit vectors belong to infinitely many bases. This constrains the possible values P or p may take, and these constraints turn out to be strong enough to enforce (2.116).

The proof of Theorem 2.28 consists of two nontrivial parts, the second of which is notoriously difficult. By exception in quantum-mechanical reasoning, both involve $\mathbb{R}^3$ as a *real* Hilbert space, whose elements $\mathbf{x} = (x,y,z)$ have standard inner product

$$\langle \mathbf{x}, \mathbf{x}' \rangle = xx' + yy' + zz', \tag{2.124}$$

with the ensuing (Pythagorean) norm and (Euclidean) notion of orthogonality.

Proposition 2.30. *If Theorem 2.28 holds for the real Hilbert space $\mathbb{R}^3$, then it holds for any complex finite-dimensional Hilbert space of dimension > 2.*

Proposition 2.31. *Theorem 2.28 holds for the real Hilbert space $\mathbb{R}^3$.*

Proposition 2.30 is a conjunction of two lemmas.

Lemma 2.32. *If* (2.123) *holds for $\mathbb{R}^3$, where ρ is some symmetric operator, then* (2.123) *holds for $\mathbb{C}^3$, where ρ is a self-adjoint operator.*

Neither positivity nor normalization of ρ play a role in the argument; once we have (2.123) in this more general sense, the conclusion that ρ be a density operator trivially follows from the definition of p. This also applies to the second sublemma.

Lemma 2.33. *If* (2.123) *holds for $\mathbb{C}^3$, then it holds for for any complex finite-dimensional Hilbert space of dimension > 2.*

It will be convenient to extend $p : H_1 \to [0,1]$ to a function $Q : H \to \mathbb{R}$ by

$$Q(0) = 0; \tag{2.125}$$

$$Q(\psi) = \|\psi\|^2 p\left(\frac{\psi}{\|\psi\|}\right) \quad (\psi \neq 0), \tag{2.126}$$

so that (2.123) is evidently equivalent to the analogous expression

$$Q(\psi) = \langle \psi, \rho\psi \rangle \quad (\psi \in H). \tag{2.127}$$

Given (2.127), the minimax principle for real symmetric matrices implies that Q is maximized on H_1 by $\psi \in H_1$ iff $\rho\psi = \lambda\psi$, where λ is the largest eigenvalue of ρ.

Proof of Lemma 2.32. Suppose $p : \mathbb{C}^3_1 \to [0,1]$ is a probability distribution (in the sense of Definition 2.23). The first step shows that p assumes a maximum on the unit sphere $\mathbb{C}^3_1$ (note that $\mathbb{C}^3_1$ is compact, but we do not know yet if p is continuous!). Since $0 \leq p(\upsilon) \leq 1$ for $\upsilon \in \mathbb{C}^3_1$, $M = \sup\{p(\upsilon), \upsilon \in \mathbb{C}^3_1\}$ exists, and there is a sequence (υ_n) in $\mathbb{C}^3_1$ for which $p(\upsilon_n) \to M$. Since $\mathbb{C}^3_1$ is compact, this sequence has a convergent subsequence, with limit $\upsilon_\infty \in \mathbb{C}^3_1$. Furthermore, we may assume that $\langle \upsilon_n, \upsilon_\infty \rangle \in \mathbb{R}$, for if not, we change to $\upsilon'_n = z_n \upsilon_n$ with $z_n = \langle \upsilon_\infty, \upsilon_n \rangle / |\langle \upsilon_n, \upsilon_\infty \rangle|$.

For each fixed n (with υ_n in the convergent subsequence in question), the real linear span of υ_∞ and υ_n is isomorphic to $\mathbb{R}^2$ as a Hilbert space (with standard inner product), embedded in any $\mathbb{R}^3 \subset \mathbb{C}^3$ one likes (where, once again, $\mathbb{R}^3$ is seen as a real Hilbert subspace in the sense that all inner products of vectors in $\mathbb{R}^3$ are real). By assumption, (2.123) holds on $\mathbb{R}^3$ and hence also on $\mathbb{R}^2 \subset \mathbb{R}^3$, so that, in particular,

$$\begin{aligned} |p(\upsilon_\infty) - p(\upsilon_n)| &= |\langle \upsilon_\infty, \rho\upsilon_\infty \rangle - \langle \upsilon_n, \rho\upsilon_n \rangle| = |\langle (\upsilon_\infty - \upsilon_n), \rho(\upsilon_\infty + \upsilon_n) \rangle| \\ &\leq \|\rho\| \|\upsilon_\infty + \upsilon_n\| \|\upsilon_\infty - \upsilon_n\| \leq 2\|\rho\| \|\upsilon_\infty - \upsilon_n\|, \end{aligned}$$

since $\|\upsilon_\infty + \upsilon_n\| \leq \|\upsilon_\infty\| + \|\upsilon_n\|$ and $\|\upsilon_\infty\| = \|\upsilon_n\| = 1$. Consequently,

$$|p(\upsilon_\infty) - M| \leq |p(\upsilon_\infty) - p(\upsilon_n)| + |p(\upsilon_n - M| \leq 2\|\rho\| \|\upsilon_\infty - \upsilon_n\| + |p(\upsilon_n) - M|,$$

so letting $n \to \infty$ makes both terms on the right-hand side vanish. Hence $p(\upsilon_\infty) = M$.

For reasons to become clear soon, we relabel $\upsilon_\infty \equiv \upsilon_1$. Take any $\upsilon_0 \in \mathbb{C}^3_1$ with $\langle \upsilon_0, \upsilon_1 \rangle = 0$ and consider the *real* Hilbert space $\mathbb{R}^2 \subset \mathbb{C}^3$ spanned by υ_1 and υ_0. By assumption, (2.127) holds, and by the minimax principle, $\rho\upsilon_1 = \lambda_1\upsilon_1 = p(\upsilon_1)\upsilon_1$, with $p(\upsilon_1) = M$. Hence for any $\upsilon = t_0\upsilon_0 + t_1\upsilon_1$, with $t_0, t_1 \in \mathbb{R}$, we have

$$Q(\upsilon) = \langle t_0\upsilon_0 + t_1\upsilon_1, \rho(t_0\upsilon_0 + t_1\upsilon_1) \rangle = |t_0|^2 p(\upsilon_0) + |t_1|^2 p(\upsilon_1). \tag{2.128}$$

We claim that this also holds for *complex* coefficients $t_0, t_1 \in \mathbb{C}$. Indeed, by (2.126),

$$Q(t_0\upsilon_0 + t_1\upsilon_1) = |t_1|^2 Q\left(\frac{|t_0|}{|t_1|}\frac{|t_1|}{|t_0|}\frac{t_0}{t_1}\upsilon_0 + \upsilon_1\right) = |t_0|^2 p(\upsilon_0) + |t_1|^2 p(\upsilon_1), \tag{2.129}$$

where we used (2.128) with $\upsilon'_0 = (t_0/t_1)/|(t_0/t_1)|\upsilon_0$ instead of υ_0; this is still a vector orthogonal to υ_1, and we also used $Q(\upsilon'_0) = p(\upsilon'_0) = p(\upsilon_0)$.

We now repeat this analysis on the part $(\mathbb{C}^3_1)_{\perp\upsilon_1}$ of $\mathbb{C}^3_1$ that consists of all unit vectors orthogonal to υ_1, which remains compact. Thus p assumes a maximum at some unit vector $\upsilon_2 \in (\mathbb{C}^3_1)_{\perp\upsilon_1}$, and we may complete the pair (υ_1, υ_2) to a basis $(\upsilon_1, \upsilon_2, \upsilon_3)$ of $\mathbb{C}^3$. With $\upsilon_0 = t_2\upsilon_2 + t_3\upsilon_3$, the above argument (on $(\mathbb{C}^3_1)_{\perp\upsilon_1}$) gives

$$p(\upsilon_0) = Q(\upsilon_0) = |t_2|^2 p(\upsilon_2) + |t_3|^2 p(\upsilon_3). \tag{2.130}$$

Combined with (2.129) at $t_0 = 1$, this gives, for any coefficients $t_1, t_2, t_3 \in \mathbb{C}$,

$$Q(t_1\upsilon_1+t_2\upsilon_2+t_3\upsilon_3)=|t_1|^2p(\upsilon_1)+|t_2|^2p(\upsilon_2)+|t_3|^2p(\upsilon_3). \tag{2.131}$$

Hence (2.127) holds on all of $\mathbb{C}^3$, with

$$\rho=p(\upsilon_1)|\upsilon_1\rangle\langle\upsilon_1|+p(\upsilon_2)|\upsilon_2\rangle\langle\upsilon_2|+p(\upsilon_3)|\upsilon_3\rangle\langle\upsilon_3|. \qquad \square$$

Proof of Lemma 2.33. Let H be a complex finite-dimensional Hilbert space of dimension ≥ 3, equipped with a probability distribution p, and define $Q: H\to\mathbb{R}$ by (2.125) - (2.126). We need to prove (2.127) for some self-adjoint operator ρ. By Propositions A.4 and A.23, this is equivalent to Q being a quadratic form. Since (A.8) evidently holds, we just need to prove (A.9). Take any three-dimensional Hilbert space $L_3\subset H$ containing v and w. By assumption, there exists a self-adjoint operator ρ_{L_3} on L_3 for which (2.127) is valid for all $\psi\in L_3$. Taking $\psi=v$, $\psi=w$, $\psi=v+w$, and $\psi=v-w$ then validates (A.9). This completes the first proof.

