Lecture 5 Handout: The cornerstones of quantum theory

Bryan W. Roberts

1. Cornerstone 1: Gleason's theorem

Why does quantum mechanics use Hilbert space vectors, norms and traces? Surely: it works to describe a certain class of experiment. But what kinds of experiments? Gleason's theorem tells us: *any probabilistic experiment*, in a certain generalised sense of probability.

1.1. **Quantum logic.** Classical probabilities are defined on a Boolean logic of propositions. They always require a *probability space*, which is a triple (S, Σ, μ) in which S is a set, Σ is a collection of subsets forming a σ-algebra, and $\mu : \Sigma \to [0, 1]$ is a measure. The σ-algebra is a Boolean logic[1], which is widely viewed as defining the axioms for classical logic. This allows one to view each subset A is a proposition describing an outcome, with $A \cap B$ meaning 'A and B', $A \cup B$ meaning 'A or B', and the complement $A^\perp$ meaning 'not A'. The additional 'sigma' property of countable closure ensures that this structure satisfies the Kolmogorov axioms for probability when S is uncountable; otherwise, difficulties associated with non-measurable abound.

The problem is a Boolean logic is that it fails to describe the statistics of quantum systems observed in nature. The problem can be isolated precisely, that a σ-algebra satisfies the distributive law,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \tag{1}$$

A toy example can be used to explain why this isn't appropriate for quantum mechanics: suppose we have a quantum system satisfying the position-momentum uncertainty principle: let $J \subset \mathbb{R}$ be a small standard deviation momentum measurement, and let $I_1, I_2 \subset \mathbb{R}$ each be a small standard deviation for position measurement in adjacent regions of space, such that if a J-measurement is made, then neither I_1 or I_2 measurements are possible, but that they *can* be measured within the accuracy of $I_1 \cup I_2$. Make the definitions:

- $A =$ 'the particle's momentum is in J'
- $B =$ 'the particle's position is in I_1'
- $C =$ 'the particle's momentum is in I_2'

[1] A *Boolean logic* is an algebra for which the two binary operations $\cap$ and $\cup$ are associative, commutative, distributive, satisfy 'absorption' $A \cup (A \cap B) = A$ and $A \cap (A \cup B) = A$, each admit an identity $A \cup 0 = A$ and $A \cap 1 = A$, and which admit a complement operation $\perp$ satisfying $A \cup (A^\perp) = 1$ and $A \cap (A^\perp) = 0$. These are widely viewed as the axioms for a classical logic of propositions.

Then $A \cap (B \cup C)$ is true, but both $(A \cap B)$ and $(A \cap C)$ are false, so that their union is false as well. That is, the distributive law fails in quantum mechanics.

Fortunately, the early developers of quantum mechanics found[2] that the problem can be solved by replacing the distributive axiom with a logically weaker alternative, called the orthomodularity axiom, which says that distributivity holds in the following particular circumstance:

$$\text{(2)} \qquad \text{If } A \subseteq B \text{ and } A^{\perp} \subseteq C, \text{ then } A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

where $A \subseteq B$ is defined to hold iff $A = A \cap B$. This solves the problem in the example above, since the antecedent $A \subseteq B$ (i.e. $A = A \cap B$) is false, and so distributivity need not hold here. The resulting structure $\mathcal{L}$ is sometimes called a orthomodular lattice, or more simply a quantum logic.

Strictly speaking, it is not possible to define a probability measure on an orthomodular lattice, since probability measures are by definition associated with a (Boolean) σ-algebra. However, it is possible to define a natural generalisation of probability on an orthomodular lattice $\mathcal{L}$. This is sometimes called a *generalised probability measure*, defined to be a function $p : \mathcal{L} \to [0, 1]$ that is σ-additive on mutually orthogonal elements.[3] A probability measure is obviously an example of a generalised probability measure, but the latter may be applied to more general algebraic structures.

