we introduce quantum mechanics, we will relentlessly capitalize on parallels with probability theory, but also take care to point out where such parallels break down.

2. Quantum theory in finite dimensions

We begin with a very brief history of quantum theory. Circa 1900 Max Planck studied blackbody radiation, and solved an inadequacy in the extant equations by stipulating that energy is quantized in units of his eponymous constant. Then in 1905, Einstein suggests that light itself is quantized as "photons", providing an explanation for the photoelectric effect. In the ensuing decade, Bohr makes a first pass at quantum theory (the so-called 'old' quantum theory), and correctly predicts the spectral lines of hydrogen. This first pass at quantum theory only goes so far, and a second pass is made in the 1920's. In 1924, de Broglie postulates that a particle with momentum p has 'wavelength' $\lambda = h/p$, which is soon confirmed by electron diffraction experiments. Thereafter, Heisenberg, Born, and Jordan developed matrix mechanics in 1925 (although they did not yet understand the connection to de Broglie). In 1926, Schrödinger leveraged de Broglie's insight to develop wave mechanics, and that same year showed the equivalence with matrix mechanics. That year as well, Born gave a 'probabilistic' interpretation of quantum mechanics which clarified its connections to measurable quantities in experiments. In 1927, Heisenberg wrote down his famous uncertainty principle. Most of the abstract mathematical foundations of quantum mechanics were consolidated by Dirac and von Neumann in the early 1930's, and Einstein-Podolsky-Rosen as well as Schrödinger highlighted the importance of entanglement in 1935. The year after in 1936, Birkoff and von Neumann investigated how quantum mechanics leads to a new form of logical reasoning that goes beyond classical Boolean logic; in hindsight this may be regarded as the first hint of the possibility of quantum computing (although it was not understood as such at the time).

Having completed our brief historical diegesis, we now turn to presenting the axioms of quantum mechanics. There are various ways of 'motivating' the axioms of quantum mechanics, although at some level they were *guessed* by very clever people and experimentally confirmed by very clever people (sometimes in the opposite order). We will, however, give some intuition. But first, a word of caution. When someone asks for a motivation for quantum mechanics in terms of classical mechanics, this is philosophically backwards; it would be like asking for a derivation of special relativity starting from Newton's equations. Indeed, just as special relativity reduces to Newtonian physics in a certain regime of validity, so too does quantum mechanics reduce to classical mechanics in a certain regime of validity. Nonetheless, we will proceed with an idiosyncratic way of 'guessing' some of the axioms of quantum mechanics starting from classical intuitions.

2.1. Mechanics on ℓ^p spaces: from classical to quantum

Let us begin by contemplating the salient mathematical structures undergirding the dynamics of probability distributions discussed above. For this, it is useful to have the following definition:

Definition 12 (Normed vector space)**.** *Let V be a vector space over a field K; we will consider either $V = \mathbb{R}^N$ (with $K = \mathbb{R}$), or $V = \mathbb{C}^N$ (with $K = \mathbb{C}$). A **normed***

vector space *is a pair* $(V, \|\cdot\|)$ *where* $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ *is the* ***norm*** *which satisfies the following three properties:*

(1) (Positive definiteness) $\|\vec{v}\| = 0$ *if and only if* $\vec{v}$ *is the zero vector.*
(2) (Absolute homogeneity) $\|a\,\vec{v}\| = |a|\|\vec{v}\|$ *for any* $a \in K$ *and any* $\vec{v} \in V$*.*
(3) (Triangle inequality) $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$ *for any* $\vec{v}, \vec{w} \in V$*.*

Then we can define a very useful class of norms as follows:

Definition 13 (ℓ^p norms)**.** *The* ℓ^p ***norm****, defined over* $\mathbb{R}^N$ *or* $\mathbb{C}^N$ *for* $p \geq 1$*, is*

$$\|\vec{v}\|_p := \left(\sum_{j=1}^{N} |v_j|^p \right)^{\frac{1}{p}}. \tag{5}$$

One can show that (5) is indeed a norm in the sense of Definition 12 above. (It is immediate to verify positive definiteness and absolute homogeneity; verifying the triangle inequality involves a more delicate proof leveraging Hölder's inequality.)

A special case of the ℓ^p norm is when $p = 1$, giving $\|\vec{v}\|_1 = \sum_{j=1}^N |v_j|$. Then when $\vec{p}$ describes a probability distribution, the normalization of probability distributions is equivalent to the condition $\|\vec{p}\|_1 = 1$. Then our characterization of Markov matrices can be equivalently phrased as follows: M is a Markov matrix if and only if

$$\|M \cdot \vec{p}\|_1 = \|\vec{p}\|_1$$

for all $\vec{p}$ describing probability distributions. In fact, using absolutely homogeneity, we also have the slightly weaker statement that M is a Markov matrix if and only if $\|M \cdot \vec{v}\|_1 = \|\vec{v}\|_1$ where all entries of $\vec{v}$ have the same sign. But then we might ask: what are the matrices A such that $\|A \cdot \vec{v}\|_1 = \|\vec{v}\|_1$ for all $\vec{v} \in \mathbb{R}^N$? Interestingly, such matrices A, called ℓ^1-isometries, are highly restricted:

Theorem 14 (ℓ^1-isometries)**.** *Let* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ *and* $A \in \mathbb{K}^{N \times N}$*. The following are equivalent:*

(1) $\|A \cdot \vec{v}\|_1 = \|\vec{v}\|_1$ *for all* $\vec{v} \in \mathbb{K}^N$*.*
(2) $A = P \cdot \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_N)$ *where* P *is a permutation matrix and* $|\varepsilon_j| = 1$ *for all* j *(so* $\varepsilon_j = \pm 1$ *if* $\mathbb{K} = \mathbb{R}$*).*

In the proof below, for a vector $\vec{v} = (v_1, \ldots, v_N) \in \mathbb{K}^N$ we write

$$\mathrm{supp}(\vec{v}) := \{\, k \in \{1, \ldots, N\} : v_k \neq 0 \,\}$$

for its *support*. We say two vectors have *disjoint supports* if their supports are disjoint sets.

Proof. Write $\vec{a}_j := A \cdot \vec{e}_j$ for the jth column of A. Then $\|\vec{a}_j\|_1 = \|A \cdot \vec{e}_j\|_1 = \|\vec{e}_j\|_1 = 1$.