This lemma may also be proved without invoking Proposition A.4, as follows.

If v and w are linearly independent, they are contained in a unique two-dimensional subspace $L_2\subset H$, which in turn is contained in a (non-unique) three-dimensional subspace $L_3\subset H$. Take ρ_{L_3} as above and define a bilinear form B on L_2 by $B(v,w)=\langle v,\rho_{L_3}w\rangle$. Defining the associated quadratic form Q by (A.7), we see that (2.125) - (2.126) hold, from which we also conclude that B is independent of the choice of $L_3\supset L_2$. If v and w are linearly dependent, a similar argument shows that B is independent of the choice of the subspace L_2 containing v and w. Hence $B: H\times H\to\mathbb{C}$ is well defined, and to conclude that it is a self-adjoint form we need to check that $B(v,\lambda w+x)=\lambda B(v,w)+B(v,x)$ for all $v,w,x\in V$, $\lambda\in\mathbb{C}$, cf. Definition A.1. If v,w, and x are linearly independent, this can be done by passing to the unique three-dimensional subspace $L_3'\subset H$ containing these vectors. If they are not, we are already done by the previous step. Finally, given that B is a bilinear form, a self-adjoint operator ρ may be reconstructed from Proposition A.23, upon which (2.127) holds by construction. $\square$

Proposition 2.31 again follows from two lemmas by *modus ponens*.

Lemma 2.34. *Any probability distribution on $\mathbb{R}^3$ (vf. Definition 2.23) is continuous.*

Lemma 2.35. *Any continuous probability distribution in $\mathbb{R}^3$ satisfies* (2.127)*, for some self-adjoint operator ρ.*

The operator ρ obtained by Lemma 2.35 is necessarily positive and automatically has unit trace. Another way to phrase this is to take the complex linear span of all probability distribution on the unit sphere $\mathbb{R}^3_1=S^2$ in $\mathbb{R}^3$; this yields a vector space $\mathscr{F}(S^2)$, whose elements are called ***frame functions***. These are *bounded* functions

$$f: S^2\to\mathbb{C},$$

with the property that for any basis $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$ of $\mathbb{R}^3$ one has

$$f(\mathbf{u}_1)+f(\mathbf{u}_2)+f(\mathbf{u}_3)=w(f), \tag{2.132}$$

where $w(f) \in \mathbb{C}$ does not depend on the basis and is called the ***weight*** of the frame function f. For a probability distribution p we obviously have $w(p) = 1$. The natural norm on $\mathscr{F}(S^2)$ is the supremum-norm inherited from $C(S^2)$, and like the latter, $\mathscr{F}(S^2)$ is closed in this norm (and hence is a Banach space in its own right, a fact that will play an important technical role in Lemma 2.40 below).

As for probability distributions, (2.132) implies a lemma that will often be used:

Lemma 2.36. *If $(\mathbf{u}_1, \mathbf{u}_2)$ is a basis of some two-dimensional linear subspace of $\mathbb{R}^3$, then $f(\mathbf{u}_1) + f(\mathbf{u}_2)$ is independent of the choice of this pair. Hence if C is some great circle in S^2 and $\mathbf{u}_1 \perp \mathbf{u}_2$ for $\mathbf{u}_1, \mathbf{u}_2 \in C$, then $f(\mathbf{u}_1) + f(\mathbf{u}_2)$ only depends on C.*

Furthermore, by similar arguments any frame function is even, i.e., $f(-\mathbf{u}) = f(\mathbf{u})$.

The proof of Lemma 2.34 will actually show that every frame function on S^2 is continuous, whilst the proof of Lemma 2.35 will establish the property that any continuous frame function on S^2 satisfies (2.127), for some self-adjoint operator ρ.

Proof of Lemma 2.34. Let $f : S^2 \to \mathbb{R}$ be a frame function (the complex-valued case follows by decomposing f into a real and an imaginary part). Since constants are frame functions, adding a constant to f if necessary we may assume

$$\inf\{f(\mathbf{x}), \mathbf{x} \in S^2\} = 0. \tag{2.133}$$

Hence for given $\varepsilon > 0$ there exists $\mathbf{p} \in S^2$ with

$$f(\mathbf{p}) < \varepsilon/2. \tag{2.134}$$

Performing a rotation if necessary, we may assume that $\mathbf{p} = (0,0,1)$ is the north pole. It is useful to introduce another frame function $g : S^2 \to \mathbb{R}^+$ by

$$g(\mathbf{x}) = f(\mathbf{x}) + f(R_z(\pi/2)\mathbf{x}), \tag{2.135}$$

where $R_z(\pi/2)$ is the (counter-clockwise) rotation around the z-axis by an angle $\pi/2$. It is easy to see that g is constant on the equator E: for $\mathbf{x} \in E$, consider the basis $(\mathbf{x}, R_z(\pi/2)\mathbf{x}, \mathbf{p})$ of $\mathbb{R}^3$, so that $g(\mathbf{x}) = w(f) - f(\mathbf{p})$ is independent of $\mathbf{x}$.

Furthermore, for any $U \subset S^2$ consider the ***oscillation*** of f at U, defined by

$$\mathrm{Osc}_U(f) = \sup_U(f) - \inf_U(f) \equiv \sup\{f(\mathbf{u}), \mathbf{u} \in U\} - \inf\{f(\mathbf{u}), \mathbf{u} \in U\}. \tag{2.136}$$

If, for given $\mathbf{x} \in S^2$, for any $\varepsilon > 0$ there is a neighbourhood $U \subset S^2$ of $\mathbf{x}$ on which $\mathrm{Osc}_U(f) < \varepsilon$, then $|f(\mathbf{x}) - f(\mathbf{u})| < \varepsilon$ for all $\mathbf{u} \in U$, so that f is continuous at $\mathbf{x}$.

The lengthier steps in the proof of Lemma 2.34 are now as follows:

Lemma 2.37. *Given that $g(\mathbf{p}) < \varepsilon$, there is an open set $U \subset S^2$ on which*

$$\mathrm{Osc}_U(g) < 3\varepsilon.$$

Lemma 2.38. *For any non-negative frame function h, if $\mathrm{Osc}_U(h) \leq \varepsilon'$ for some open U, then each point $\mathbf{x} \in S^2$ has a neighborhood V where*

$$\mathrm{Osc}_V(h) \leq 4\varepsilon'.$$

Assuming these lemmas (to be proved below), continuity of f easily follows:

1. Lemmas 2.37 and 2.38 applied to $h = g$ and $\mathbf{x} = \mathbf{p}$ yield $\mathrm{Osc}_V(g) < 12\varepsilon$ for some neighbourhood V of $\mathbf{p}$. Now $g(\mathbf{p}) < \varepsilon$, hence $\inf\{g(\mathbf{v}), \mathbf{v} \in V\} < \varepsilon$, hence
$$\sup_V(f) \leq \sup_V(g) \leq \mathrm{Osc}_V(g) + \inf_V(g) < 13\varepsilon.$$
2. Since $f \geq 0$ and hence $0 \leq \inf_V(f) \leq \sup_V(f)$, this yields $\mathrm{Osc}_V(f) < 13\varepsilon$.
3. Applying Lemmas 2.38 to $h = f$ and $U = V$ gives that each point $\mathbf{x} \in S^2$ has a neighborhood W where $\mathrm{Osc}_W(f) < 52\varepsilon$.
4. Hence $|f(\mathbf{x}) - f(\mathbf{w})| < 52\varepsilon$ for all $\mathbf{w} \in W$. Since $\varepsilon > 0$ was arbitrary, it follows that f is continuous at $\mathbf{x}$, and since $\mathbf{x}$ was arbitrary, f is continuous on all of S^2.

For $\mathbf{p} \neq \mathbf{u} \in N$, i.e., the open northern hemisphere, let $C_{\mathbf{u}}$ be the unique great circle through $\mathbf{u}$ with one (and hence both) of the following equivalent properties:

- the point of greatest latitude on $C_{\mathbf{u}}$ is $\mathbf{u}$;
- $C_{\mathbf{u}}$ cuts the equator E at two points that are both orthogonal to $\mathbf{u}$.