1.2. **Hilbert spaces for generalised probabilities.** Calling the structure above a 'quantum logic' is motivated in part by the fact that the standard framework for quantum theory, Hilbert spaces, provide orthomodular lattice. In particular, we have:

Fact 1 (Hilbert space orthomodularity)**.** *The subspaces of a (possibly ∞-dimensional) Hilbert space $\mathcal{H}$ form an orthmodular lattice (where $\cap$ is subspace intersection, and $\cup$ is the linear closure under subspace union, and $\perp$ is the complement).*[4]*, which are the standard framework for quantum theory.*

This fact obviously applies as well to the set of projection operators $\mathcal{P}(\mathcal{H})$ associated with a Hilbert space $\mathcal{H}$. For this reason, we sometimes refer to the projections as the (orthomodular) 'Hilbert lattice of projection operators'. Not only that: the Born rule, which is the conventional prescription for 'probabilities' in quantum mechanics is in fact a generalised probability measure on this lattice:

[2](Birkhoff and Neumann; 1936)

[3]Additivity on mutually orthogonal elements means that if $A \subseteq B^{\perp}$, then $p(A \cup b) = p(A) + p(B)$. Note that this implies that since 0 and 1 are orthogonal, $p(1) = p(0 \cup 1) = p(0) + p(1)$, and hence that $p(0) = 0$ and $p(1) = 1$.

[4]Check this! For a reference, see Jauch (1968, §2-5).

Fact 2 (Hilbert space 'probabilities'). *For each density operator*[5] $\rho : \mathcal{H} \to \mathcal{H}$*, the 'Born rule map' on projections* $p : \mathcal{P}(\mathcal{H}) \to [0,1]$ *defined by,*

$$p(E) = \mathrm{Tr}(\rho E)$$

is a generalised probability measure. (Note that, when ρ *is a vector-state associated with a unit-norm vector* $\psi \in \mathcal{H}$*, this is just the expectation value,* $p(E) = \langle \psi, E\psi \rangle$*.)*

What is remarkable is that there is a sense in which the converse is also true. That is, every generalised probability measure can be implemented on a Hilbert space using the Born rule. This is the content of a representation theorem due to Gleason. Its proof is deep and non-trivial.

Fact 3 (Gleason's theorem). *For every generalised probability measure* $\rho : \mathcal{L} \to [0,1]$ *on an orthomodular lattice* $\mathcal{L}$ *with* $\dim(\mathcal{L}) > 2$*, there exists an isomorphism to a lattice* $\mathcal{P}(\mathcal{H})$ *of Hilbert space projections and a fixed density operator* ρ *such that, writing the association of the isomorphism as* $A \mapsto E_A$*, we have,*

$$p(A) = \mathrm{Tr}(\rho E_A)$$

for all $A \in \mathcal{L}$*.*

The upshot of Gleason's theorem is that we can always use Hilbert spaces and the Born rule to represent the kinds of statistical phenomena associated with a generalised probability measure. The result thus provides a substantial justification for the standard formalism.

That said, when we are collecting direct statistical information in a laboratory, we are normally using *classical* probabilities. So, an explanation of this practice is still needed. That explanation comes from the second cornerstone of quantum theory.

2. Cornerstone 2: The spectral theorem

2.1. **Projection valued measures.** We learned from Gleason's theorem that the outcomes of a statistical experiment can always be viewed as associated with elements of a lattice of projections $\mathcal{P}(\mathcal{H})$. But, a statistical experiment also normally involves a set of outcomes that are 'mutually exclusive' and constitute 'all that is possible'. We can formalise these assumptions in the context of a Boolean logic or an orthomodular lattice $\mathcal{L}$ in a straightforward way. Suppose that for a statistical experiment:

- (**finite dimensions**) Each outcome is a projection in sequence $E_1, E_2, \ldots, E_n$ that are mutually orthogonal ($E_i E_j = 0$ for $i \neq j$), form a decomposition of unity ($\sum_i E_i = I$).
- (**infinite dimensions**) Each outcome is a projection in a *projection-valued measure* (PVM) on a σ-algebra over a set X (usually the Borel sets over the reals), i.e., $\Delta \mapsto E_\Delta$ that satisfies additivity for countable[6] disjoint sets

[5] A *density operator* $\rho : \mathcal{H} \to \mathcal{H}$ is a positive linear operator of unit trace, $\mathrm{Tr}(\rho) = 1$.