Fix $i \neq j$. In the real case,

$$\|\vec{a}_i \pm \vec{a}_j\|_1 = \|A \cdot (\vec{e}_i \pm \vec{e}_j)\|_1 = \|\vec{e}_i \pm \vec{e}_j\|_1 = 2.$$

By the triangle inequality we always have $\|\vec{a}_i \pm \vec{a}_j\|_1 \leq \|\vec{a}_i\|_1 + \|\vec{a}_j\|_1 = 2$; equality of sums forces equality *coordinate-wise*. Thus for every coordinate k,

$$|a_i(k) \pm a_j(k)| = |a_i(k)| + |a_j(k)|.$$

For reals, the '+' equality enforces same sign (or a zero), and the '−' equality enforces opposite sign (or a zero); both can hold only if $a_i(k)a_j(k) = 0$. Hence $\mathrm{supp}(\vec{a}_i) \cap \mathrm{supp}(\vec{a}_j) = \varnothing$.

In the complex case, use

$$\|\vec{a}_i + \vec{a}_j\|_1 = \|\vec{a}_i + i\,\vec{a}_j\|_1 = 2.$$

Again equality is coordinate-wise, so with $z = a_i(k)$ and $w = a_j(k)$,

$$|z + w| = |z| + |w|, \qquad |z + i\,w| = |z| + |w|\,.$$

Each equality in $\mathbb{C}$ holds if and only if the summands share an argument; the first says z and w are collinear, the second says z and $i\,w$ are collinear. This is impossible unless $z = 0$ or $w = 0$. Thus the supports of distinct columns are disjoint in the complex case as well.

We now have N nonempty, pairwise-disjoint subsets $S_j := \mathrm{supp}(\vec{a}_j) \subseteq \{1, \ldots, N\}$. Therefore

$$N \;\le\; \sum_{j=1}^{N} |S_j| = \Big|\bigcup_{j=1}^{N} S_j\Big| \;\le\; N,$$

so $|S_j| = 1$ for all j. Hence $\vec{a}_j = \varepsilon_j\,\vec{e}_{\sigma(j)}$ for some permutation σ and some $\varepsilon_j \neq 0$. From $\|\vec{a}_j\|_1 = |\varepsilon_j| = 1$ we get $|\varepsilon_j| = 1$, and writing P for the permutation matrix of σ gives $A = P \cdot \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_N)$. The converse is immediate. □

Remark 15. *Equivalently, the ℓ^1-isometries are the* ***signed permutation matrices*** *when $\mathbb{K} = \mathbb{R}$ and the* ***monomial matrices*** *with unimodular entries (i.e. their absolute value equals one) when $\mathbb{K} = \mathbb{C}$. If one further assumes $A_{ij} \ge 0$, then necessarily $\varepsilon_j = 1$ for all j, so A is a permutation matrix.*

The upshot of Theorem 14 is that the only linear maps that preserve the ℓ^1 norm on all of $\mathbb{R}^N$ (or $\mathbb{C}^N$) are signed-permutation (or monomial) matrices. Thus, if we insist on global ℓ^1–isometries, the dynamics amount only to relabeling coordinates and multiplying by signs (or phases). A nontrivial theory appears when we restrict attention to the positive cone and, in particular, to the probability simplex Δ_N: requiring a linear map M to send probability vectors to probability vectors yields precisely the column–stochastic (Markov) matrices introduced above. Moreover, Theorem 14 generalizes as follows:

Theorem 16 (ℓ^p-isometries for $p \neq 2$)**.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and $A \in \mathbb{K}^{N\times N}$. Then for $p \ge 1$ and $p \neq 2$, the following are equivalent:*

(1) $\|A \cdot \vec{v}\|_p = \|\vec{v}\|_p$ for all $\vec{v} \in \mathbb{K}^N$.

(2) $A = P \cdot \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_N)$ where P is a permutation matrix and $|\varepsilon_j| = 1$ for all j (so $\varepsilon_j = \pm 1$ if $\mathbb{K} = \mathbb{R}$).

A proof can be found in [**Aar04**] (although the original proof goes back to at least Banach). The theorem above shows that for $p \neq 2$ the only linear $\|\cdot\|_p$–isometries are monomial matrices, so there is no norm-preserving linear dynamics that mixes coordinates beyond permutations (and multiplicative sign or phase factors). The case $p = 1$ is special only in that, after restricting to the positive cone, we can relax from "isometry on all vectors" to the weaker requirement "maps the probability simplex to itself"; this yields the rich class of Markov matrices. For $p > 1$ and not equal to 2, no analogous stochastic family exists. By contrast, when $p = 2$ the isometries form a continuous group providing genuinely nontrivial linear dynamics.

We have already explicated how the $p = 1$ case corresponds to classical mechanics; we will see that the $p = 2$ case corresponds to quantum mechanics.

First let us give a structure theorem for the ℓ^2-isometries. We start with the following definition.

Definition 17 (Orthogonal and unitary groups)**.** *A matrix $R \in \mathbb{R}^{N\times N}$ is an **orthogonal matrix** if and only if it satisfies $R^T R = RR^T = \mathbb{1}$. This set of matrices is closed under multiplication and inverses, and forms the **orthogonal group** $O(N)$. Similarly, a matrix $U \in \mathbb{C}^{N\times N}$ is a **unitary matrix** if and only if it satisfies $U^\dagger U = UU^\dagger = \mathbb{1}$. This set of matrices is closed under multiplication and inverses, and forms the **unitary group** $U(N)$.*

Then our structure theorem for ℓ^2-isometries is as follows.

Theorem 18 (ℓ^2-isometries)**.** *Let $R \in \mathbb{R}^{N\times N}$. The following are equivalent.*

(1) $\|R \cdot \vec{v}\|_2 = \|\vec{v}\|_2$ for all $\vec{v} \in \mathbb{R}^N$.
(2) $R \in O(N)$.

Similarly, let $U \in \mathbb{C}^{N\times N}$. The following are equivalent.

(1) $\|U \cdot \vec{v}\|_2 = \|\vec{v}\|_2$ for all $\vec{v} \in \mathbb{C}^N$.
(2) $U \in U(N)$.

We defer the proof until later, when additional mathematical tools will allow us to present it more simply.

