# Lecture Notes on
# Foundations of Quantum Mechanics

Roderich Tumulka*

Winter semester 2017/18

These notes will be updated as the course proceeds.

*Department of Mathematics, Eberhard-Karls University, Auf der Morgenstelle 10, 72076 Tübingen, Germany. Email: roderich.tumulka@uni-tuebingen.de

# 1 Course Overview

**Learning goals of this course:** To understand the rules of quantum mechanics; to understand several important views of how the quantum world works; to understand what is controversial about the orthodox interpretation and why; to be familiar with the surprising phenomena and paradoxes of quantum mechanics.

Quantum mechanics is the field of physics concerned with (or the post-1900 theory of) the motion of electrons, photons, quarks, and other elementary particles, inside atoms or otherwise. It is distinct from classical mechanics, the pre-1900 theory of the motion of physical objects. Quantum mechanics forms the basis of modern physics and covers most of the physics under the conditions on Earth (i.e., not-too-high temperatures or speeds, not-too-strong gravitational fields). "Foundations of quantum mechanics" is the topic concerned with what exactly quantum mechanics means and how to explain the phenomena described by quantum mechanics. It is a controversial topic. Here are some voices critical of the traditional, orthodox view:

> "With very few exceptions (such as Einstein and Laue) [...] I was the only sane person left [in theoretical physics]."
>
> (E. Schrödinger in a 1959 letter)

> "I think I can safely say that nobody understands quantum mechanics."
>
> (R. Feynman, 1965)

> "I think that conventional formulations of quantum theory [...] are unprofessionally vague and ambiguous."
>
> (J. Bell, 1986)

In this course we will be concerned with what kinds of reasons people have for criticizing the orthodox understanding of quantum mechanics, what the alternatives are, and which kinds of arguments have been put forward for or against important views. We will also discuss the rules of quantum mechanics for making empirical predictions; they are uncontroversial. The aspects of quantum mechanics that we discuss also apply to other fields of quantum physics, in particular to quantum field theory.

**Topics** of this course:

- The Schrödinger equation
- The Born rule
- Self-adjoint matrices, axioms of the quantum formalism, collapse of the wave function, decoherence
- The double-slit experiment and variants thereof, interference and superposition

- Spin, the Stern-Gerlach experiment, the Pauli equation, representations of the rotation group
- The Einstein-Podolsky-Rosen argument, entanglement, non-locality, and Bell's theorem
- The paradox of Schrödinger's cat and the quantum measurement problem
- Heisenberg's uncertainty relation
- Interpretations of quantum mechanics (Copenhagen, Bohm's trajectories, Everett's many worlds, spontaneous collapse theories, quantum logic, perhaps others)
- Views of Bohr and Einstein
- POVMs and density matrices
- No-hidden-variables theorems
- Identical particles and the non-trivial topology of their configuration space, bosons and fermions

**Mathematical tools** that will be needed in this course:

- Complex numbers
- Vectors in $n$ dimensions, inner product
- Matrices, their eigenvalues and eigenvectors
- Multivariable calculus
- Probability; continuous random variables, the Gaussian (normal) distribution

The course will involve advanced mathematics, as appropriate for a serious discussion of quantum mechanics, but will not focus on technical methods of problem-solving (such as methods for calculating the ground state energy of the hydrogen atom). Mathematical topics we will discuss in this course:

- Differential operators (such as the Laplace operator) and their analogy to matrices
- Eigenvalues and eigenvectors of differential (and other) operators
- The Hilbert space of square-integrable functions, norm and inner product
- Projection operators
- Fourier transform of a function
- Positive operators and positive-operator-valued measures (POVMs)

- Tensor product of vector spaces
- Trace of a matrix or an operator, partial trace
- Special ordinary and partial differential equations, particularly the Schrödinger equation
- Exponential random variables and the Poisson process

**Philosophical questions** that will come up in this course:

- Is the world deterministic, or stochastic, or neither?
- Can and should logic be revised in response to empirical findings?
- Are there in principle limitations to what we can know about the world (its laws, its state)?
- Which theories are meaningful as fundamental physical theories? In particular:
- If a statement cannot be tested empirically, can it be meaningful? (Positivism versus realism)
- Does a fundamental physical theory have to provide a coherent story of what happens?
- Does that story have to contain elements representing matter in 3-dimensional space in order to be meaningful?

Physicists usually take math classes but not philosophy classes. That doesn't mean, though, that one doesn't use philosophy in physics. It rather means that physicists learn the philosophy they need in physics classes. Philosophy classes are not among the prerequisites of this course, but we will sometimes make connections with the history of philosophy.

# 2 The Schrödinger Equation

One of the fundamental laws of quantum mechanics is the *Schrödinger equation*

$$i\hbar\frac{\partial\psi}{\partial t} = -\sum_{i=1}^{N}\frac{\hbar^2}{2m_i}\nabla_i^2\psi + V\,\psi\,. \tag{2.1}$$

It governs the time evolution of the *wave function* $\psi = \psi_t = \psi(t, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$. (It can be expected to be valid only in the *non-relativistic* regime, i.e., when the speeds of all particles are small compared to the speed of light. In the general case (the *relativistic* case) it needs to be replaced by other equations, such as the *Klein–Gordon equation* and the *Dirac equation.*) We focus first on *spinless* particles and discuss the phenomenon of *spin* later. I use boldface symbols such as $\boldsymbol{x}$ for 3d vectors.