[6] If it is infinite, then this sum is the limit of partial sums.

$E_{\Delta\cup\Delta'\cup\cdots} = E_\Delta + E_{\Delta'} + \cdots$, and is a decomposition of unity in that $E_X = I$. (This implies that $E_{\Delta\cap\Delta'} = E_\Delta E_{\Delta'} = E_{\Delta'}E_\Delta$, and that $\Delta \cap \Delta' = \varnothing$ only if $E_\Delta E_{\Delta'} = 0$.)

Note that the former is in fact a special cases of the latter, associated with a finite partition of $\mathbb{R}$.

A projection-valued measure is a σ-algebra capable of supporting a classical probability measure. Gleason's theorem ensures that this can always be captured using Hilbert space and the Born rule. What remains is to introduce quantitative values into our statistical experiments.

2.2. **Outcome values and the spectral theorem.** In physics, the outcomes of an experiment often involve measurement of a quantitative value, like '671,000,000 mph'. Such numerical values are also required if we want to consider the *expected value* of a random variable; for example, when we associated the outcome 'coin lands heads' with 1 and 'coin lands tails' with 0, then the expected value of a fair coin toss can be identified with 1/2.

If our outcomes are represented by a numbers, then we can achieve this by summing all elements of a projection-valued measure $\Delta \mapsto E_\Delta$, with each projection 'weighted' by a numerical value associated with that outcome. That is: let the outcomes of our experiment be $E_1, E_2, \ldots$, and let the corresponding numerical outcomes be $\alpha_1, \alpha_2, \ldots$; then define an operator,

$$A = \alpha_1 E_1 + \alpha_2 E_2 + \cdots \tag{3}$$

where the E_i form a projection-valued measure. (For a continuous PVM on Borel sets $\Delta \mapsto E_\Delta$, we integrate $A = \int_{\mathbb{R}} \lambda dE_{(-\infty,\lambda)}$. Then, for any unit-norm vector state $\psi \in \mathcal{H}$, the Born rule for our statistics will produce the expected value, i.e. the sum of the outcome values 'weighted' by their respective probabilities:

$$\langle \psi, A\psi\rangle = \alpha_1 \underbrace{\langle \psi, E_1\psi\rangle}_{\Pr(\text{outcome 1})} + \alpha_2 \underbrace{\langle \psi, E_2\psi\rangle}_{\Pr(\text{outcome 2})} + \cdots . \tag{4}$$

A PVM $\Delta \mapsto E_\Delta$ has the structure of a σ-algebra with the operations of addition, multiplication and complement. With a unit-norm state ψ, the Born rule provides a classical probability measure on that algebra. So, the operator A thus has a special property: for each state ψ, it encodes the expectation value for a classical statistical experiment. If an operator can be decomposed into a projection-valued measure in this way, then it is called its *spectral decomposition* or *spectral measure*. It is of interest to experimental practice to know: which operators have this property?

You may know that the self-adjoint operators[7] have it; this is sometimes presented as the 'spectral theorem'. However, the more general class of operators that have a spectral decomposition as well. A linear operator A is called *normal* if it satisfies either of the two (equivalent[8]) properties:

- $AA^* = A^*A$
- $A = B + iC$ for some pair of commuting self-adjoint operators B, C.

Many familiar operators are normal; for example, every unitary operator is normal. In the Schrödinger (position) representation, so is the operator $Q + iQ$. However, $Q + iP$ in that representation is not normal, since $[Q, P] = i\hbar \neq 0$.

The full statement of the spectral theorem is a property of normal operators (Conway; 1990, Thm. X.4.11):

Fact 4 (Spectral theorem)**.** *If A is a normal operator on a Hilbert space, then there exists a unique projection-valued measure on Borel sets of $\mathbb{C}$, written in the discrete case as $i \mapsto E_i$, such that,*

$$A = \sum_i^n \zeta_i G,$$

where each $\zeta_i \in \mathbb{C}$ is an eigenvalue of A. An operator B commutes with A iff it commutes with each projection E_i. The eigenvalues ζ_i are all real numbers iff A is self-adjoint.