We write $D_{\mathbf{u}} = C_{\mathbf{u}} \cap N$, and for each $\mathbf{z} \in N$, we introduce the set

$$DD_{\mathbf{z}} = \{\mathbf{x} \in N \mid \exists \mathbf{y} \in D_{\mathbf{x}}, \mathbf{z} \in D_{\mathbf{y}}\}. \tag{2.137}$$

Geometrically, $DD_{\mathbf{z}}$ consists of the points $\mathbf{x}$ on the northern hemisphere from which $\mathbf{z}$ can be reached by "double descent", where we say that $\mathbf{y} \in N$ may be reached from some point $\mathbf{x}$ at higher latitude by (single) descent if $\mathbf{y} \in C_{\mathbf{x}}$. The proof of our lemmas relies on the following two facts from spherical geometry (stated without proof, as they have nothing to do with frame functions, though the second is easy).

Lemma 2.39. *1. The set $DD_{\mathbf{z}}$ in* (2.137) *has open interior.*
2. For any $\mathbf{x} \in S^2$ there exists $\mathbf{y} \in E$ such that $\mathbf{x}$ lies on the equator $E_{\mathbf{y}}$ relative to $\mathbf{y}$ regarded as the north pole (so in this terminology, $E = E_{\mathbf{p}}$).

Proof of Lemma 2.37. By definition of the infimum, for each $\varepsilon > 0$ there exists $\mathbf{z} \in N$ such that
$$\inf_N g \leq g(\mathbf{z}) \leq \inf_N g + \varepsilon. \tag{2.138}$$

The open U in question will be the interior of $DD_{\mathbf{z}}$. The crucial inequality is

$$g(\mathbf{x}) < g(\mathbf{z}) + 2\varepsilon \;\; (\mathbf{x} \in DD_{\mathbf{z}}), \tag{2.139}$$

which together with (2.138) yields $\inf_N g \leq g(\mathbf{x}) \leq \inf_N g + 3\varepsilon$ for each $\mathbf{x} \in DD_{\mathbf{z}}$, whence $\mathrm{Osc}_U(g) \leq 3\varepsilon$. So we need to prove (2.139), given the assumption $g(\mathbf{p}) < \varepsilon$, which is immediate from (2.134) and (2.135).

To prove (2.139), take $\mathbf{r} \in N$ and $\mathbf{s} \in C_{\mathbf{r}} \cap E$, so $\mathbf{r} \perp \mathbf{s}$ and hence

$$g(\mathbf{r}) + g(\mathbf{s}) \leq w(g). \tag{2.140}$$

Furthermore, take $\mathbf{t}, \mathbf{u} \in E$, $\mathbf{t} \perp \mathbf{u}$, so that $(\mathbf{t}, \mathbf{u}, \mathbf{p})$ is a basis and, g being a frame function, we have

$$g(\mathbf{t}) + g(\mathbf{u}) + g(\mathbf{p}) = w(g). \tag{2.141}$$

But by construction g is constant on the equator E, so $g(\mathbf{t}) = g(\mathbf{u}) = k$, hence $2k + g(\mathbf{p}) = w(g)$, and (2.140) yields

$$g(\mathbf{r}) \leq w(g) - g(\mathbf{s}) = 2k + g(\mathbf{p}) - g(\mathbf{s}) = k + g(\mathbf{p}),$$

from which

$$k - g(\mathbf{r}) \geq -g(\mathbf{p}). \tag{2.142}$$

Furthermore, for $\mathbf{q} \in N$, $\mathbf{x}, \mathbf{r} \in D_{\mathbf{q}}$, $\mathbf{x} \perp \mathbf{r}$, there exists $\mathbf{q}' \in D_{\mathbf{q}} \cap E$ such that

$$g(\mathbf{x}) + g(\mathbf{r}) = g(\mathbf{q}) + g(\mathbf{q}') = g(\mathbf{q}) + k,$$

from which, using (2.142), we obtain

$$g(\mathbf{x}) = g(\mathbf{q}) + k - g(\mathbf{r}) \geq g(\mathbf{q}) - g(\mathbf{p}),$$

and hence

$$g(\mathbf{q}) \leq g(\mathbf{x}) + g(\mathbf{p}),\ \mathbf{q} \in N, \mathbf{x} \in D_{\mathbf{q}}. \tag{2.143}$$

Aplying this twice to the double descent definition domain (2.137), we find

$$g(\mathbf{x}) \leq g(\mathbf{y}) + g(\mathbf{p}) \leq g(\mathbf{z}) + 2g(\mathbf{p}),\ \mathbf{y} \in D_{\mathbf{x}}, \mathbf{z} \in D_{\mathbf{y}}. \tag{2.144}$$

Since (2.134) and (2.135) imply $g(\mathbf{p}) < \varepsilon$, this yields (2.139). □

Proof of Lemma 2.38. We may assume $\mathbf{p} \in U \equiv U_{\mathbf{p}}$. Using Lemma 2.39.2, by the argument to come we then move $U_{\mathbf{p}}$ to a neighborhood of $\mathbf{y}$ called $U_{\mathbf{y}}$, and subsequently repeat the argument so as to move $U_{\mathbf{y}}$ to $U_{\mathbf{x}} \equiv V$ as specified in the lemma.

We use spherical coordinates (ϕ, θ) for $\mathbf{x} = (x, y, z) \in S^2$, given by

$$(x = \cos\phi \sin\theta, y = \sin\phi \sin\theta, z = \cos\theta),\ \phi \in [0, 2\pi),\ \theta \in [0, \pi]. \tag{2.145}$$

Hence the north pole $\mathbf{p} = (0, 0, 1)$ has $\theta = 0$ and ϕ undefined (note that (ϕ, θ) are essentially (longitude, latitude), except that the latter usually starts counting downwards from $\frac{1}{2}\pi$ to $-\frac{1}{2}\pi$, with the north pole having latitude $\frac{1}{2}\pi$). Since U is open, there exists $\delta > 0$ such that all points with $0 \leq \theta < \delta$ belong to U. Pick $\mathbf{y} \in E$ as above, and define $\mathbf{r}$ as the point with the same ϕ as $\mathbf{y}$ but $\theta_{\mathbf{r}} = \theta_{\mathbf{y}} + \frac{1}{2}\delta$ (so that $\mathbf{r}$ lies a little south of $\mathbf{y}$). Then inspection of S^2 shows that one can find a neighborhood $U_{\mathbf{y}}$ of $\mathbf{y}$ with the following property: for any $\mathbf{u} \in U_{\mathbf{y}}$ there exists a great circle C through $\mathbf{r}$ and $\mathbf{u}$ that contains two further points $\mathbf{r}' \in U_{\mathbf{p}}$ and $\mathbf{u}' \in U_{\mathbf{p}}$ such that $\mathbf{r} \perp \mathbf{r}'$ and $\mathbf{u} \perp \mathbf{u}'$. Hence $h(\mathbf{r}) + h(\mathbf{r}') = h(\mathbf{u}) + h(\mathbf{u}')$. Doing this for two different points $\mathbf{u} = \mathbf{u}_1$ and $\mathbf{u} = \mathbf{u}_2$ gives

$$\begin{aligned} h(\mathbf{r}) + h(\mathbf{r}'_1) &= h(\mathbf{u}_1) + h(\mathbf{u}'_1); \\ h(\mathbf{r}) + h(\mathbf{r}'_2) &= h(\mathbf{u}_2) + h(\mathbf{u}'_2). \end{aligned}$$

Hence $h(\mathbf{u}_1) - h(\mathbf{u}_2) = h(\mathbf{r}'_1) - h(\mathbf{r}'_2) - (h(\mathbf{u}'_1) - h(\mathbf{u}'_2))$, from which we obtain

$$|h(\mathbf{u}_1) - h(\mathbf{u}_2)| \leq |h(\mathbf{r}'_1) - h(\mathbf{r}'_2)| + |(h(\mathbf{u}'_1) - h(\mathbf{u}'_2)| \leq \mathrm{Osc}_U(h) + \mathrm{Osc}_U(h) \leq 2\varepsilon',$$

for by assumption, $\mathrm{Osc}_U(h) \leq \varepsilon'$. Since $\mathbf{u}_1$ and $\mathbf{u}_2$ in $U_{\mathbf{y}}$ were arbitrary, this gives

$$\mathrm{Osc}_{U_{\mathbf{y}}}(h) \leq 2\varepsilon'. \tag{2.146}$$

Repeating this with $\mathbf{y}$ as the north pole gives $\mathrm{Osc}_{U_{\mathbf{x}}}(h) \leq 4\varepsilon'$, i.e., the lemma. □

To prove Lemma 2.35, following Gleason himself we consider the natural action of the rotation group $SO(3)$ (with positive determinant) on $\mathbb{R}^3$, written $R : \mathbf{x} \mapsto R\mathbf{x}$. This action maps S^2 onto itself and hence induces an action U on $C(S^2)$ by pullback:

$$U(R)f(\mathbf{u}) = f(R^{-1}\mathbf{u}). \tag{2.147}$$

By Lemma 2.34 we have inclusions

$$\mathscr{F}(S^2) \subset C_e(S^2) \subset C(S^2), \tag{2.148}$$

where $\mathscr{F}(S^2)$ are the frame functions and $C_e(S^2)$ consists of the even functions in $C(S^2)$; both spaces are obviously stable under the action (2.147). The following facts, due to Weyl, which we state without proof, follow from elementary representation theory, but they are also quite easily verified by explicit computation. Let

$$\psi_\ell(x,y,z) = (x+iy)^\ell,\ \ell \in \mathbb{N}, \tag{2.149}$$

and restrict this function to S^2, still calling it ψ_ℓ. Let $H_\ell \subset C(S^2)$ be the vector space spanned by all transforms $U(R)\psi_\ell$, $R \in SO(3)$. This vector space:

- consists of all homogeneous polynomials of degree ℓ that are orthogonal (with respect to the inner product in $L^2(S^2)$) to any such polynomials of degree $\ell - 2$;
- has a basis consisting of the spherical harmonics Y_ℓ^m, $m = -\ell, -\ell+1, \ldots, \ell-1, \ell$;
- accordingly, has *finite* dimension equal to $\dim(H_\ell) = 2\ell + 1$;
- is irreducible under the natural $SO(3)$-action (2.147).