In the same way that

$$\vec{p}\,' = M_k \cdots M_2 \cdot M_1 \cdot \vec{p}$$

for the M_i being Markov matrices constitutes ℓ^1-preserving dynamics on $\Delta_N \subset \mathbb{R}^N$, then e.g.

$$\vec{\Psi}' = U_k \cdots U_2 \cdot U_1 \cdot \vec{\Psi} \tag{6}$$

for $\vec{\Psi}, \vec{\Psi}' \in \mathbb{C}^N$ and the U_i being unitary matrices constitutes ℓ^2-preserving dynamics on $\mathbb{C}^N$. Just as probability distributions $\vec{p} \in \Delta_N \subset \mathbb{R}^N$ play a starring role in classical mechanics, the **wavefunction** plays a starring role in quantum mechanics. In its simplest form, a wavefunction is a vector $\vec{\Psi} \in \mathbb{C}^N$. (The fact that $\vec{\Psi}$ lives on $\mathbb{C}^N$ as opposed to $\mathbb{R}^N$ is an empirical fact with measurable consequences.) In particular, the wavefunction will provide a description of the 'state' of a quantum system, and so often the words 'wavefunction' and 'state' are used interchangeably.

Quantum mechanics is essentially the study of dynamics of the form (6) on $\mathbb{C}^N$, along with additional physical input that relates that dynamics to observable reality. Other physical inputs can constrain the form of the unitaries which are used. Before delving into these 'physical' considerations below, it is first worth explicating a bit more of the mathematical structure of ℓ^2 spaces, since they will be our stomping grounds for the entirety of this book.[2]

So far we have introduced the structure of an ℓ^2 norm on $\mathbb{C}^N$, in Definitions 12 and 13 (taking $p = 2$ in the latter). A nice feature of the ℓ^2 norm is that it gives us a very nice additional structure on $\mathbb{C}^N$, namely an inner product space. We define inner product spaces below, and then explain how the ℓ^2 norm allows us to define a canonical inner product space.

[2]They are also, more generally, the stomping grounds for our physical reality.

Definition 19 (Inner product and inner product space)**.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be a vector space over $\mathbb{K}$. An* ***inner product*** *on V is a map*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}$$

such that for all $u, v, w \in V$ and $a, b \in \mathbb{K}$:

(1) (Conjugate symmetry) $\langle v, w\rangle = \overline{\langle w, v\rangle}$.
(2) (Sesquilinearity) $\langle u, av + bw\rangle = a\langle u, v\rangle + b\langle u, w\rangle$ *and* $\langle au + bv, w\rangle = \overline{a}\langle u, w\rangle + \overline{b}\langle v, w\rangle$. *Equivalently, the inner product is linear in its* second *argument and conjugate–linear in its first.*[3]
(3) (Positive definiteness) $\langle v, v\rangle \geq 0$, *with equality if and only if $v = 0$.*

A pair $(V, \langle \cdot, \cdot \rangle)$ is called an ***inner product space****. The inner product induces a norm by*

$$\|v\| := \sqrt{\langle v, v\rangle}\,.$$

To fully bring you into the fold, we introduce a slightly more refined notion of inner product spaces due to Hilbert.

Definition 20 (Hilbert space)**.** *A (complex)* ***Hilbert space*** *is a complex inner product space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ that is complete*[4] *with respect to the induced norm $\|v\| = \sqrt{\langle v, v\rangle}$.*

Remark 21 (Finite-dimensional case and notation)**.** *When $\dim \mathcal{H} < \infty$, completeness is automatic, so every complex inner product space is a Hilbert space. In this book we work exclusively with finite-dimensional Hilbert spaces, typically written $\mathcal{H} \simeq \mathbb{C}^N$ equipped with the ℓ^2 inner product. We will often write $\vec{\Psi} \in \mathcal{H}$ for a state vector ("wavefunction"), and linear maps on $\mathcal{H}$ are represented by matrices; those that preserve the inner product are precisely the unitary operators $U : \mathcal{H} \to \mathcal{H}$.*

As such, an inner product space can be thought of as a normed space, with additional structure. Below we explain how the ℓ^2 norm is induced by an inner product.

Definition 22 (ℓ^2 inner product)**.** *On $\mathbb{C}^N$ we take the standard inner product to be the* ***ℓ^2 inner product***

$$\langle v, w\rangle := v^\dagger w = \sum_{j=1}^{N} \overline{v}_j w_j,$$

which on $\mathbb{R}^N$ reduces to $v^T w$. The induced norm is $\|v\| = \sqrt{\langle v, v\rangle} = \left(\sum_{j=1}^N |v_j|^2\right)^{1/2} = \|\vec{v}\|_2$, which is precisely the ℓ^2 norm.

A useful notion is (Hermitian) conjugation, which we define below.

Definition 23 (Conjugation and Hermitian adjoint)**.** *For a complex number $a \in \mathbb{C}$, its complex conjugate is a^*. For a vector $\vec{v} \in \mathbb{C}^N$, write $\vec{v}^*$ for entrywise conjugation and define the* ***conjugate transpose*** *(or* ***Hermitian conjugate****) by*

$$\vec{v}^\dagger := (\vec{v}^*)^T\,.$$

[3]This is the convention commonly used in physics. Over $\mathbb{R}$ it reduces to bilinearity.

[4]"Complete" means that every Cauchy sequence in $\mathcal{H}$ (with respect to the metric $d(u, v) = \|u - v\|$ induced by the inner product) converges to a limit *in* $\mathcal{H}$: for all $\varepsilon > 0$ there exists N such that $m, n \geq N$ implies $\|x_m - x_n\| < \varepsilon$, and there is $x \in \mathcal{H}$ with $\|x_n - x\| \to 0$. Intuitively, there are no 'holes' in the space.