Eq. (19.2) applies to a system of $N$ particles in $\mathbb{R}^3$. The word "particle" is traditionally used for electrons, photons, quarks, etc.. Opinions diverge whether electrons actually are particles in the literal sense (i.e., point-shaped objects, or little grains). A *system* is a subset of the set of all particles in the world. A *configuration* of $N$ particles is a list of their positions; *configuration space* is thus, for our purposes, the Cartesian product of $N$ copies of physical space, or $\mathbb{R}^{3N}$. The wave function of quantum mechanics, at any fixed time, is a function on configuration space, either complex-valued or spinor-valued (as we will explain later); for spinless particles, it is complex-valued, so

$$\psi : \mathbb{R}_t \times \mathbb{R}^{3N}_q \to \mathbb{C}. \tag{2.2}$$

The subscript indicates the variable: $t$ for time, $q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ for the configuration. Note that $i$ in (19.2) either denotes $\sqrt{-1}$ or labels the particles, $i = 1, \ldots, N$; $m_i$ are positive constants, called the *masses* of the particles; $\hbar = h/2\pi$ is a constant of nature, $h$ is called Planck's quantum of action or Planck's constant, $h = 6.63 \times 10^{-34}\,\mathrm{kg\ m^2 s^{-1}}$;

$$\nabla_i = \Big(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial y_i}, \frac{\partial}{\partial z_i}\Big) \tag{2.3}$$

is the derivative operator with respect to the variable $\boldsymbol{x}_i$, $\nabla_i^2$ the corresponding *Laplace operator*

$$\nabla_i^2\psi = \frac{\partial^2\psi}{\partial x_i^2} + \frac{\partial^2\psi}{\partial y_i^2} + \frac{\partial^2\psi}{\partial z_i^2}\,. \tag{2.4}$$

$V$ is a given real-valued function on configuration space, called the *potential energy* or just *potential*.

Fundamentally, the potential in non-relativistic physics is

$$V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sum_{1\le i<j\le N}\frac{e_i e_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} - \sum_{1\le i<j\le N}\frac{G m_i m_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|}\,, \tag{2.5}$$

where

$$|\boldsymbol{x}| = \sqrt{x^2 + y^2 + z^2} \text{ for } \boldsymbol{x} = (x, y, z) \tag{2.6}$$

denotes the Euclidean norm in $\mathbb{R}^3$, $e_i$ are constants called the *electric charges* of the particles (which can be positive, negative, or zero); the first term is called the *Coulomb potential*, the second term is called the *Newtonian gravity potential*, $G$ is a constant of nature called Newton's constant of gravity $G = 6.67 \times 10^{-11}\,\mathrm{kg}^{-1}\mathrm{m}^3\mathrm{s}^{-2}$, and $m_i$ are again the masses. However, when the Schrödinger equation is regarded as an *effective equation* rather than as a fundamental law of nature then the potential $V$ may contain terms arising from particles outside the system interacting with particles belonging to the system. That is why the Schrödinger equation is often considered for rather arbitrary functions $V$, also time-dependent ones. The operator

$$H = -\sum_{i=1}^{N} \frac{\hbar^2}{2m_i} \nabla_i^2 + V \tag{2.7}$$

is called the *Hamiltonian operator*, so the Schrödinger equation can be summarized in the form

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi \,. \tag{2.8}$$

The Schrödinger equation is a *partial differential equation* (PDE). It determines the time evolution of $\psi_t$ in that for a given initial wave function $\psi_0 = \psi(t=0) : \mathbb{R}^{3N} \to \mathbb{C}$ it uniquely fixes $\psi_t$ for any $t \in \mathbb{R}$. The initial time could also taken to be any $t_0 \in \mathbb{R}$ instead of 0.

So far I have not said anything about what this new physical object $\psi$ has to do with the particles. One fundamental connection is

**Born's rule.** *If we measure the system's configuration at time $t$ then the outcome is random with probability density*

$$\rho(q) = \big|\psi_t(q)\big|^2. \tag{2.9}$$

This rule refers to the concept of *probability density*, which means the following. The probability that the random outcome $X \in \mathbb{R}^{3N}$ is any particular point $x \in \mathbb{R}^{3N}$ is zero. However, the probability that $X$ lies in a set $B \subseteq \mathbb{R}^{3N}$ is given by

$$\mathbb{P}(X \in B) = \int_B \rho(q)\, d^{3N}q \tag{2.10}$$

(a $3N$-dimensional volume integral). Instead of $d^{3N}q$, we will often just write $dq$. A density function $\rho$ must be non-negative and normalized,

$$\rho(x) \geq 0\,, \qquad \int_{\mathbb{R}^{3N}} \rho(q)\, dq = 1\,. \tag{2.11}$$

A famous density function in 1 dimension is the *Gaussian density*

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,. \tag{2.12}$$

A random variable with Gaussian density is also called a *normal* (or *normally distributed*) random variable. It has *mean* $\mu \in \mathbb{R}$ and *standard deviation* $\sigma > 0$. The *mean value* or *expectation value* $\mathbb{E}X$ of a random variable $X$ is its average value

$$\mathbb{E}X = \int_{\mathbb{R}} x\, \rho(x)\, dx\,. \tag{2.13}$$

The *standard deviation* of $X$ is defined to be $\sqrt{\mathbb{E}(X - \mathbb{E}X)^2}$.