Indeed, all (necessarily finite-dimensional) irreducible representations of $SO(3)$ arise in this way. Now $\mathscr{F}(S^2)$ is closed under the $SO(3)$-action (2.147), hence so must be $\mathscr{F}(S^2) \cap H_\ell$. Since H_ℓ is irreducible, there are merely two possibilities:

$$H_\ell \subset \mathscr{F}(S^2); \tag{2.150}$$
$$H_\ell \cap \mathscr{F}(S^2) = \{0\}. \tag{2.151}$$

Since for even/odd values of ℓ the space H_ℓ consist of even/odd functions, and $\mathscr{F}(S^2)$ only has even elements, we immediately see that (2.151) applies if ℓ is *odd*. For *even* values of ℓ, we see at once that (2.150) holds for:

- $\ell = 0$, where the constant frame function $f(x,y,z) = c = \frac{1}{3}w(f) \neq 0$ is obviously induced by the operator $\rho = c \cdot 1_3$ (where 1_3 is the 3×3 unit matrix), cf. (2.127);
- $\ell = 2$, which corresponds to frame functions f with weight $w(f) = 0$.

The latter functions are induced by operators ρ with zero trace. To see this, diagonalize ρ in $\mathbb{C}^3$ as in (2.6), without the constraints on p_i. This yields

$$f(\mathbf{x}) = \langle \mathbf{x}, \rho \mathbf{x} \rangle = \sum_{i=1}^{3} p_i |\langle \mathbf{x}, \upsilon_i \rangle|^2. \tag{2.152}$$

For $f \in H_2$, since $H_2 \perp H_0$ in $L^2(S^2)$ we must have

$$\langle 1_{\mathbb{R}^3}, f \rangle_{L^2(S^2)} = \int_{S^2} d^2\mathbf{x}\, f(\mathbf{x}) = 0. \tag{2.153}$$

For any $\upsilon \in \mathbb{C}^3$, we have

$$\int_{S^2} d^2\mathbf{x}\, |\langle \mathbf{x}, \upsilon \rangle|^2 = \frac{4\pi}{3} \|\upsilon\|^2; \tag{2.154}$$

to see this, write $|\langle \mathbf{x}, \upsilon \rangle|^2 = |\upsilon_x|^2 x^2 + |\upsilon_y|^2 y^2 + |\upsilon_z|^2 z^2$, and use the surface element $d^2\mathbf{x} = d\phi d\theta \sin\theta$ associated to the spherical coordinates (2.145) to compute

$$\int_{S^2} d^2\mathbf{x}\, x^2 = \int_{S^2} d^2\mathbf{x}\, y^2 = \int_{S^2} d^2\mathbf{x}\, z^2 = \frac{4\pi}{3}. \tag{2.155}$$

Therefore, from (2.152), noting that $\|\upsilon_i\|^2 = 1$ for each $i = 1, 2, 3$, we obtain

$$\int_{S^2} d^2\mathbf{x}\, f(\mathbf{x}) = \frac{4\pi}{3} \sum_{i=1}^{3} p_i = \frac{4\pi}{3} \mathrm{Tr}(\rho). \tag{2.156}$$

To settle the case $\ell \geq 4$, all we need to know about the spherical harmonics is that if ℓ is even, then, once again using spherical coordinates, one has

$$Y_\ell^m(x, y, z = 0) \sim e^{im\phi} \ (m \text{ even}); \tag{2.157}$$
$$Y_\ell^m(x, y, z = 0) = 0 \ (m \text{ odd}). \tag{2.158}$$

If (2.150) holds, then $Y_\ell^m \in \mathscr{F}(S^2)$ for each $m = -\ell, -\ell+1, \ldots, \ell-1, \ell$. But for any (even) $\ell \geq 4$, there are values of m for which Y_ℓ^m cannot be a frame function. To see this, take the following family of bases of $\mathbb{R}^3$, indexed by ϕ:

$$u_1 = (\cos\phi, \sin\phi, 0); \tag{2.159}$$
$$u_2 = (-\sin\phi, \cos\phi, 0); \tag{2.160}$$
$$u_3 = (0, 0, 1). \tag{2.161}$$

For any frame function f, the value of $f(u_1) + f(u_2) = w(f) - f(u_3)$ must therefore be independent of ϕ. However, from (2.157) - (2.158), we find

$$Y_\ell^m(u_1) + Y_\ell^m(u_2) \sim e^{im\phi} + e^{im(\phi+\pi/2)} = e^{im\phi}(1 + i^m),$$

which is independent of ϕ iff $m = 0$ or $m = 2$ (mod 4). For $\ell = 0,2$ these are indeed the only values that occur, but as soon as $\ell \geq 4$, the value $m = 4$ (among others) will ruin it. So (2.150) holds only for $\ell = 0$ and $\ell = 2$, whereas (2.151) is the case for all other $\ell \in \mathbb{N}$. Since H_0 and H_2 occur in $C(S^2)$ with multiplicity one, they cannot have greater multiplicity in $\mathscr{F}(S^2) \subset C(S^2)$, so the above argument suggests that

$$\mathscr{F}(S^2) = H_0 \oplus H_2, \tag{2.162}$$

which would prove the lemma. Fortunately, this is indeed the case, but to complete the argument we need the following technical results (left out by Gleason himself):

Lemma 2.40. *1. Frame functions are uniformly continuous.*

2. *The representation* (2.147) *of* $SO(3)$ *on* $\mathscr{F}(S^2)$ *is continuous (in the usual sense that the map* $(R, f) \mapsto U(R)f$ *from* $SO(3) \times \mathscr{F}(S^2)$ *to* $\mathscr{F}(S^2)$ *is continuous) with respect to the supremum-norm on* $\mathscr{F}(S^2)$.
3. *A continuous representation of a compact group G on a Banach space B is completely reducible (in that B is the closure of the direct sum of all irreducible representations of G that it contains).*

Proof. 1. The first claim follows because S^2 is compact. Another proof starts from the proof of Lemma 2.38, which has the feature that for given $\varepsilon' > 0$, if $y, y' \in E$ with $y' = R_z(\phi)$ for some angle ϕ, then $U_{y'} = R_z(\phi)U_y$ (this is immediately clear from the geometry). Similarly, as $x \in S^2$, different neighborhoods $V = U_x$ are related by a rotation. Hence the size of U_x is independent of x, so that the above proof of continuity established uniform continuity of frame functions also.