For a matrix $A \in \mathbb{C}^{M\times N}$*, write* A^* *for entrywise conjugation and define its* ***Hermitian adjoint*** *(conjugate transpose) by*

$$A^\dagger := (A^*)^T \in \mathbb{C}^{N\times M}.$$

Equivalently, using the ℓ^2 *inner product* $\langle u, v\rangle = u^\dagger v = \sum_{j=1}^N u_j^* v_j$*, we have that* $A^\dagger$ *is the unique linear map satisfying*

$$\langle x, Ay\rangle = \langle A^\dagger x, y\rangle \quad \text{for all } x \in \mathbb{C}^M,\ y \in \mathbb{C}^N.$$

The adjoint obeys, for all compatible A, B *and scalars* $\alpha, \beta \in \mathbb{C}$*,*

$$(AB)^\dagger = B^\dagger A^\dagger, \qquad (\alpha A + \beta B)^\dagger = \alpha^* A^\dagger + \beta^* B^\dagger, \quad (A^\dagger)^\dagger = A.$$

Over $\mathbb{R}$*, complex conjugation is trivial (*$a^* = a$*), so* $A^\dagger = A^T$*. Additionally, a matrix* H *is* ***Hermitian*** *(or* ***self–adjoint****) if* $H^\dagger = H$*.*

Having defined the ℓ^2 inner product as well as the Hermitian adjoint, we can rephrase Theorem 18 as:

Theorem 24 (ℓ^2-isometries, reprise)**.** *Let* $R \in \mathbb{R}^{N\times N}$*. The following are equivalent.*

(1) $\langle R\vec{v}, R\vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle$ *for all* $\vec{v} \in \mathbb{R}^N$*.*
(2) $R \in O(N)$*.*

Similarly, let $U \in \mathbb{C}^{N\times N}$*. The following are equivalent.*

(1) $\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle$ *for all* $\vec{v} \in \mathbb{C}^N$*.*
(2) $U \in U(N)$*.*

With our inner product definitions at hand, we are now equipped to provide a simple proof of Theorem 24 and thus Theorem 18.

Proof. We give the argument for $\mathbb{C}^N$; the real case is analogous with † replaced by ^{T} and i replaced by ± 1.

Assume (1): $\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle$ for all $\vec{v} \in \mathbb{C}^N$. Write

$$\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, U^\dagger U\, \vec{v}\rangle,$$

so for every $\vec{v}$,

$$\langle \vec{v},\, (U^\dagger U - \mathbb{1})\, \vec{v}\rangle = 0.$$

Set $H := U^\dagger U - \mathbb{1}$. Then $\langle \vec{v}, H\vec{v}\rangle = 0$ for all $\vec{v}$. For arbitrary $\vec{x}, \vec{y}$ we compute (using conjugate–linearity in the first argument and linearity in the second):

$$\begin{aligned} 0 &= \langle \vec{x} + \vec{y},\, H(\vec{x} + \vec{y})\rangle = \langle \vec{x}, H\vec{x}\rangle + \langle \vec{x}, H\vec{y}\rangle + \langle \vec{y}, H\vec{x}\rangle + \langle \vec{y}, H\vec{y}\rangle \\ &= \langle \vec{x}, H\vec{y}\rangle + \langle \vec{y}, H\vec{x}\rangle, \\ 0 &= \langle \vec{x} + i\vec{y},\, H(\vec{x} + i\vec{y})\rangle = \langle \vec{x}, H\vec{x}\rangle + i\langle \vec{x}, H\vec{y}\rangle - i\langle \vec{y}, H\vec{x}\rangle + \langle \vec{y}, H\vec{y}\rangle \\ &= i\langle \vec{x}, H\vec{y}\rangle - i\langle \vec{y}, H\vec{x}\rangle. \end{aligned}$$

Solving these two equations gives $\langle \vec{x}, H\vec{y}\rangle = \langle \vec{y}, H\vec{x}\rangle = 0$ for all $\vec{x}, \vec{y}$. Fixing $\vec{y}$ and taking $\vec{x} = H\vec{y}$ yields $\|H\vec{y}\|^2 = 0$, so $H\vec{y} = 0$ for all $\vec{y}$ and hence $H = 0$. Therefore $U^\dagger U = \mathbb{1}$. In particular, the columns of U are orthonormal, so U is invertible and $U^{-1} = U^\dagger$; hence also $UU^\dagger = \mathbb{1}$ and $U \in U(N)$, establishing (2).

Conversely, if $U \in U(N)$ then $U^\dagger U = \mathbb{1}$, and for all $\vec{v}$,

$$\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, U^\dagger U\, \vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle\,,$$

which is (1). This completes the proof. □

Let us pause to summarize what we have done so far in this Subsection. First, we recognized that dynamics on (finite) probability distributions is dynamics that preserves ℓ_1. We then contemplated what dynamics would look like that preserves ℓ_p for $p > 1$, and found that the only interesting option is $p = 2$, for which unitary dynamics does the job. We then explained that the ℓ_2 is produced by a natural inner product, which also interfaces nicely with unitary dynamics. Below, we will show how ℓ_2-preserving dynamics is essentially (finite-dimensional) quantum mechanics, along with some additional mathematical baggage which relates the dynamics to observable measurements. Then let us commence below with the axioms of quantum mechanics.

2.2. The axioms of quantum mechanics

Quantum mechanics was presented essentially its contemporary form by Paul Dirac in 1930 [**Dir81**] and placed on a rigorous Hilbert space footing by John von Neumann in 1932 [**VN18**]. The reader might be surprised to discover that Dirac's book [**Dir81**] remains foundational for quantum-mechanics courses nearly a century later.

2.2.1. *Bra-ket notation*

Before giving the axioms, we introduce Dirac's famous **bra-ket notation**, much beloved by physicists (and sometimes unfairly despised by mathematicians). Consider the $\mathbb{C}^N$, viewed as a Hilbert space with ℓ^2 inner product. In the future, we will simply say "consider the Hilbert space $\mathcal{H} \simeq \mathbb{C}^N$". Recall that if $\vec{\psi}, \vec{\phi} \in \mathcal{H}$ then their inner product is

$$\langle \vec{\psi}, \vec{\phi} \rangle = \sum_{j=1}^{N} \psi_j^* \phi_j = \vec{\psi}^\dagger \cdot \vec{\phi}.$$

The far right-hand side demonstrates that we can think of the inner product as a bilinear map from $\mathcal{H}^* \otimes \mathcal{H} \to \mathbb{C}$, where $\mathcal{H}^*$ is the space of row vectors. There is a canonical isomorphism from $\mathcal{H}$ to $\mathcal{H}^*$ given by Hermitian conjugation. This is all just a fancy way of saying the following: to take the inner product $\langle \vec{\psi}, \vec{\phi} \rangle$ of $\vec{\psi}$ and $\vec{\phi}$, we just take the Hermitian conjugate of $\vec{\psi}$ and dot that with $\vec{\phi}$.