For the Born rule to make sense, we need that

$$\int_{\mathbb{R}^{3N}} |\psi_t(q)|^2\, dq = 1. \tag{2.14}$$

And indeed, the Schrödinger equation guarantees this relation: If it holds for $t = 0$ then it holds for any $t \in \mathbb{R}$. More generally, the Schrödinger equation implies that

$$\int dq\, |\psi_t|^2 = \int dq\, |\psi_0|^2 \tag{2.15}$$

for any $\psi_0$. One says that $\int dq\, |\psi_t|^2$ satisfies a *conservation law*. Indeed, the Schrödinger equation implies a *local conservation law* for $|\psi|^2$; that is, it implies the *continuity equation*[1]

$$\frac{\partial |\psi(t,q)|^2}{\partial t} = -\sum_{i=1}^{N} \nabla_i \cdot \boldsymbol{j}_i(t,q) \quad \text{with} \quad \boldsymbol{j}_i(t,q) = \tfrac{\hbar}{m_i} \mathrm{Im}\Big(\psi^*(t,q) \nabla_i \psi(t,q)\Big)\,, \tag{2.16}$$

where Im means imaginary part, because

$$\frac{\partial}{\partial t}\Big(\psi^*\, \psi\Big) = 2\mathrm{Re}\Big(\psi^*\, \tfrac{-i}{\hbar} H\psi\Big) \tag{2.17}$$

$$= \tfrac{2}{\hbar}\mathrm{Im}\Big(-\sum_{i=1}^{N} \tfrac{\hbar^2}{2m_i} \psi^*\, \nabla_i^2 \psi + \underbrace{V(q)|\psi|^2}_{\text{real}}\Big) \tag{2.18}$$

$$= -\sum_{i=1}^{N} \tfrac{\hbar}{m_i} \mathrm{Im}\Big(\psi^*\, \nabla_i^2 \psi + \underbrace{(\nabla_i \psi^*) \cdot (\nabla_i \psi)}_{\text{real}}\Big) = -\sum_{i=1}^{N} \nabla_i \cdot \boldsymbol{j}_i. \tag{2.19}$$

The continuity equation expresses that the amount of $|\psi|^2$ cannot be created or destroyed, only moved around, and in fact flows with the *current* $(\boldsymbol{j}_1, \ldots, \boldsymbol{j}_N)$. To see this, note that it asserts that the $(3N+1)$-dimensional (configuration-space-time) vector field $j = (|\psi|^2, \boldsymbol{j}_1, \ldots, \boldsymbol{j}_N)$ has vanishing divergence. By the Ostrogradski–Gauss integral theorem (divergence theorem), the surface integral of a vector field equals the volume integral of its divergence, so the surface integral of a divergence-less vector field

[1] I don't know where this name comes from. It has nothing to do with being continuous. It should be called conservation equation.

vanishes. Let the surface be the boundary of a $(3N+1)$-dimensional cylinder $[0,T]\times S$, where $S\subseteq\mathbb{R}^{3N}$ is a ball or any set with smooth boundary $\partial S$. Then the surface integral of $j$ is

$$0 = -\int_S |\psi_0|^2 + \int_S |\psi_t|^2 + \int_0^T dt \int_{\partial S} dA\, n_{\partial S}\cdot j \tag{2.20}$$

with $n_{\partial S}$ the unit normal vector field in $\mathbb{R}^{3N}$ on the boundary of $S$. That is, the amount of $|\psi|^2$ in $S$ at time $T$ differs from the initial amount of $|\psi|^2$ in $S$ by the flux of $j$ across the boundary of $S$ during $[0,T]$—a conservation law. If (and this is indeed the case) there is no flux to infinity, i.e., if the last integral becomes arbitrarily small by taking $S$ to be a sufficiently big ball, then the total amount of $|\psi|^2$ remains constant, see (2.15).

Since the quantity $\int dq\,|\psi|^2$ occurs frequently, it is useful to abbreviate it: The $L^2$ *norm* is defined to be

$$\|\psi\| = \Bigl(\int_{\mathbb{R}^{3N}} dq\,|\psi(q)|^2\Bigr)^{1/2}. \tag{2.21}$$

Thus, $\|\psi_t\| = \|\psi_0\|$, and the Born rule is consistent with the Schrödinger equation, provided the initial datum $\psi_0$ has norm 1, which we will henceforth assume. The wave function $\psi_t$ will in particular be square-integrable, and this makes the space $L^2(\mathbb{R}^{3N})$ of square-integrable functions a natural arena. It is also called the *Hilbert space*, and is the space of all wave functions.