This means that there is a sense in which all the normal operators can be viewed as "observables", if by that we mean, "operators with a PVM" that encode the quantitative outcomes and expectation values of a statistical experiment using classical probabilities. For a more general discussion of extending "observables" beyond self-adjoint operators, see Roberts (2018).

3. Cornerstone 3: Wigner's theorem and Stone's theorem

3.1. **Symmetries and Wigner's theorem.** We now need to have a grip on the notion of a "symmetry" in quantum theory. This is typically given by Wigner's theorem,[9], which we state it here in a more general form due to Uhlhorn (1963); see also Varadarajan (2007, Thm. 4.29). It makes use of the concept of a *Hilbert space ray*, or equivalence class of vectors related by a phase factor, $\Psi = \{e^{i\theta}\psi \mid \psi \in \mathcal{H}, \theta \in \mathbb{R}\}$.

[7] An operator A on a (possibly infinite-dimensional) Hilbert space $\mathcal{H}$ is called *symmetric* if $A\psi = A^*\psi$ for all ψ in the common domain of A and A^*; it is called *self-adjoint* if in addition, A and A^* have the same domain. In finite dimensions the latter condition is automatically satisfied, but not with infinite dimensions.

[8] Every linear operator A can be written in the form $A = B + iC$ by setting $B = (A^* + A)/2$ and $B = i(A^* - A)/2$. Exercise: check that $[B, C] = 0$ if and only if $AA^* = A^*A$.

[9] Wigner (1931) originally gave a heuristic proof of this theorem. Work making it precise and generalising it is an interesting discussion in philosophy of physics in its own right; for a recent survey, see Chevalier (2007).

Rays admit an inner product given by,

$$\langle \Psi, \Phi \rangle := |\langle \psi, \phi \rangle|^2 \text{ where } \psi \in \Psi, \phi \in \Phi,$$

and if two rays are *orthogonal* ($\langle \Psi, \Phi \rangle = 0$) then we write $\Psi \perp \Phi$. As propositions, orthogonal rays are "mutually exclusive" outcomes. Suppose we take "symmetry" to be a transformation that preserves orthogonal outcomes. This provides a strong constraint on what symmetries in quantum mechanics can be:

Fact 5 (Wigner-Uhlhorn theorem)**.** *Let* $\boldsymbol{S}$ *be a bijection on the rays* Ψ *of a Hilbert space* $\mathcal{H}$ *of countable dimension* > 2*. Suppose* $\langle \Psi, \Phi \rangle = 0$ *if and only if* $\langle S\Psi, S\Phi \rangle = 0$*. Then for all rays* Ψ, Φ*,*

$$\langle S\Psi, S\Phi \rangle = \langle \Psi, \Phi \rangle.$$

Moreover, there exists a unique (up to a constant) $S : \mathcal{H} \to H$ *that implements* $\boldsymbol{S}$ *on* $\mathcal{H}$ *in that* $\psi \in \Psi$ *if and only if* $S\psi \in \mathbf{S}\Psi$*, where* S *is either unitary or antiunitary and satisfies* $|\langle S\psi, S\phi \rangle| = |\langle \psi, \phi \rangle|$ *for all* $\psi, \phi \in \mathcal{H}$*.*

As a reminder: a *unitary* operator U is one that satisfies:

(1) (linearity) $U(a\psi + b\phi) = aU\psi + bU\phi$ for all $a, b \in \mathbb{R}$ and all $\psi, \phi \in \mathcal{H}$;
(2) (unitary property) $U^*U = UU^* = I$.

These properties are jointly equivalent[10] to the statement that $\langle U\psi, U\phi \rangle = \langle \psi, \phi \rangle$ for all $\psi, \phi \in \mathcal{H}$. An *antiunitary* operator T is one that satisfies:

(1) (anti-linearity) $T(a\psi + b\phi) = a^*T\psi + b^*T\phi$ for all $a, b \in \mathbb{R}$ and all $\psi, \phi \in \mathcal{H}$;
(2) (unitary property) $T^*T = TT^* = I$.