The far left-hand side of 2.2.1 takes the notational form of a 'bracket'. Dirac suggests that we enclose vectors in $\mathcal{H}$ by $|\cdot\rangle$, so that instead of writing $\vec{\phi}$ we write $|\phi\rangle$. Such an object is called a 'ket'. In similar spirit, a column vector $\vec{\phi}^\dagger \in \mathcal{H}^*$ is enclosed by $\langle \cdot|$, so that instead of writing $\vec{\psi}^\dagger$ we write $\langle\psi|$. Such an object is called a 'bra'. Then bras and kets are related via Hermitian conjugation, namely

$$|\psi\rangle^\dagger = \langle\psi| \,.$$

Finally, we can put together bras and kets to form

$$\langle\psi|\phi\rangle := \langle \vec{\psi}, \vec{\phi} \rangle = \sum_{j=1}^{N} \psi_j^* \phi_j = \vec{\psi}^\dagger \cdot \vec{\phi} \,,$$

which is a...(drum roll please) 'bra-ket'! Get it?[5]

[5]Famously, Dirac was not known for his sense of humor.

Besides being somewhat whimsical, Dirac's bra-ket notation is in fact extremely useful. The reason is not so much mathematical, but rather visual. As you yourself will experience, bra-ket notation is visually suggestive of how to organize and manipulate certain equations (especially compared with arrows and daggers), and eases the mind towards simplifying complicated expressions in multi-linear algebra. That is, Dirac found a notation which resonates with the structure of our minds.

Let us develop Dirac's notation a bit further. In addition to forming 'inner products' $\langle\psi|\phi\rangle = \vec{\psi}^\dagger \cdot \vec{\phi}$, we can also form 'outer products' $|\phi\rangle\langle\psi| = \vec{\phi} \cdot \vec{\psi}^\dagger$. Here $|\phi\rangle\langle\psi|$ is evidently a rank 1, $N \times N$ matrix. Then the trace of this matrix is evidently

$$\mathrm{tr}(|\phi\rangle\langle\psi|) = \langle\psi|\phi\rangle\,.$$

Since Hermitian conjugation for a scalar is the same as complex conjugation, we have the useful identity

$$(\langle\psi|\phi\rangle)^\dagger = (\langle\psi|\phi\rangle)^* = \langle\phi|\psi\rangle\,,$$

where we observe that the ψ and ϕ have switched sides.

It is useful to show a few examples to get you fully acquainted with bra-ket notation. Consider the standard orthonormal basis $\{\vec{e}_i\}_{i=1}^N$ of $\mathbb{C}^N$, which we denote by $\{|i\rangle\}_{i=1}^N$ in our new notation. The orthonormality of the basis elements can be expressed as

$$\langle i|j\rangle = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}\,,$$

where δ_{ij} is called the **Kronecker delta**. Recalling that the identity matrix is $\mathbb{1} = \sum_{i=1}^N \vec{e}_i \cdot \vec{e}_i^{\,T}$, in bra-ket notation we have

$$\mathbb{1} = \sum_{i=1}^N |i\rangle\langle i|\,.$$

Then given a state $|\psi\rangle$, we have

$$|\psi\rangle = \mathbb{1}|\psi\rangle = \left(\sum_{i=1}^N |i\rangle\langle i|\right)|\psi\rangle = \sum_{i=1}^N |i\rangle \underbrace{\langle i|\psi\rangle}_{=:\psi_i} = \sum_{i=1}^N \psi_i|i\rangle\,, \tag{7}$$

where ψ_i are the coefficients of $|\psi\rangle$ in the $|i\rangle$-basis. (Note also that $(\langle i|\psi\rangle)^\dagger = \langle i|\psi\rangle)^* = \langle\psi|i\rangle = \psi_i^*$, and so the coefficients of $\langle\psi|$ in the $\langle i|$-covector basis are ψ_i^*.) Similarly, for a matrix M, we have

$$M = \mathbb{1}\cdot M \cdot \mathbb{1} = \left(\sum_{i=1}^N |i\rangle\langle i|\right) M \left(\sum_{j=1}^N |j\rangle\langle j|\right) = \sum_{i,j=1}^N |i\rangle \underbrace{\langle i|M|j\rangle}_{=:M_{ij}}\langle j| = \sum_{i,j=1}^N M_{ij}\,|i\rangle\langle j|\,, \tag{8}$$

where M_{ij} are the matrix elements of M in the $|i\rangle$-basis. As a check of our notation, let us compute $M|\psi\rangle$ using the far-right hand sides of both (7) and (8):

$$M|\psi\rangle = \left(\sum_{i,j=1}^N M_{ij}\,|i\rangle\langle j|\right)\sum_{k=1}^N \psi_k|k\rangle = \sum_{i,j,k=1}^N M_{ij}\psi_k|i\rangle \underbrace{\langle j|k\rangle}_{=\delta_{jk}} = \sum_{i=1}^N \left(\sum_{j=1}^N M_{ij}\psi_j\right)|i\rangle\,.$$

So we see that the coefficients of $M|\psi\rangle$ in the $|i\rangle$-basis are $\sum_{j=1}^N M_{ij}\psi_j$, exactly as expected using the standard rules of matrix multiplication.

For our final flourish, we present the **spectral theorem** in finite dimensions in bra-ket notation. The spectral theorem will play a crucial role in the formulation of quantum mechanics.

Theorem 25 (Spectral theorem for normal matrices in finite dimensions)**.** *Let $A : \mathcal{H} \to \mathcal{H}$ be a linear operator on a finite dimensional complex Hilbert space $\mathcal{H} \simeq \mathbb{C}^N$. The following are equivalent*

(1) A is normal, meaning $A^\dagger A = AA^\dagger$.
(2) There exists an orthonormal basis of eigenstates $|v_1\rangle, \ldots, |v_N\rangle$ and complex numbers $\lambda_1, \ldots, \lambda_N$ such that $A = \sum_{j=1}^N \lambda_j\, |v_j\rangle\langle v_j|$ Equivalently, if U is the unitary with columns $|v_j\rangle$ then $U^\dagger AU = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$.