# 3 Unitary Operators in Hilbert Space

In the following, we will often simply write $L^2$ for $L^2(\mathbb{R}^{3N})$. We will leave out many mathematical details (which will be discussed in the course *Mathematical Quantum Theory*).

## 3.1 Existence and Uniqueness of Solutions of the Schrödinger Equation

The Schrödinger equation defines the time evolution of the wave function $\psi_t$. In mathematical terms, this means that for every choice of initial wave function $\psi_0(q)$ there is a unique solution $\psi(t,q)$ of the Schrödinger equation. This leads to the question what exactly is meant by "every" wave function. Remarkably, even when $\psi_0$ is not differentiable, there is still a natural sense in which a "weak solution" or "$L^2$ solution" can be defined. This sense allows for a particularly simple statement:

**Theorem 3.1.** [2] *For a large class of potentials $V$ (including Coulomb, Newton's gravity, every bounded measurable function, and linear combinations thereof) and for every $\psi_0 \in L^2$, there is a unique weak solution $\psi(t,q)$ of the Schrödinger equation with potential $V$ and initial datum $\psi_0$. Moreover, at every time $t$, $\psi_t$ lies again in $L^2$.*

## 3.2 The Time Evolution Operators

Let $U_t : L^2 \to L^2$ be the mapping defined by

$$U_t\psi_0 = \psi_t \,. \tag{3.1}$$

$U_t$ is called the *time evolution operator* or *propagator*. Often, it is not possible to write down an explicit closed formula for $U_t$, but it is nevertheless useful to consider $U_t$. It has the following properties.

First, $U_t$ is a *linear operator*, i.e.

$$U_t(\psi + \phi) = (U_t\psi) + (U_t\phi) \tag{3.2}$$

$$U_t(z\psi) = z\,(U_t\psi) \tag{3.3}$$

for any $\psi, \phi \in L^2$, $z \in \mathbb{C}$. This follows from the fact that the Schrödinger equation is a *linear equation*, or, equivalently, that $H$ is a linear operator. It is common to say *operator* for linear operator.

Second, $U_t$ preserves norms:

$$\|U_t\psi\| = \|\psi\| \,. \tag{3.4}$$

[2] This follows from Stone's theorem and Kato's theorem together. See, e.g., Theorem VIII.8 in M. Reed and B. Simon: *Methods of Modern Mathematical Physics, Vol. 1 (revised edition)*, Academic Press (1980), and Theorem X.16 in M. Reed and B. Simon: *Methods of Modern Mathematical Physics, Vol. 2*, Academic Press (1975).

This is just another way of expressing Eq. (2.15). Operators with this property are called *isometric*.

Third, they obey a *composition law*:

$$U_s U_t = U_{t+s}\,, \quad U_0 = I\,, \tag{3.5}$$

where $I$ denotes the *identity operator*

$$I\psi = \psi\,. \tag{3.6}$$

It follows from (3.5) that $U_t^{-1} = U_{-t}$. In particular, $U_t$ is a bijection. An isometric bijection is also called a *unitary operator*; so $U_t$ is unitary. A family of operators satisfying (3.5) is called a *one-parameter group* of operators. Thus, the propagators form a unitary 1-parameter group.

Fourth,

$$U_t = e^{-iHt/\hbar}\,. \tag{3.7}$$

The exponential of an operator $A$ can be defined by the *exponential series*

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} \tag{3.8}$$

if $A$ is a so-called *bounded operator*; in this case, the series converges. Unfortunately, the Hamiltonian of the Schrödinger equation (19.2) is unbounded. But mathematicians agree about how to define $e^A$ for unbounded operators (of the type that $H$ is); we will not worry about the details of this definition.

Eq. (3.7) is easy to understand: after defining

$$\phi_t := e^{-iHt/\hbar}\psi_0\,, \tag{3.9}$$

one would naively compute as follows:

$$i\hbar\frac{d}{dt}\phi_t = i\hbar\frac{d}{dt}e^{-iHt/\hbar}\psi_0 \tag{3.10}$$

$$= i\hbar\Big(-\frac{iH}{\hbar}\Big)e^{-iHt/\hbar}\psi_0 \tag{3.11}$$

$$= H\phi_t\,, \tag{3.12}$$

so $\phi_t$ is a solution of the Schrödinger equation with $\phi_0 = e^0\psi_0 = \psi_0$, and thus $\phi_t = \psi_t$. The calculation (3.10)–(3.12) can actually be justified for all $\psi_0$ in the domain of $H$, a dense set in $L^2$; we will not go into details here.