These properties are jointly equivalent to the statement that $\langle U\psi, U\phi \rangle = \langle \psi, \phi \rangle^*$ for all $\psi, \phi \in \mathcal{H}$.

3.2. **The Noether correspondence and Stone's theorem.** Classical mechanics in Lagrangian and Hamiltonian form both have a Noether-like correspondence, which says something like:

> (symmetry-observable correspondence) *Every observable admits a continuous group of symmetries along which it is conserved, and vice versa: every continuous group of symmetries generates an observable that it conserves.*

Such a correspondence exists for quantum theory as well. Suppose we view "observables" as self-adjoint operators, and "symmetries" as unitary or antiunitary operators; the only continuous groups of symmetries are unitary, so we will restrict attention to unitary operators.

Stone's theorem makes use of the concept of a 1-parameter representation of the real numbers under addition, namely, a set of Hilbert space operators $\{U_s \mid s \in \mathbb{R}\}$ such that $U_{s+t} = U_sU_t$. We can now state (Blank et al.; 2008, Thm. 5.9.2):

[10]Exercise: Check this. Hint: Use the linearity and skew-linearity of the Hilbert space inner product.

Fact 6 (Stone's theorem)**.** *For any strongly continuous one-parameter unitary representation of* $(\mathbb{R}, +)$*, written* $s \mapsto U_s$*, there exists a unique self-adjoint operator* A *such that* $U_s = e^{-isA}$ *for all* $s \in \mathbb{R}$*. Conversely, for each self-adjoint operator the association* $s \mapsto e^{-isA} := U_s$ *defines a strongly continuous one-parameter unitary representation.*

This result in particular allows us to motivate the dynamics of quantum theory, using the assumptions that,

- Change occurs continuously in time; and
- Experiments are time-translation invariant.

Interpreting these assumptions as the antecedent of Stone's theorem imply that time evolution is represented by a unitary group $t \mapsto U_t = e^{-itH}$ for some self-adjoint operator H. Writing $\psi(t) = e^{-itH}\psi$ for some initial vector ψ and differentiating both sides then gives the famous Schrödinger equation,

$$i\tfrac{d}{dt}\psi(t) = H\psi(t).$$

That is, the Schrödinger equation is the "differential form" off unitary time evolution.

References

Birkhoff, G. and Neumann, J. V. (1936). The logic of quantum mechanics, *The Annals of Mathematics* **37**(4): 823–843.
URL: *http://www.jstor.org/stable/1968621*

Blank, J., Exner, P. and Havlíček, M. (2008). *Hilbert Space Operators in Quantum Physics*, 2nd edn, Springer Science and Business Media B.V.

Chevalier, G. (2007). Wigner's theorem and its generalizations, *in* K. Engesser, D. M. Gabbay and D. Lehmann (eds), *Handbook of Quantum Logic and Quantum Structures: Quantum Structures*, Amsterdam, Oxford: Elsevier B.V.

Conway, J. B. (1990). *A Course in Functional Analysis*, Vol. 96 of *Graduate Texts in Mathematics*, 2nd edn, Springer-Verlag New York, Inc.

Jauch, J. M. (1968). *Foundations of Quantum Mechanics*, Addison-Wesley Series in Advanced Physics, Reading, MA, Menlo Park, CA, London, Don Mills, ON: Addison-Wesley Publishing Company, Inc.

Roberts, B. W. (2018). Observables, disassembled, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **63**: 150–162. Preprint: `http://philsci-archive.pitt.edu/14449/`.

Uhlhorn, U. (1963). Representation of symmetry transformations in quantum mechanics, *Arkiv för Fysik* **23**: 307–340.

Varadarajan, V. S. (2007). *Geometry of Quantum Theory*, 2nd edn, New York: Springer Science and Business Media, LLC.

Wigner, E. P. (1931). *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*, New York: Academic Press (1959).

Email address: b.w.roberts@lse.ac.uk