We will break up the proof into a few lemmas:

Lemma 26. *Let $A : \mathcal{H} \to \mathcal{H}$ be a normal matrix for $\mathcal{H} \simeq \mathbb{C}^N$. Then A has at least one eigenvector $|v\rangle$. Moreover, if $A|v\rangle = \lambda|v\rangle$, then $A^\dagger|v\rangle = \lambda^*|v\rangle$.*

Proof. By the fundamental theorem of algebra the characteristic polynomial $p_A(\lambda) = \det(A - \lambda\mathbb{1})$ has a complex root. If λ is such a root, then $A - \lambda\mathbb{1}$ has a non-trivial nullspace, meaning that A has an eigenvalue λ and at least one nonzero eigenstate $|v\rangle$ with $A|v\rangle = \lambda|v\rangle$. Without loss of generality we take $|v\rangle$ to be normalized so that $\langle v|v\rangle = 1$. Now notice that

$$\underbrace{\langle v|A^\dagger}_{=\lambda^*\langle v|}\,|v\rangle = \lambda^*\,. \tag{9}$$

Recall that the Cauchy-Schwarz inequality $|\langle\psi|\phi\rangle| \leq \sqrt{\langle\psi|\psi\rangle}\sqrt{\langle\phi|\phi\rangle}$ achieves equality only when $|\psi\rangle$ is proportional to $|\phi\rangle$. Assuming A is normal, we have

$$\begin{aligned}|\lambda| &= |\langle v|A^\dagger|v\rangle| \\ &\leq \sqrt{\langle v|v\rangle}\sqrt{\langle v|A^\dagger A|v\rangle} \\ &\leq \sqrt{\langle v|AA^\dagger|v\rangle} \\ &= |\lambda|\,,\end{aligned}$$

where we have used Cauchy-Schwarz in the first equality and normality of A in the equality thereafter. We thus see that Cauchy-Schwarz is tight in the above setting, implying that $A^\dagger|v\rangle$ is proportional to $|v\rangle$. In light of (9), we find that $A^\dagger|v\rangle = \lambda^*|v\rangle$, and so $|v\rangle$ is an eigenstate of $A^\dagger$ with eigenvalue λ^*. □

Lemma 27. *Let $A : \mathcal{H} \to \mathcal{H}$ be a normal matrix for $\mathcal{H} \simeq \mathbb{C}^N$. If A has two eigenvectors $|v\rangle, |w\rangle$ with distinct eigenvalues λ, μ, then $\langle v|w\rangle = 0$, i.e. $|v\rangle$ and $|w\rangle$ are orthogonal.*

Proof. Without loss of generality we can take $\langle v|v\rangle = \langle w|w\rangle = 1$. Using Lemma 26, $A^\dagger|w\rangle = \mu^*|w\rangle$. Then

$$(\lambda - \mu)\langle v|w\rangle = \langle (A - \mu\mathbb{1})v\,|\,w\rangle = \langle v\,|\,(A^\dagger - \mu^*\mathbb{1})w\rangle = 0\,,$$

and so we find $\langle v|w\rangle = 0$. Thus eigenstates with distinct eigenvalues are orthogonal. □

Lemma 28 (Invariance of an eigenspace and its orthogonal complement)**.** *Let A be a normal operator on a finite–dimensional complex Hilbert space $\mathcal{H}$ and let*

$$E_\lambda := \ker\big(A - \lambda\,\mathbb{1}\big)$$

be the λ–eigenspace of A. Then E_λ and $E_\lambda^\perp$ are each invariant under both A and $A^\dagger$. In particular, the restriction

$$A\big|_{E_\lambda^\perp}$$

is normal on the Hilbert space $E_\lambda^\perp$.

Proof. By Lemma 26, if $|y\rangle \in E_\lambda$ then $A^\dagger|y\rangle = \lambda^*|y\rangle$. Now let $|x\rangle \in E_\lambda^\perp$ and $|y\rangle \in E_\lambda$. Then we have $\langle y|A|x\rangle = \langle A^\dagger y|x\rangle = \lambda^*\langle y|x\rangle = 0$. Since $\langle y|A|x\rangle = 0$ for every $|y\rangle \in E_\lambda$, we have $A|x\rangle \in E_\lambda^\perp$. Thus A leaves $E_\lambda^\perp$ invariant. The same calculation with A and $A^\dagger$ interchanged shows $A^\dagger$ leaves $E_\lambda^\perp$ invariant as well. Trivially A leaves E_λ invariant and from the first step $A^\dagger$ leaves E_λ invariant too.

Finally set $B := A\big|_{E_\lambda^\perp}$. Since both A and $A^\dagger$ leave $E_\lambda^\perp$ invariant, the adjoint of B with respect to the inner product on $E_\lambda^\perp$ is $B^\dagger = A^\dagger\big|_{E_\lambda^\perp}$. Hence

$$B^\dagger B = \big(A^\dagger A\big)\big|_{E_\lambda^\perp} = \big(AA^\dagger\big)\big|_{E_\lambda^\perp} = BB^\dagger,$$

so B is normal on $E_\lambda^\perp$. □

With the above lemmas at hand, we finally turn to the proof of Theorem 25.

Proof of Theorem 25. We prove (1) ⇒ (2) by induction on N. The case $N = 1$ is immediate. Assume the claim holds for all dimensions smaller than N.

By Lemma 26 the operator A has an eigenvalue λ and a nonzero eigenstate. Let $E_\lambda = \ker(A - \lambda\,\mathbb{1})$ and choose an orthonormal basis $\{|v_1\rangle, \ldots, |v_r\rangle\}$ of E_λ. By Lemma 28 the orthogonal complement $E_\lambda^\perp$ is invariant under both A and $A^\dagger$. Hence the restriction

$$B := A\big|_{E_\lambda^\perp}$$

is a normal operator on the Hilbert space $E_\lambda^\perp$ whose dimension is $N - r$. By the induction hypothesis there exists an orthonormal basis $\{|v_{r+1}\rangle, \ldots, |v_N\rangle\}$ of $E_\lambda^\perp$ consisting of eigenstates of B, hence of A. Together with $\{|v_1\rangle, \ldots, |v_r\rangle\}$ this gives an orthonormal eigenbasis of $\mathcal{H}$. Writing A in this basis yields

$$A = \sum_{j=1}^{N} \lambda_j\, |v_j\rangle\langle v_j|\,,$$

with $\lambda_j = \lambda$ for $j \le r$ and λ_j equal to the eigenvalues of B for $j > r$. This proves (2).