## 3.3 Unitary Matrices and Rotations

The space $L^2$ is infinite-dimensional. As a finite-dimensional analog, consider the functions on a finite set, $\psi : \{1, \dots, n\} \to \mathbb{C}$, and the norm

$$\|\psi\| = \left(\sum_{i=1}^{n} \psi(i)\right)^{1/2} \tag{3.13}$$

instead of the $L^2$ norm

$$\|\psi\| = \left( \int |\psi(q)|^2 \, dq \right)^{1/2} . \tag{3.14}$$

A function on $\{1, \ldots, n\}$ is always square-summable (its norm cannot be infinite). It can be written as an $n$-component vector

$$\big(\psi(1), \ldots, \psi(n)\big) , \tag{3.15}$$

and the space of these functions can be identified with $\mathbb{C}^n$.

The linear operators on $\mathbb{C}^n$ are given by the complex $n \times n$ matrices. If a matrix preserves the norm (3.13) as in (3.4), it is automatically bijective and thus unitary. A matrix $U_{ij}$ is unitary iff[3]

$$U^\dagger = U^{-1} , \tag{3.16}$$

where $U^\dagger$, the *adjoint matrix of* $U$, is defined by

$$U^\dagger_{ij} = (U_{ji})^* . \tag{3.17}$$

The norm (3.13) is analogous to the norm (= magnitude = length) of a vector in $\mathbb{R}^3$,

$$|\boldsymbol{u}| = \left( \sum_{i=1}^{3} u_i^2 \right)^{1/2} . \tag{3.18}$$

The norm-preserving operators in $\mathbb{R}^3$ are exactly the *orthogonal matrices*, i.e., those matrices $A$ with

$$A^t = A^{-1} , \tag{3.19}$$

where $A^t$ denotes the transposed matrix, $A^t_{ij} = A_{ji}$. They have a geometric meaning: Each orthogonal matrix is either a rotation around some axis passing through the origin, or a reflection across some plane through the origin, followed by a rotation. The set of orthogonal $3 \times 3$ matrices is denoted $O(3)$. The set of those orthogonal matrices which do not involve a reflection is denoted $SO(3)$ for "special orthogonal matrices"; they correspond to rotations and can be characterized by the condition $\det A > 0$ in addition to (3.19).

In dimension $d > 3$, one can show that the special orthogonal matrices are still compositions (i.e., products) of 2-dimensional rotation matrices such as (for $d = 4$)

$$\begin{pmatrix} \cos\alpha & \sin\alpha & & \\ -\sin\alpha & \cos\alpha & & \\ & & 1 & \\ & & & 1 \end{pmatrix} . \tag{3.20}$$

This rotation does not rotate around an axis, it rotates around a $(d-2)$-dimensional subspace (spanned by the 3rd and 4th axes). However, in $d \geq 4$ dimensions, not every

[3]iff = if and only if

special orthogonal matrix is a rotation around a $(d-2)$-dim. subspace through a certain angle, but several such rotations can occur together, as the following example shows:

$$\begin{pmatrix} \cos\alpha & \sin\alpha & & \\ -\sin\alpha & \cos\alpha & & \\ & & \cos\beta & \sin\beta \\ & & -\sin\beta & \cos\beta \end{pmatrix}. \tag{3.21}$$

We will simply call every special orthogonal $d \times d$ matrix a "rotation."

Since $\mathbb{C}^n$ can be regarded as $\mathbb{R}^{2n}$, and the norm (3.13) then coincides with the $2n$-dimensional version of (3.18), every unitary operator then corresponds to an orthogonal operator, in fact a special orthogonal one. So if you can image $2n$-dimensional space, every unitary operator is geometrically a rotation. Also in $L^2$ it is appropriate to think of a unitary operator as a rotation.

## 3.4 Inner Product

In analogy to the dot product in $\mathbb{R}^3$,

$$\boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{3} u_i v_i \tag{3.22}$$

one defines the *inner product* of two functions $\psi, \phi \in L^2$ to be

$$\langle\psi|\phi\rangle = \int_{\mathbb{R}^{3N}} \psi(q)^* \, \phi(q) \, dq \,. \tag{3.23}$$

It has the following properties:

1. It is *anti-linear* (or *semi-linear* or *conjugate-linear*) in the first argument,

$$\langle\psi + \phi|\chi\rangle = \langle\psi|\chi\rangle + \langle\phi|\chi\rangle \,, \quad \langle z\psi|\phi\rangle = z^* \langle\psi|\phi\rangle \tag{3.24}$$

   for all $\psi, \phi, \chi \in L^2$ and $z \in \mathbb{C}$.

2. It is linear in the second argument,

$$\langle\psi|\phi + \chi\rangle = \langle\psi|\phi\rangle + \langle\psi|\chi\rangle \,, \quad \langle\psi|z\phi\rangle = z\langle\psi|\phi\rangle \tag{3.25}$$

   for all $\psi, \phi, \chi \in L^2$ and $z \in \mathbb{C}$. Properties 1 and 2 together are called *sesqui-linear* (from Latin *sesqui* $= 1\frac{1}{2}$).

3. It is *conjugate-symmetric*,

$$\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^* \tag{3.26}$$

   for all $\psi, \phi \in L^2$.