2. Let $R_n \to R$ in $SO(3)$ and $f_m \to f$ uniformly in $\mathscr{F}(S^2)$, i.e., $\|f_m - f\|_\infty \to 0$. Then, subtracting and adding a term $U(R_n)f$ and using isometricity of U, i.e.,
$$\|U(R_n)(f_m - f)\|_\infty = \|f_m - f\|_\infty,$$
we obtain the estimate
$$\|U(R_n)f_m - U(R)f\|_\infty \leq \|f_m - f\|_\infty + \|U(R_n)f - U(R)f\|_\infty,$$
cf. (2.147). As $m \to \infty$ the first term on the right-hand side vanishes by assumption, whilst the second vanishes as $n \to \infty$ by uniform continuity of f.
3. This is a Banach space version of the Peter–Weyl theorem, applied to the Banach space of frame functions equipped with the supremum-norm (see Notes). □

Something like this is necessary, because one needs to rule out the possibility that although (by the Stone–Weierstrass Theorem) the polynomial functions on $\mathbb{R}^3$, restricted to S^2, are uniformly dense in $C(S^2)$, so that the linear span of all spherical harmonics and hence of all H_ℓ is uniformly dense in $C(S^2)$, some frame functions might lie in the closure of this direct sum (or, in other words, they are given by uniformly convergent infinite sums of certain Y_ℓ^m). Lemma 2.40 clinches the proof of (2.162), since the third part implies that $\mathscr{F}(S^2)$ would contain all irreducible representations that contribute to the potential infinite sums; but we have already proved that it only contains H_0 and H_2. Thus Lemma 2.35 now also follows. □

2.9 Effects and Busch's Theorem

Gleason's Theorem is easy to state but difficult to prove; ***Busch's Theorem*** is a variation of it, which is more difficult to state but much easier to prove. Logically, Busch's Theorem is weaker than Gleason's, as the assumptions of the latter are contained in those of the former, but physically it appears to be more useful, as it covers more situations. To wit, Busch's Theorem revolves around certain generalizations of projections (which took the centre stage in Gleason's Theorem) called ***effects***: these are (necessarily self-adjoint) operators $a \in B(H)$ that satisfy $0 \leq a \leq 1_H$, in the sense defined after Proposition A.22. Thus $a \in B(H)$ is an effect iff

$$0 \leq \langle \psi, a\psi \rangle \leq 1 \;\; (\psi \in H). \tag{2.163}$$

The set of effects on a Hilbert space H is denoted by $\mathscr{E}(H)$ or by $[0,1]_{B(H)}$. By Theorem A.10, we have (2.163) iff $a^* = a$ and the eigenvalues λ of a lie in the interval $[0,1]$ (i.e., $\sigma(a) \subset [0,1]$). This implies that $\|a\| \leq 1$, and conversely, if $a \geq 0$, using the bound $a \leq \|a\| \cdot 1_H$ for any self-adjoint operator a, which easily follows from (A.47), we see that for $a \geq 0$, the condition $\|a\| \leq 1$ is equivalent to $a \in \mathscr{E}(H)$. In particular, it follows that both projections and density operators are effects.

Proposition 2.41. *1. The set $\mathscr{E}(H)$ of effects on H is a compact convex subset of $B(H)$ in its σ-weak topology, with extreme boundary*

$$\partial_e \mathscr{E}(H) = \mathscr{P}(H), \tag{2.164}$$

i.e., the set of all projections on H (including 0).
2. Each $a \in \mathscr{E}(H)$ has a (typically non-unique) extremal decomposition

$$a = \sum_{i=0}^{m} t_i f_i, \tag{2.165}$$

in which $t_i \geq 0$ and $\sum_i t_i = 1$, and the f_i are projections.

The σ-weak topology on $B(H)$, defined after Corollary A.31, is the right one in this context, but if H is finite-dimensional, as we assume here, this technicality may be ignored, as the claim is even true with respect to the norm topology.

Proof. In Part 1, compactness and convexity are easily checked.

The inclusion $\partial_e \mathscr{E}(H) \subseteq \mathscr{P}(H)$ is equivalent to the claim that any $a \in \mathscr{E}(H)$, $a \notin \mathscr{P}(H)$, does not lie in $\partial_e \mathscr{E}(H)$ and hence admits a convex decomposition

$$a = ta_1 + (1-t)a_2, \; t \in (0,1), a_1, a_2 \in \mathscr{E}(H), a_1 \neq a \neq a_2, \tag{2.166}$$

or, equivalently, a has a nontrivial decomposition $a = \sum_i t_i a_i$, for certain $t_i > 0$ with $\sum_i t_1 = 1$. Indeed, the latter follows from the spectral resolution (A.37), in which the spectral projections e_λ should be rescaled if necessary to as to make the coefficients sum to unity (note that $te \in \mathscr{E}(H)$ for any projection e and any $t \in [0,1]$).

To show the opposite inclusion $\mathscr{P}(H) \subseteq \partial_e \mathscr{E}(H)$, again assume (2.166), where this time $a = e \in \mathscr{P}(H)$ is a projection. "Sandwiching" between $\psi \in H_1$, this yields

$$\langle \psi, a_1 \psi \rangle = \langle \psi, a_2 \psi \rangle = 0, \; \psi \in (eH)^{\perp}; \tag{2.167}$$
$$\langle \psi, a_1 \psi \rangle = \langle \psi, a_2 \psi \rangle = 1, \; \psi \in eH. \tag{2.168}$$

Using $0 \leq a_i \leq 1$, $i = 1,2$, and (A.37), these equations imply that $a_1 = a_2 = e$.

The claim of part 2 is satisfied by picking the t_i and f_i in terms of the spectral data associated to a (cf. Theorem A.10), as follows: with $m = |\sigma(a)|$, order the eigenvalues $\lambda \in \sigma(a)$ according to $\lambda_1 < \cdots < \lambda_m$, and take:

$$t_0 = 1 - \lambda_m; \tag{2.169}$$
$$t_1 = \lambda_1; \tag{2.170}$$
$$t_i = \lambda_i - \lambda_{i-1} \; (i \geq 2); \tag{2.171}$$
$$f_0 = 0; \tag{2.172}$$
$$f_1 = 1_H; \tag{2.173}$$
$$f_i = \sum_{j=i}^{m} e_{\lambda_i} \; (i \geq 2). \tag{2.174}$$

The validity of (2.165) is then a trivial verification. □

Note that, in general, the extremal decomposition of *a as an effect* differs from its spectral resolutions (A.37) or (A.38) *as a self-adjoint operator*. If $a = \rho$ is a density operator, then the latter, i.e., (2.6), does provide an extremal decomposition of a construed as an effect also, which differs from the one in (2.165). This example shows that extremal decompositions in $\mathscr{E}(H)$ are not necessarily unique. Also, observe that te, for $e \in \mathscr{P}(H)$ and $t \in (0,1)$, does not lie in $\partial_e \mathscr{E}(H)$, since it admits a nontrivial decomposition $te = te + (1-t) \cdot 0$, recalling that $0 \in \mathscr{P}(H) \subset \mathscr{E}(H)$.

Busch's Theorem classifies the following objects.

Definition 2.42. *A* **probability distribution** *on $\mathscr{E}(H)$ is a function $p : \mathscr{E}(H) \to [0,1]$ that satisfies the following two conditions:*

1. *$p(1_H) = 1$;*
2. *If a (finite) family (a_i) of effects satisfies $\sum_i a_i \leq 1_H$, then*

$$p\left(\sum_i a_i\right) = \sum_i p(a_i). \tag{2.175}$$

Lemma 2.43. *if a (finite) family (a_i) of effects satisfies $\sum_i a_i = 1$, then $\sum_i p(a_i) = 1$.*

This trivial observation implies that a probability distribution on $\mathscr{E}(H)$ induces a probability distribution on $\mathscr{P}(H) \subset \mathscr{E}(H)$ by restriction, cf. Definition 2.23. Another way to see this from the perspective of probability measures is to note that any family (e_i) of projections that satisfies $\sum_i e_i \leq 1$ is automatically orthogonal.

Therefore, restricted to $\mathscr{P}(H)$, Definition 2.42 reduces to Definition 2.23.2. To see this, fix j and pick $\psi \in e_j H$. The condition $\sum_i e_i \leq 1$ gives

$$\sum_{i \neq j} \langle \psi, e_i \psi \rangle = \sum_{i \neq j} \|e_i \psi\|^2 \leq 0,$$

but since each term is positive, this implies $e_i \psi = 0$ for each $i \neq j$. Putting $\psi = e_j \varphi$, where $\varphi \in H$ is arbitrary, this gives $e_i e_j \varphi = 0$ for all φ and hence $e_i e_j = 0$.

Clearly, any state ω on $B(H)$ induces a probability distribution p_ω on $\mathscr{E}(H)$ by

$$p_\omega(a) = \omega(a). \tag{2.176}$$

Busch's Theorem shows the converse.