For (2) ⇒ (1) we compute

$$A^\dagger = \sum_{j=1}^{N} \lambda_j^*\, |v_j\rangle\langle v_j| \quad \text{and} \quad A^\dagger A = \sum_{j=1}^{N} |\lambda_j|^2\, |v_j\rangle\langle v_j| = AA^\dagger\,,$$

and so A is normal. This completes the proof. □

Remark 29 (Hermitian and unitary cases)**.** *If $A = A^\dagger$ then every λ_j is real and $A = \sum_j \lambda_j |v_j\rangle\langle v_j|$. If A is unitary then every λ_j has $|\lambda_j| = 1$ and $A = \sum_j e^{i\theta_j}|v_j\rangle\langle v_j|$*

As an application, consider the following definition.

Definition 30 (Projector)**.** *A **projector** P on $\mathcal{H}$ is a Hermitian idempotent: $P = P^\dagger = P^2$. Equivalently, $P \succeq 0$ and its eigenvalues lie in $\{0, 1\}$.*

Hermiticity implies that all of the eigenvalues of P are real and positive semi-definiteness implies that all of the eigenvalues are nonnegative. Then $P^2 = P$ means that the eigenvalues are either 0 or 1. Supposing $\mathcal{H} \simeq \mathbb{C}^N$, we can use the spectral decomposition to write P as

$$P = \sum_{i=1}^{r} 1 \cdot |v_i\rangle\langle v_i| + \sum_{i=r+1}^{N} 0 \cdot |v_i\rangle\langle v_i| = \sum_{i=1}^{r} |v_i\rangle\langle v_i|$$

for some orthonormal basis $\{|v_i\rangle\}_{i=1}^N$, where r is the rank of the projector. Then P is a projector onto the r-dimensional subspace of $\mathcal{H}$ spanned by $\{|v_i\rangle\}_{i=1}^r$. We can check that $P^\perp = \mathbb{1} - P$ is also a projector onto the orthogonal complement.

Having covered the essence of bra-ket notation, we turn to presenting the axioms of quantum mechanics a la Dirac (with some refinements).

2.2.2. *The axioms*

Here we give the standard axioms of quantum mechanics, with some commentary. The axioms describe the basic mathematical objects of quantum theory, and tether them to observable reality. In the form presented below, the axioms may seem somewhat abstract, and we will discuss this unusual feature shortly. We have tailored the axioms to the finite-dimensional setting for clarity.

(1) **(Quantum states fully describe a system at fixed time.)** At a fixed moment in time, a quantum system about which we have maximal information is fully described by some state vector $|\psi\rangle$ with unit norm living in a Hilbert space $\mathcal{H}$.
(2) **(Time evolution of a closed system is unitary.)** If a quantum system is closed (i.e. it is not interacting with any external system) and starts in an initial state $|\psi_0\rangle$, then at any later time T the state $|\psi_T\rangle$ will be related to the original one by some unitary, that is $|\psi_T\rangle = U|\psi_0\rangle$ for some unitary U that may depend on T.
(3) **(Physical properties have associated projectors.)** Any measurable physical property (such as "spin-up along the z-axis", or "the particle is in region R") has an associated projector P. Such an operator P is an example of an **observable**.
(4) **(Measurement and the Born rule.)** Suppose we have a property corresponding to a projector P, and measure whether or not a system with state vector $|\psi\rangle$ (with unit norm) has that property. Then the probability that we measure $|\psi\rangle$ to have the given property is $\langle\psi|P|\psi\rangle$. This is called the **Born rule**. If $|\psi\rangle$ is measured to have the property, then after measurement the new state of the system is

$$\frac{P|\psi\rangle}{\sqrt{\langle\psi|P^\dagger P|\psi\rangle}} = \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}}\,,$$

which also has unit norm (assuming $P|\psi\rangle \neq 0$, in which case we would never have measured $|\psi\rangle$ to have the given property anyway.)

Now we have a number of comments to make in order to unpack the axioms. The first two axioms were motivated by our previous discussions, in which quantum mechanics is framed as norm-preserving dynamics on ℓ^2. The first axiom codifies that a (normalized) vector in a Hilbert space contains everything there is to know about a quantum state, and the second axioms explains that the dynamics of an isolated system is described by unitary dynamics. Unitary dynamics is reversible since (i.e. unitary matrices are invertible), and so in a closed system the future is completely determined by the past and the past is completely determined by the future. An interesting feature of the second axiom is that it does not tell us *which* unitaries we should use. Indeed, given a classical system, we might wonder what kinds of quantum unitary dynamics can reproduce the the classical dynamics in the appropriate regime. This is a subtle question which goes beyond the axioms, and requires additional empirical input.

The first axiom's proviso "about which we have maximal information" deserves explanation. Consider flipping an unbiased coin to decide whether to prepare a system in state $|\psi_0\rangle$ or $|\psi_1\rangle$. After the flip, the system is in state $|\psi_0\rangle$ with probability $1/2$ or state $|\psi_1\rangle$ with probability $1/2$. This probabilistic description reflects our classical ignorance, not any fundamental quantum uncertainty. The system is definitely in one state or the other; *we* simply do not know which. There is a useful formalism for handling such incomplete knowledge, which we will introduce later.

The second axiom's restriction to "closed" systems is similarly important. A closed system does not interact with any external environment. If such interactions were present, we would need to account for our incomplete knowledge of the environment, which we will address later. When a system couples to an external environment, its dynamics can become non-unitary: information leaks irreversibly from our system into the environment, where it becomes inaccessible to us. Despite being non-unitary, these dynamics can be nicely characterized.

While the first and second axioms specify the basic mathematical objects at play, the third and fourth axioms tether those mathematical objects to empirical reality. This is differently structured than e.g. Newton's axioms of classical mechanics, which specify properties like position and momentum but do not explain what it means to measure them, or how to do so.[6]

Now we turn to the third axiom. The third axiom assigns yes/no properties of a quantum system to linear subspaces of the Hilbert space, via projectors onto those subspaces. For instance, the property 'the spin points up in the z-direction' corresponds to some projector P. The opposite property corresponds to the projector $P^\perp = \mathbb{1} - P$ onto the orthogonal complement. If we have a collection of properties corresponding to projectors $P_1, \ldots, P_k$, we call them **compatible** if the corresponding subspaces are mutually orthogonal, i.e. $P_i P_j = 0$ for $i \neq j$. This orthogonality implies $[P_i, P_j] = 0$. Under orthogonality, if a state answers 'yes' to one property

[6]Part of the reason is that position and momentum, at least in some informal sense, were already known to empiricists in Newton's time. Thus people already knew how to measure them. Interestingly, as we all know, one can use Newton's laws to build devices to better measure position and momentum. You might wonder if this would lead to a circular argument: can we use devices, built using principles from Newton's laws, to then do experiments to test Newton's laws? In short, the answer is 'yes', if we (correctly) conceive of such experiments as testing the *consistency* of Newton's laws with empirical reality. Indeed, since measurements of quantities in Newton's theory require Newton's theory for their specification and possibly design, and there is no clear sense in which one can use empirical findings to test Newton's laws *ex nihilo*.