4. It is *positive definite*,[4]
$$\langle\psi|\psi\rangle > 0 \text{ for } \psi \neq 0\,. \tag{3.27}$$

Note that the dot product in $\mathbb{R}^3$ has the same properties, the *properties of an inner product*, except that the scalars involved lie in $\mathbb{R}$, not $\mathbb{C}$. Another inner product with these properties is defined on $\mathbb{C}^n$ by

$$\langle\psi|\phi\rangle = \sum_{i=1}^{n} \psi(i)^*\,\phi(i)\,. \tag{3.28}$$

The norm can be expressed in terms of the inner product according to

$$\|\psi\| = \sqrt{\langle\psi|\psi\rangle}\,. \tag{3.29}$$

Note that the radicand is $\geq 0$. Conversely, the inner product can be expressed in terms of the norm according to the *polarization identity*

$$\langle\psi|\phi\rangle = \tfrac{1}{4}\Big(\|\psi+\phi\|^2 - \|\psi-\phi\|^2 - i\|\psi+i\phi\|^2 + i\|\psi-i\phi\|^2\Big). \tag{3.30}$$

Its proof is a homework exercise. It follows from the polarization identity that every unitary operator $U$ preserves inner products,

$$\langle U\psi|U\phi\rangle = \langle\psi|\phi\rangle\,. \tag{3.31}$$

(Likewise, every $A \in SO(3)$ preserves dot products, which has the geometrical meaning that a rotation preserves the angle between any two vectors.)

In analogy to the dot product, two functions $\psi, \phi$ with $\langle\psi|\phi\rangle = 0$ are said to be *orthogonal*.

## 3.5 Abstract Hilbert Space

The general and abstract definition of a *vector space* (over $\mathbb{R}$ or over $\mathbb{C}$) is that it is a set $S$ (whose elements are called vectors) together with a prescription for how to add elements of $S$ and a prescription for how to multiply an element of $S$ by a scalar, such that the usual algebraic rules of addition and scalar multiplication are satisfied. Similarly, a *Hilbert space* is a vector space over $\mathbb{C}$ together with an inner product satisfying the *completeness* property: every Cauchy sequence converges. One can then prove the

**Theorem 3.2.** *$L^2$ is a Hilbert space.*

[4] Another math subtlety: This will be true only if we identify two functions $\psi, \phi$ whenever the set $\{q \in \mathbb{R}^{3N} : \psi(q) \neq \phi(q)\}$ has volume 0. It is part of the standard definition of $L^2$ to make these identifications.

# 4 Classical Mechanics

*Classical physics* means pre-quantum (pre-1900) physics. I describe one particular version that could be called *Newtonian mechanics* (even though certain features were not discovered until after Isaac Newton's death). This version is over-simplified in that it leaves out magnetism, electromagnetic fields (which play a role for electromagnetic waves and thus the classical theory of light), and relativity theory.

## 4.1 Definition of Newtonian Mechanics

According to Newtonian mechanics, the world consists of a space, which is a 3-dimensional Euclidean space, and particles moving around in space with time. Here, a *particle* means a material point—a point-shaped physical object. Let us suppose there are $N$ particles in the world (say, $N \approx 10^{80}$), and let us fix a Cartesian coordinate system in Euclidean space. At every time $t$, particle number $i$ ($i = 1, \ldots, N$) has a position $\boldsymbol{Q}_i(t) \in \mathbb{R}^3$. These positions are governed by the *equation of motion*

$$m_i \frac{d^2\boldsymbol{Q}_i}{dt^2} = -\nabla_i V(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_N) \tag{4.1}$$

with $V$ the fundamental potential function of the universe as given in Eq. (2.5). This completes the definition of Newtonian mechanics.

The equation of motion (4.1) is an *ordinary differential equation* of second order (i.e., involving second time derivatives). Once we specify, as initial conditions, the initial positions $\boldsymbol{Q}_i(0)$ and velocities $(d\boldsymbol{Q}_i/dt)(0)$ of every particle, the equation of motion (4.1) determines $\boldsymbol{Q}_i(t)$ for every $i$ and every $t$.

Written explicitly, (4.1) reads

$$m_i \frac{d^2\boldsymbol{Q}_i}{dt^2} = -\sum_{j\neq i} e_i e_j \frac{\boldsymbol{Q}_j - \boldsymbol{Q}_i}{|\boldsymbol{Q}_j - \boldsymbol{Q}_i|^3} + \sum_{j\neq i} G m_i m_j \frac{\boldsymbol{Q}_j - \boldsymbol{Q}_i}{|\boldsymbol{Q}_j - \boldsymbol{Q}_i|^3}\,. \tag{4.2}$$

The right hand side is called the *force* acting on particle $i$; the $j$-th term in the first sum (with the minus sign in front) is called the Coulomb force exerted by particle $j$ on particle $i$; the $j$-th term in the second sum is called the gravitational force exerted by particle $j$ on particle $i$.

Newtonian mechanics is empirically wrong. For example, it entails the absence of interference fringes in the double-slit experiment (and entails wrong predictions about everything that is considered a quantum effect). Nevertheless, it is a coherent theory, a "theory of everything," and often useful to consider as a hypothetical world to compare ours to.