Theorem 2.44. *Any probability distribution p on $\mathscr{E}(H)$ takes the form $p = p_\omega$ for some state ω on $B(H)$, establishing a bijective correspondence between probability distributions on $\mathscr{E}(H)$ and states on $B(H)$.*

Proof. If $p : \mathscr{E}(H) \to [0,1]$ can be extended to a linear map $\omega : B(H) \to \mathbb{C}$, then ω is automatically a state, for normalization is assumed and positivity follows from the fact that any $0 \neq b \geq 0$ has the form $b = ra$ for some $r \in \mathbb{R}^+$ and $0 \leq a \leq 1_H$, namely with $r = \|b\|$ and $a = b/\|b\|$; then $a \geq 0$ and $\|a\| = 1$, so that, as explained earlier, a is an effect. Hence $\omega(b) = \omega(ra) = rp(a) \geq 0$. To achieve this extension:

1. We show that $p(ra) = rp(a)$ for all $r \in \mathbb{Q} \cap [0,1]$ and $0 \leq a \leq 1_H$. Indeed, for any such a and $n \in \mathbb{N}$ we write $a = (a + \cdots + a)/n$ (n terms), so that by (2.175), $p(a) = np(a/n)$. Similarly, for any $m \in \mathbb{N}$ and $0 \leq b \leq 1_H/m$, we have $p(mb) = mp(b)$. Take integers m, n such that $(m/n) \in [0,1]$ and put $b = a/n$, so that
$$p\left(\frac{m}{n}a\right) = mp\left(\frac{a}{n}\right) = \frac{m}{n}p(a). \tag{2.177}$$
2. We next prove that $p(ta) = tp(a)$ for all $t \in [0,1]$ and $0 \leq a \leq 1_H$. Positivity of p yields $p(a) \leq p(a')$ whenever $0 \leq a \leq a' \leq 1_H$. Given $t \in [0,1]$, take an increasing sequences of rationals (r_n) with $r_n \leq t$, as well as a decreasing sequence of rationals (s_n) with $t \leq s_n$, such that $r_n \uparrow t$ and $s_n \downarrow t$ in $\mathbb{R}$. With step 1, this gives
$$r_n p(a) = p(r_n a) \leq p(ta) \leq p(s_n a) \leq s_n p(a).$$
Letting $n \to \infty$, this gives $tp(a) \leq p(ta) \leq tp(a)$, and hence equality.
3. Now extend p to all $a \geq 0$, calling the extension ω, by $\omega(a) = \|a\| p(a/\|a\|)$ at $a \neq 0$ and $\omega(0) = 0$; the previous step then easily yields the compatibility property $\omega_{|[0,1]_{B(H)}} = p$ and the scaling property $\omega(ta) = t\omega(a)$ for each $t \geq 0$.
4. For $a \geq 0$ and $b \geq 0$, rescaling and (2.175) yield $\omega(a+b) = \omega(a) + \omega(b)$.
5. For general $a^* = a$ we write $a = a_+ - a_-$, with $a_\pm \geq 0$, as in Proposition A.24, and define ω on all of $B(H)_{\text{sa}}$ by $\omega(a) = \omega(a_+) - \omega(a_-)$. This is well defined despite the lack of uniqueness of (A.74), for if $a = a_+ - a_- = a'_+ - a'_-$, with $a'_\pm \geq 0$, then $a_+ + a'_- = a'_+ + a_-$, whence $\omega(a_+) - \omega(a_-) = \omega(a'_+) - \omega(a'_-)$.

This argument also shows that ω remains linear on general self-adjoint a and b, since $a+b=(a_{+}+b_{+})-(a_{-}+b_{-})$ is a decomposition with $(a_{\pm}+b_{\pm})\geq 0$.
6. Finally, for general $c\in B(H)$ we (uniquely) decompose $c=a+ib$, $a^{*}=a$, $b^{*}=b$, cf. the proof of Corollary A.20, and put $\omega(c)=\omega(a)+i\omega(b)$. □

To close, we give a very brief and superficial introduction to effects as they arise from modern ("operational") quantum measurement theory. This theory associates quantum data to classical data through the concept of a ***Positive Operator Valued Measure*** or ***POVM***. Relative to some given "classical" space X (taken finite here) and Hilbert space H (assumed finite-dimensional), a POVM is defined as a map

$$\mathsf{A}:\mathscr{P}(X)\rightarrow\mathscr{E}(H) \tag{2.178}$$

that satisfies $\mathsf{A}(X)=1_{H}$ as well as $\mathsf{A}(U\cup V)=\mathsf{A}(U)+\mathsf{A}(V)$ whenever $U\cap V=\emptyset$, cf. Definition 1.1. Equivalently, a POVM is a map

$$\mathsf{a}:X\rightarrow\mathscr{E}(H) \tag{2.179}$$

that satisfies

$$\sum_{x\in X}\mathsf{a}(x)=1_{H}. \tag{2.180}$$

As in the classical case, these notions are trivially equivalent through

$$\mathsf{a}(x)=\mathsf{A}(\{x\}); \tag{2.181}$$

$$\mathsf{A}(U)=\sum_{x\in U}\mathsf{a}(x). \tag{2.182}$$

The motivating special case of a POVM is given by some self-adjoint operator $a\in B(H)$, which yields $X=\sigma(a)$ and $\mathsf{a}(\lambda)=e_{\lambda}$. In that case, each density operator ρ induces a probability distribution on $\sigma(a)$ through the Born rule (2.8). More generally, a probability distribution p on $\mathscr{E}(H)$ and a POVM (2.179) jointly determine a probability distribution p_{a} on X, given by

$$p_{\mathsf{a}}(x)=p(\mathsf{a}(x)). \tag{2.183}$$

Indeed, $p_{\mathsf{a}}(x)\geq 0$ because $a\geq 0$, and $\sum_{x\in X}p_{\mathsf{a}}(x)=1$ by (2.180) and Lemma 2.43. The idea, then, is that a measurement of some POVM a has (classical) outcome x with probability $p_{\mathsf{a}}(x)$; this generalizes the traditional dogma that a measurement of an observable a has outcome $\lambda\in\sigma(a)$ with (Born) probability (2.8). Indeed, combined with (2.33), Busch's Theorem shows that we necessarily have

$$p_{\mathsf{a}}(x)=\mathrm{Tr}\,(\rho\mathsf{a}(x)), \tag{2.184}$$

for some density operator ρ. So nothing has been gained by introducing Definition 2.42, expect perhaps for the insight that, as in Gleason's Theorem, it is the non-contextuality of a probability distribution on $\mathscr{E}(H)$—in that $p(a(x))$ is independent of the POVM a which $a(x)$ forms part of—that eventually enforces (2.184).

2.10 The quantum logic of Birkhoff and von Neumann

In §1.4 we showed that *classical* mechanics has a *classical* logical structure, in which (equivalence classes of) propositions correspond to subsets of phase space. These subsets form a Boolean lattice in which the logical connectives $\neg$, $\wedge$, and $\vee$ fornegation, disjunction, and conjunction, respectively, are interpreted as their natural set-theoretic counterparts (i.e., complementation, intersection, and union).

In 1936, Birkhoff and von Neumann proposed a strikingly similar *quantum* logic for *quantum* mechanics, in which (closed) linear subspaces of Hilbert space play the role of (measurable) subsets of phase space, and the basic logical connectives (except implication, which is queerly lacking in this setting) are interpreted as:

$$\neg L = L^{\perp}; \tag{2.185}$$
$$L \wedge M = L \cap M; \tag{2.186}$$
$$L \vee M = L + M, \tag{2.187}$$

where $L^{\perp}$ is the orthogonal complement of L, see (A.29), $L \cap M$ is the (set-theoretic) intersection of L and M, and $L+M$ is the (closed) linear span of L and M. If $\dim(H) < \infty$, as we continue to assume, any linear subspace of H is automatically closed, and the infinite-dimensional case an attractive operator-algebraic and lattice-theoretic structure arises only if the events are taken to be *closed* linear subspaces.

Although the Brouwer–Hilbert debate on the foundations of mathematics had somewhat subsided in 1936, with hindsight it may be argued that the quantum logic of Birkhoff and von Neumann (who had been a "postdoc" *avant la lettre* with Hilbert) was predicated on their desire to preserve not only the ***law of contradiction***

$$\alpha \wedge \neg \alpha = \perp, \tag{2.188}$$

where α is any proposition and $\perp$ is the proposition that is identically false, but also, against Brouwer, the ***law of excluded middle*** (or ***tertium non datur***)

$$\alpha \vee \neg \alpha = \top, \tag{2.189}$$

where $\top$ is the proposition that is identically true. Indeed, in the Birkhoff–von Neumann model (2.185) - (2.187), where $\perp = \{0\}$ and $\top = H$, these are identities. Similarly, their model satisfies the ***law of double negation***

$$\neg\neg\alpha = \alpha, \tag{2.190}$$

which both in classical logic (where it is a tautology) and in intuitionistic logic (where it is rejected in general) is equivalent to (2.189). Also, ***De Morgan's Laws***:

$$\neg(\alpha \vee \beta) = \neg\alpha \wedge \neg\beta; \tag{2.191}$$
$$\neg(\alpha \wedge \beta) = \neg\alpha \vee \neg\beta, \tag{2.192}$$

hold in their quantum logic (despite their origin in *classical* propositional logic).

We will now derive the Birkhoff–von Neumann structure along similar lines as its classical counterpart (cf. §1.4), except that in the absence of the necessary structure for a classical propositional calculus we now rely on semantic entailment alone.

In quantum theory, the role of functions $f : X \to \mathbb{R}$ as observables in classical physics is played by self-adjoint operators $a : H \to H$ on some Hilbert space H, and hence the quantum analogue of an elementary proposition $f \in \Delta$ of classical physics is $a \in \Delta$ (where $\Delta \subset \mathbb{R}$), with special case $a = \lambda$ for $a \in \{\lambda\}$ (with $\lambda \in \mathbb{R}$).