(i.e. $P_i|\psi\rangle = |\psi\rangle$), it automatically answers 'no' to all others (i.e. $P_j|\psi\rangle = 0$ for $j \neq i$). The projectors $P_1, \ldots, P_k$ are **complete** if their corresponding subspaces span all of $\mathcal{H}$, which is equivalent to $P_1 + \cdots + P_k = \mathbb{1}$. Completeness means that a state will always answer 'yes' to at least one property. Then compatibility and completeness together mean that the state will answer 'yes' to exactly one property in the list (and thus 'no' to all others in the list). The following remark captures some useful nomenclature.

Remark 31 (Hermitian observables)**.** *Let $P_1, \ldots, P_k$ correspond to compatible and complete properties. Suppose that my detector registers the real number a_j to indicate 'yes' for property j. (For instance, if the jth property is 'the particle is at position j', then the detector might just output the number j for the position.) Then we can construct the Hermitian observable*

$$A = \sum_{j=1}^{k} a_j \, P_j \tag{10}$$

which encodes measurement outcomes with respect to our list of properties. In particular,

$$\langle\psi|A|\psi\rangle = \sum_{j=1}^{k} a_j \, \langle\psi|P_j|\psi\rangle = \sum_{j=1}^{k} a_j \, \text{Prob}[\text{measure outcome } j] \, ,$$

where in the last equality we used the Born rule from the fourth axiom. The resulting number is the expectation value of the output of our detector. Since by the spectral theorem all Hermitian operators A can be written in the form (10), we call Hermitian operators ***observables****, with the understanding that their physical interpretation in terms of properties comes from their spectral decomposition.*

A consequence of our discussion above is that certain properties may be *incompatible*, i.e. correspond to non-orthogonal subspaces. For instance, properties corresponding to projectors P and Q are said to be incompatible if $[P, Q] \neq 0$. In this case the two measurements do not admit a common eigenbasis, so in general one cannot ascribe sharp values to both properties simultaneously. Typically, if a state has a definite value for the property corresponding to P, then measuring the property corresponding to Q will yield (in light of the fourth axiom) probabilistic results, and the act of measurement can disturb the system so that P is no longer definite. This lack of joint sharpness is the essence of incompatibility, underlies the uncertainty principle, and is one of the distinguishing features of quantum mechanics vis-à-vis classical mechanics.

The fourth axiom is, in a sense, the most mysterious. While the third axioms abstractly explains the relationship between properties of a system and the quantum state of a system, the fourth axiom tethers these properties to probabilistic observable outcomes. To begin, recall that we said that a state $|\psi\rangle$ has the property corresponding to P if $P|\psi\rangle = |\psi\rangle$, and so not have the property if $(\mathbb{1} - P)|\psi\rangle = P^{\perp}|\psi\rangle = |\psi\rangle$ (or equivalently $P|\psi\rangle = 0$). So far we have accounted for the possibilities $P|\psi\rangle = |\psi\rangle$ or 0, but if $|\psi\rangle$ is neither in the subspace corresponding to P or orthogonal to it, then $P|\psi\rangle \neq |\psi\rangle$ and $\neq 0$. The Born rule tells us that we should interpret the norm squares of the projection of $|\psi\rangle$ into P, namely $\langle\psi|P^{\dagger}P|\psi\rangle = \langle\psi|P|\psi\rangle$, as the probability that $|\psi\rangle$ has that property. More peculiar is that when we affirmatively measure $|\psi\rangle$ to have that property, the $|\psi\rangle$ assumes the new state $\frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}}$. This state now *has* the property, since

$$P \cdot \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}} = \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}}$$

Said another way, if we measure a state to affirmatively have a property (whether or not it definitely had the property before), it subsequently *assumes* that property. This is different from classical mechanics: for example, classical mechanics stipulates that if we measure a particle to have position x then it definitely had position x before. In quantum mechanics, by contrast, measuring a particle to be in position x just tells us that the particle is in position x now, even though it might not 'definitively' have had that property before.

We notice another peculiarity of the fourth axiom, which is that the map

$$|\psi\rangle \longmapsto \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}} \tag{11}$$

is not in general unitary (unless $P = \mathbb{1}$ in which case the map is the identity since $|\psi\rangle$ has unit norm). This would appear to violate the second axiom, which necessitates unitary dynamics. However, we were careful in the second axiom to specify that unitary dynamics happens for *closed* systems; in ordinary circumstances, the measurement apparatus is external to the system that it interrogates, and so the non-unitary of (11) is not in conflict with the second axiom. However, the fourth axiom tempts us to consider the following: if we described the detector (which itself is quantum-mechanical) as *part of* the closed system, then the total detector-system dynamics must be unitary; then can the fourth axiom somehow be derived from the other three? This question is both challenging and profound. Its core difficulty is that the first three axioms do not speak of probability whereas the fourth axioms does speak of probability; as such, the question posed would mandate that probability is *emergent* in quantum mechanics. There have been a vast number of attempts to weaken the fourth axiom or to in some sense 'derive' it from the other three (which often involves covertly bringing in a weakening of the fourth axiom anyway). For our purposes, we can think of the fourth axiom is *pragmatic*, in that it tells us what happens, *in practice*, when we measure a quantum system with an external measurement device.[7]

Having abstractly discussed the axioms, some examples are in order.

Example 9 (Dynamics and projective measurements for a single qubit). We work in the two-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$ with the *computational basis*

[7]Related to the previous footnote, we might wonder how we can test quantum mechanics as a theory if we require quantum theory to build the measurement apparatus needed for the tests themselves. As before, the answer is that we are testing the *consistency* of quantum mechanics, and its alignment with empirical reality. One cannot generally test quantum mechanics with detectors solely intelligible through Newtonian mechanics, i.e. you cannot solely use classical to test quantum. But it is fine to use quantum to test quantum, so long as it all works out empirically. And it very much does.