Newtonian mechanics is to be understood in the following way: Physical objects such as tables, baseballs, or dogs consist of huge numbers (such as $10^{24}$) of particles, and they must be regarded as just such an agglomerate of particles. Since Newtonian mechanics governs unambiguously the behavior of each particle, it also completely dictates the behavior of tables, baseballs, and dogs. Put differently, after (4.1) has been given, there is no need to specify any further laws for tables, baseballs, or dogs. Any physical law concerning tables, baseballs, or dogs, is a *consequence* of (4.1). This scheme is called *reductionism.* It makes chemistry and biology sub-fields of physics. (This does not mean, though, that it would be of practical use to try to solve (4.1) for $10^{24}$ or $10^{80}$ particles in order to study the behavior of dogs.) Can everything be reduced to (4.1)? It seems that conscious experiences are an exception—presumably the only one.

When we consider a baseball, we are often particularly interested in the motion of its center $\boldsymbol{Q}(t)$ because we are interested in the motion of the whole ball. It is often possible to give an *effective equation* for the behavior of a variable like $\boldsymbol{Q}(t)$, for example

$$M\frac{d^2\boldsymbol{Q}}{dt^2} = -\gamma\frac{d\boldsymbol{Q}}{dt} - Mg\begin{pmatrix}0\\0\\1\end{pmatrix}, \tag{4.3}$$

where $M$ is the mass of the baseball, the first term on the right hand side is called the *friction force*, the second the *gravitional force of Earth*, $\gamma$ is the *friction coefficient* of the baseball and $g$ the *gravitational field strength of Earth.* The effective equation (4.3) looks quite similar to the fundamental equation (4.1) but (i) it has a different status (it is not a fundamental law), (ii) it is only *approximately* valid, (iii) it contains a term that is not of the form $-\nabla V$ (the friction term), (iv) forces that do obey the form $-\nabla V(\boldsymbol{Q})$ (such as the second force) can have other functions for $V$ (such as $V(\boldsymbol{x}) = Mgx_3$) instead of (2.5).

The theory I call Newtonian mechanics was never actually proposed to give the correct and complete laws of physics (although we can imagine a hypothetical world where it does); for example, it leaves out magnetism. An extension of this theory, which we will not consider further but which is also considered "classical physics," includes electromagnetic fields (governed by Maxwell's field equations) and gravitational fields (governed by Einstein's field equations, also known as the theory of general relativity).

The greatest contributions from a single person to the development of Eq. (4.1) came from Isaac Newton (1643–1727), who suggested (in his *Philosophiae Naturalis Principia Mathematica* 1687) considering ODEs, in fact of second order, suggested "forces" and the form $m\frac{d^2\boldsymbol{Q}}{dt^2} = \text{force}$, and introduced the form of the gravitational force, now known as "Newton's law of universal gravity." Eq. (4.2) was first written down, without the Coulomb term, by Leonhard Euler (1707–1783). The first term was proposed in 1784 by Charles Augustin de Coulomb (1736–1806). Nevertheless, we will call (4.1) and (4.2) "Newton's equation of motion."

## 4.2 Properties of Newtonian Mechanics

If $t \mapsto q(t)$ is a solution of Newton's equation of motion (4.1), then so is $t \mapsto q(-t)$, which is called the *time reverse.* This property is called *time reversal invariance* or *reversibility.* It is a rather surprising property, in view of the irreversibility of many phenomena. But since it has been explained, particularly by Ludwig Boltzmann, how reversibility of the microscopic laws and irreversibility of macroscopic phenomena can be compatible,[5] time reversal invariance has been widely accepted. This was also because time reversal invariance also holds in other, more refined theories after Newtonian mechanics, such as Maxwell's equations of classical electromagnetism, general relativity, and the Schrödinger equation.

**Definition 4.1.** Let $\boldsymbol{v}_i(t) = d\boldsymbol{Q}_i/dt$ denote the velocity of particle $i$. The *energy*, the *momentum*, and the *angular momentum* of the universe are defined to be, respectively,

$$E = \sum_{k=1}^{N} \frac{m_k}{2}\boldsymbol{v}_k^2 - \sum_{\substack{j,k=1\\ j<k}}^{N} \Big(Gm_j m_k - \frac{e_j e_k}{4\pi\varepsilon_0}\Big)\frac{1}{|\boldsymbol{Q}_j - \boldsymbol{Q}_k|} \tag{4.4}$$

$$\boldsymbol{p} = \sum_{k=1}^{N} m_k \boldsymbol{v}_k \tag{4.5}$$

$$\boldsymbol{L} = \sum_{k=1}^{N} m_k \boldsymbol{Q}_k \times \boldsymbol{v}_k\,, \tag{4.6}$$

where $\boldsymbol{v}^2 = \boldsymbol{v}\cdot\boldsymbol{v} = |\boldsymbol{v}|^2$, and $\times$ denotes the cross product in $\mathbb{R}^3$. The first term in (4.4) is called *kinetic energy*, the second one *potential energy*.