In analogy to the points $x \in X$ of phase space, pure states ω_ψ as in (2.42), or the corresponding density operators e_ψ (where $\psi \in H$ is a unit vector), yield truth assignments to elementary propositions. To start with the simplest case, $a = \lambda$ is:

- ***true*** with respect to ω_ψ iff $p_a^\psi(\lambda) = 1$, see (2.10), or, equivalently, iff $\psi \in H_\lambda$, where $H_\lambda \subseteq H$ is the eigenspace of a for eigenvalue λ, cf. (A.36);
- ***false*** with respect to ω_ψ iff $p_a^\psi(\lambda) = 0$, or, equivalently, iff $\psi \perp H_\lambda$.

The underlying idea here is arguably that, according to some naive operational interpretation of quantum mechanics, a measurement of a in a state ω_ψ would give outcome λ with probability one (zero) iff $a = \lambda$ is true (false) with respect to ω_ψ. If $0 < p_a^\psi(\lambda) < 1$, the "truthmaker" ω_ψ actually *fails to assign a truth value* to $a = \lambda$; the *partial* nature of truthmakers marks a significant difference with the classical case, as does the closely related distinction between *false* and *not true*. Similarly, we say that an elementary proposition $a \in \Delta$ is ***true*** in some state ω_ψ iff

$$P_a^\psi(\Delta) \equiv \|e_\Delta \psi\|^2 = 1, \tag{2.193}$$

cf. (2.9) and (A.42), and ***false*** if $P_a^\psi(\Delta) = 0$. In other words, $a \in \Delta$ is true in ω_ψ iff $\psi \in H_\Delta$, and false if $\psi \perp H_\Delta$,see (A.43). Such propositions may formally be combined using the connectives $\neg$, $\wedge$, and $\vee$ (whose meaning is unfortunately far from clear in this new setting) according to the same (inductive) formation rules as in classical propositional logic. However, the classical truth tables for $\wedge$ and $\vee$ are unsound with regard to the above rules, at least if one eventually wants to arrive at (2.185) - (2.187). For example, ω_ψ may validate neither α nor β, yet it might make $\alpha \vee \beta$ true (assuming that α and β correspond to L and M, respectively, this is the case if $\psi \notin L$ and $\psi \notin M$, yet $\psi \in L + M$). Similarly, ω_ψ may render neither α nor β false, yet it may falsify $\alpha \wedge \beta$. Due to this complication, the approach of §1.4 has to be modified, as follows. Our goal remains to define a ***semantic equivalence relation*** $\sim_H$, which is predicated on an inductive definition of truth we first give.

Definition 2.45. *1. $a \in \Delta$ is* **true** *in ω_ψ iff $P_a^\psi(\Delta) = 1$, and* **false** *if $P_a^\psi(\Delta) = 0$.*
2. The negation $\neg(a \in \Delta)$ of an elementary proposition $a \in \Delta$ is given by $a \in \Delta^c$.
3. The negation $\neg\alpha$ is true iff α is false.
4. The conjunction $\alpha \wedge \beta$ is true iff both α and β are true.
5. De Morgan's Laws (2.191) - (2.192) *and the law of double negation* (2.190) *hold; in particular, the disjunction $\alpha \vee \beta$ is true iff $\neg(\neg\alpha \wedge \neg\beta)$ is true (as per 1–4).*
6. We write $\alpha \models_H \beta$ iff the truth of α implies the truth of β, for each state ω_ψ.
7. We write $\alpha \sim_H \beta$ iff $\alpha \models_H \beta$ and $\beta \models_H \alpha$.
8. If $\alpha \sim_H \beta$, then $\neg\alpha \sim_H \neg\beta$.

Lemma 2.46. *Definition 2.45 implies the following rules:*

1. *Our earlier truth attributions for the case* $a \in \Delta$ *with* $\Delta = \{\lambda\}$. *In particular,* $a = \lambda$ *is always false when* $\lambda \notin \sigma(a)$, *and so is* $a \in \Delta$ *whenever* $\Delta \cap \sigma(a) = \emptyset$.
2. $a \in \Delta$ *is false relative to* ω_ψ *iff* $\psi \perp H_\Delta$.
3. $(a \in \Delta) \wedge (b \in \Gamma)$ *is true in* ω_ψ *iff* $\psi \in H_\Delta^{(a)} \cap H_\Gamma^{(b)}$.
4. $(a \in \Delta) \vee (b \in \Gamma)$ *is true in* ω_ψ *iff* $\psi \in H_\Delta^{(a)} + H_\Gamma^{(b)}$.

Hence conjunctions behave classically, as part 3 states that $(a \in \Delta) \wedge (b \in \Gamma)$ is true iff $a \in \Delta$ and $b \in \Gamma$ are true). The proof of this lemma uses the following notation.

Definition 2.47. *If* e *and* f *are projections on a Hilbert space* H, *then:*

- $e \wedge f$ *is the projection onto* $eH \cap fH$;
- $e \vee f$ *is the projection onto* $eH + fH$, *i.e., the (closed) linear span of* eH *and* fH.

Note that if e and f commute, these reduce to the algebraic expressions

$$e \wedge f = ef; \tag{2.194}$$
$$e \vee f = e + f - ef. \tag{2.195}$$

Furthermore, in case of potential ambiguity we will write $e_\Delta^{(a)}$ for the spectral projection e_Δ as defined by a, and analogously $e_\Gamma^{(b)}$, etc. Similarly for $H_\Delta^{(a)}$ etc.

Proof. The first and third claims are immediate. The second one follows from the relation $e_{\Delta^c} = e_\Delta^\perp = 1 - e_\Delta$, or, equivalently, $H_{\Delta^c} = H_\Delta^\perp$. For the fourth, use Definition 2.45.6, 3, and 2 to infer that $(a \in \Delta) \vee (b \in \Gamma)$ is true iff $(a \in \Delta^c) \wedge (b \in \Gamma^c)$ is false. From the third claim, we note that

$$(a \in \Delta) \wedge (b \in \Gamma) \sim_H \left(e_\Delta^{(a)} \wedge e_\Gamma^{(b)} = 1\right), \tag{2.196}$$

so by Definition 2.45.5, $(a \in \Delta^c) \wedge (b \in \Gamma^c)$ is false iff $e_{\Delta^c}^{(a)} \wedge e_{\Gamma^c}^{(b)} = 1$ is false. Since $e_{\Delta^c}^{(a)} \wedge e_{\Gamma^c}^{(b)} = 1$ is true iff $\psi \in H_{\Delta^c}^{(a)} \cap H_{\Gamma^c}^{(b)}$, claim 2 implies $e_{\Delta^c}^{(a)} \wedge e_{\Gamma^c}^{(b)} = 1$ is false iff

$$\psi \in (H_{\Delta^c}^{(a)} \cap H_{\Gamma^c}^{(b)})^\perp = ((H_\Delta^{(a)})^\perp \cap (H_\Gamma^{(b)})^\perp)^\perp = (H_\Delta^{(a)})^{\perp\perp} + (H_\Gamma^{(b)})^{\perp\perp} = H_\Delta^{(a)} + H_\Gamma^{(b)},$$

which finishes the proof. □

Quite analogously to the classical case, Definition 2.45 implies

$$(a \in \Delta) \models_H (b \in \Gamma) \text{ iff } e_\Delta^{(a)} \subseteq e_\Gamma^{(b)}, \tag{2.197}$$

which, once again, immediately yields $(a \in \Delta) \sim_H (b \in \Gamma)$ iff $e_\Delta^{(a)} = e_\Gamma^{(b)}$. Taking $b = e_\Delta^{(a)}$ and $\Gamma = \{1\}$, analogously to (1.53), as in the above proof we have

$$a \in \Delta \sim_H e_\Delta^{(a)} = 1. \tag{2.198}$$

Furthermore, as in the proof of Lemma 2.46 we find

$$(a \in \Delta) \wedge (b \in \Gamma) \sim_H \left(e_\Delta^{(a)} \wedge e_\Gamma^{(b)} = 1\right); \tag{2.199}$$

$$(a \in \Delta) \vee (b \in \Gamma) \sim_H \left(e_\Delta^{(a)} \vee e_\Gamma^{(b)} = 1\right). \tag{2.200}$$

Consequently, we have the following counterpart of Lemma 1.19:

Lemma 2.48. *Any elementary or composite proposition is semantically equivalent (relative to H) to one of the form $e = 1$, for some projection e. Furthermore,*

$$\neg(e = 1) \sim_H \left(e^\perp = 1\right); \tag{2.201}$$

$$(e = 1) \wedge (f = 1) \sim_H (e \wedge f = 1); \tag{2.202}$$

$$(e = 1) \vee (f = 1) \sim_H (e \vee f = 1). \tag{2.203}$$

At last, the quantum version of Theorem 1.20 reads as follows:

Theorem 2.49. *The set $\mathscr{Q}(H)$ of equivalence classes $[\cdot]_H$ of propositions generated by the elementary propositions $a \in \Delta$ and the logical connectives $\neg$, $\vee$, and $\wedge$, is isomorphic to the set $\mathscr{L}(H)$ of linear subspaces of H, under the map*

$$\varphi : \mathscr{Q}(H) \overset{\cong}{\to} \mathscr{L}(H); \tag{2.204}$$

$$\varphi([a \in \Delta]_H) = e_\Delta^{(a)} H. \tag{2.205}$$

Under this isomorphism, the logical connectives $\neg$, $\wedge$ and $\vee$ turn into orthogonal complementation $(-)^\perp$, intersection $\cap$, and linear span $+$, respectively, in that

$$\varphi([\neg\alpha]_H) = \varphi([\alpha]_X)^\perp; \tag{2.206}$$

$$\varphi([\alpha \wedge \beta]_H) = \varphi([\alpha]_H) \cap \varphi([\beta]_H; \tag{2.207}$$

$$\varphi([\alpha \vee \beta]_H) = \varphi([\alpha]_H) + \varphi([\beta]_H), \tag{2.208}$$

Furthermore, if we define a partial order $\leq$ on $\mathscr{Q}(X)$ by saying that $[\alpha]_H \leq [\beta]_H$ iff $\alpha \models_H \beta$ (which is well defined), then φ maps $\leq$ into set-theoretic inclusion $\subseteq$, i.e.,

$$[\alpha]_H \leq [\beta]_H \text{ iff } \varphi([\alpha]_H) \subseteq \varphi([\beta]_H). \tag{2.209}$$

With respect to these operations, $\mathscr{L}(H)$ is a **modular lattice** *(granted that* $\dim(H) < \infty$*; otherwise, the lattice is merely* **orthomodular***, cf. §D.1 for terminology).*

Proof. Most of this is immediate from Lemma 2.48, expect for the last claim, which follows from simple computations (and from the Amemiya–Araki Theorem). □

As in the classical case, there is an algebraic reformulation of this result, obtained from the bijective correspondence between (closed) linear subspaces L of H and projections e on H, given by $L = eH$ (see Proposition A.8).

Theorem 2.50. *The set $\mathscr{Q}(H)$ of equivalence classes $[\cdot]_H$ of propositions generated by the elementary propositions $a \in \Delta$ and the logical connectives $\neg$, $\vee$, and $\wedge$, is isomorphic to the set $\mathscr{P}(H)$ of projections on H, under the map*

$$\varphi' : \mathscr{Q}(H) \stackrel{\cong}{\to} \mathscr{P}(H); \tag{2.210}$$
$$\varphi'([a \in \Delta]_H) = e^{(a)}_{\Delta}, \tag{2.211}$$

where (once again) $\mathscr{P}(H)$ is the set of all projections on H.

Under this map, the logical connectives $\neg$, $\wedge$ and $\vee$ turn into (cf. Definition 2.47):

$$\varphi'([\neg\alpha]_H) = 1 - \varphi'([\alpha]_X) \tag{2.212}$$
$$\varphi'([\alpha \wedge \beta]_H) = \varphi'([\alpha]_H) \wedge \varphi'([\beta]_H); \tag{2.213}$$
$$\varphi'([\alpha \vee \beta]_H) = \varphi'([\alpha]_H) \vee \varphi'([\beta]_H), \tag{2.214}$$

Furthermore, φ' maps the partial order $\leq$ on $\mathscr{Q}(H)$ into the partial order on $\mathscr{P}(H)$ defined by $e \leq f$ iff $eH \subseteq fH$, or equivalently, iff $ef = e$.

Finally, with respect to these operations, $\mathscr{P}(H)$ is an (ortho)modular lattice.

However, unlike (1.65) - (1.68), this result is somewhat unsatisfactory in not being purely algebraic. This may partly be remedied through expressions like

$$e \wedge f = \lim_{n \to \infty} (e \circ f)^n; \tag{2.215}$$
$$e \vee f = 1 - ((1-e) \wedge (1-f)), \tag{2.216}$$

where $e \circ f = ef + fe$, and the (strong) limit in (2.215) should be taken on fixed vectors $\psi \in H$ (upon which it exists in the norm-topology of H). Even so, this specific limit still relies on the underlying Hilbert space, and in any case the expressions fail to be purely algebraic and look pretty artificial. Indeed, the same may be said about Definition 2.45, which, of course, has been fine-tuned with hindsight in order to obtain the "desired" answer in the form of Theorem 1.20, which in turn vindicates the mathematically sweet Birkhoff–von Neumann *Ansatz* (2.185) - (2.187).

In addition, there are serious *conceptual* objections to this kind of quantum logic:

1. Conjunction $\wedge$ and disjunction $\vee$ do not distribute over each other, rendering their interpretation as "and" and "or" obscure.
2. There are propositions α and β (namely those for which $\varphi'([\alpha]_H)$ and $\varphi'([\beta]_H)$ do not commute) for which the conjunction $\alpha \wedge \beta$ is physically undefined.
3. There are states in which $\alpha \vee \beta$ is true whilst neither α nor β is true.
4. There are states in which $\alpha \wedge \beta$ is false whilst neither α nor β is false.
5. In view of Schrödinger's Cat, one would expect the law of excluded middle (2.189) to *fail* in quantum mechanics, yet it *holds* in quantum logic (and this is possible because neither $\vee$ nor $\neg$ has any familiar logical meaning in it).
6. Finally, nothing is said or done about propositions that are neither true nor false.

In Chapter 12, we will therefore replace the doomed quantum logic of Birkhoff and von Neumann by the intuitionistic logic of Brouwer and Heyting.

Notes

All operator theory for this chapter may be found in Kadison & Ringrose (1983).

§2.1. **Quantum probability theory and the Born rule**

The Born rule was first stated by Born (1926b) in the context of scattering theory, following the earlier paper (Born, 1926a) in which Born omitted the absolute value squared signs (corrected in a footnote added in proof). The application to the position operator is due to Pauli (1927), who merely spent a footnote on it. The general formulation is due to von Neumann (1932, §III), following earlier contributions by Dirac (1926b) and Jordan (1927). Both Born and Heisenberg acknowledge the profound influence of Einstein on the probabilistic formulation of quantum mechanics. However, Born and Heisenberg as well as Bohr, Dirac, Jordan, Pauli and von Neumann differed with Einstein about the fundamental nature of the Born probabilities and hence on the issue of determinism. Indeed, whereas Born and the others just listed after him believed the outcome of any individual quantum measurement to be unpredictable in principle, Einstein felt this unpredictability was just caused by the incompleteness of quantum mechanics (as he saw it). See, for example, the invaluable correspondence between Einstein and Born (2005).

Mehra & Rechenberg (2000) provide a very detailed reconstruction of the historical origin of the Born rule within the context of quantum mechanics, whereas von Plato (1994) embeds a briefer historical treatment of it into the more general setting of the emergence of modern probability theory and probabilistic thinking. For the earlier history of probability see Hacking (1975, 1990). See also Landsman (2009).

§2.2. **Quantum observables and states**

Proposition 2.10 is due to von Neumann; see also Chapter 6.

§2.3. **Pure states in quantum mechanics**

This kind of thinking goes back to von Neumann (1932) and Segal (1947ab).

§2.4. **The GNS-construction for matrices**

Again, see §C.12 for the GNS-construction in general.

§2.5. **The Born rule from Bohrification**

See notes to §4.1.

§2.6. **The Kadison–Singer Problem**

The Kadison–Singer Problem was first discussed in Kadison & Singer (1959). See the Notes to §4.3 for more information.

§2.7. **Gleason's Theorem**

§2.8. **Proof of Gleason's Theorem**

Gleason's Theorem is due to Gleason (1957), whose proof we largely follow, with some simplifications due to Varadarajan (1985) and Hamhalter (2004). Lemma 2.40.3 or some analogous result is lacking from these references; it may be found in Lyubich (1988), Chapter 4, §2, Theorem. It is often claimed that Gleason's proof has been superseded by the more elementary one due to Cooke, Keane, & Moran (1985), which avoids all use of harmonic analysis. A similar proof, following up on Cooke et al but using constructive analysis only, was given by Richman & Bridges (1999). However, both because Gleason's use of rotation invariance is very natural, and also since the proof of Cooke et al has already been presented and simplified in two monographs entirely devoted to Gleason's Theorem, viz. Dvurečenskij (1993) and Hamhalter (2004), as well as in the highly efficient book by Kalmbach (1998), we prefer to return to the original source (and add some technical details).

§2.9. **Effects and Busch's Theorem**

Busch's Theorem is from Busch (2003), whose proof we follow almost *verbatim*. See also Caves et al (2004). For the use of POVM's in quantum physics see, e.g., Busch, Grabowski, & Lahti (1998), Davies (1976), Holevo (1982), Kraus (1983), Landsman (1998a, 1999), de Muynck (2002), and Schroeck (1996).

§2.10. **The quantum logic of Birkhoff and von Neumann** Our discussion is based on Rédei (1998), with some modifications though. The original source is Birkhoff & von Neumann (1936).