**Proposition 4.2.** *$E$, $\boldsymbol{p}$, and $\boldsymbol{L}$ are conserved quantities, i.e., they are time independent.*

Proof: exercise.

## 4.3 Hamiltonian Systems

A *dynamical system* is another name for an ODE. A dynamical system in $\mathbb{R}^n$ can be characterized by specifying the function $F : \Omega \to \mathbb{R}^n$ in

$$\frac{dX}{dt} = F(X,t)\,, \tag{4.7}$$

with $\Omega \subseteq \mathbb{R}^n \times \mathbb{R}$. $F$ can be called a time-dependent vector field on (a possibly time-dependent domain in) $\mathbb{R}^n$. (One often considers a more general concept of ODE, in which $F$ is a time-dependent vector field on a differentiable manifold $M$.)

Newtonian mechanics has a time evolution that belongs to the class of dynamical systems, with $n = 6N$, $X = (\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_N, \boldsymbol{v}_1, \dots, \boldsymbol{v}_N)$, and $\Omega$ (the *phase space*) $= \mathbb{R}^{6N}$ or rather $\Omega = \{(\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_N, \boldsymbol{v}_1, \dots, \boldsymbol{v}_N) \in \mathbb{R}^{6N} : \boldsymbol{Q}_i \neq \boldsymbol{Q}_j \forall i \neq j\}$.

It also belongs to a narrower class, called *Hamiltonian systems.* Simply put, these are dynamical systems for which the vector field $F$ is a certain type of derivative of a scalar

[5] For discussion see, e.g., J. L. Lebowitz: From Time-symmetric Microscopic Dynamics to Time-asymmetric Macroscopic Behavior: An Overview. Pages 63–88 in G. Gallavotti , W. L. Reiter, J. Yngvason (editors): *Boltzmann's Legacy.* Zürich: European Mathematical Society (2008) `http://arxiv.org/abs/0709.0724`

function $H$ called the *Hamiltonian function* or simply the *Hamiltonian.* Namely, $n$ is assumed to be even, $n = 2r$, and denoting the $n$ components of $x$ by $(q_1, \ldots, q_r, p_1, \ldots, p_r)$, the ODE is of the form

$$\frac{dq_i}{dt} = \quad \frac{\partial H}{\partial p_i} \tag{4.8}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}\,. \tag{4.9}$$

Newtonian mechanics fits this definition with $r = 3N$, $q_1, \ldots, q_r$ the $3N$ components of $q = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N)$, $p_1, \ldots, p_r$ the $3N$ components of $p = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)$ (the momenta $\boldsymbol{p}_k = m_k \boldsymbol{v}_k$), and $H = H(q, p)$ the energy (4.4) expressed as a function of $q$ and $p$, that is,

$$H(q,p) = \sum_{k=1}^{N} \frac{\boldsymbol{p}_k^2}{2m_k} - \sum_{\substack{j,k=1\\ j\neq k}}^{N} \Big( G m_j m_k - \frac{e_j e_k}{4\pi\varepsilon_0} \Big) \frac{1}{|\boldsymbol{q}_j - \boldsymbol{q}_k|}\,. \tag{4.10}$$

For readers familiar with manifolds I mention that the natural definition of a Hamiltonian system on a manifold $M$ is as follows. $M$ plays the role of phase space. Let the dimension $n$ of $M$ be even, $n = 2r$, and suppose we are given a symplectic form $\omega$ on $M$, i.e., a non-degenerate differential 2-form whose exterior derivative vanishes. (Non-degenerate means that it has full rank $n$ at every point.) The equation of motion for $t \mapsto x(t) \in M$ reads

$$\omega\Big(\frac{dx}{dt}, \cdot\Big) = dH\,, \tag{4.11}$$

where $dH$ means the exterior derivative of $H$. To make the connection with the case $M = \mathbb{R}^n$ just described, $dH$ is then the gradient of $H$ and $\omega$ the $n \times n$ matrix

$$\omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \tag{4.12}$$

with $I$ the $r \times r$ unit matrix and 0 the $r \times r$ zero matrix; $\omega(dx/dt, \cdot)$ becomes the transpose of $\omega$ applied to the $n$-vector $dx/dt$, and (4.11) reduces to (4.8) and (4.9).

# 5 The Double-Slit Experiment

A few remarks about Feynman's text:

- The word "interference" means the constructive or destructive cooperation (i.e., addition) of waves. The word "diffraction" means more or less the same as interference. The same phenomena arise when using more than two slits. In practice, it is common to use dozens of slits or more (called a diffraction grating). One speaks of "interference fringes" when referring to the bright and dark stripes of an interference patttern.

- Feynman's statement on page 1,

  > [The double slit experiment] has in it the heart of quantum mechanics. In reality, it contains the *only* mystery.

  is a bit too strong. Other mysteries can claim to be on equal footing with this one. Feynman weakened his statement later.

- Feynman's statements

  > We cannot make the mystery go away by "explaining" how it works. (page 1)
  >
  > Many ideas have been concocted to try to explain the curve for $P_{12}$ [...] None of them has succeeded. (page 6)
  >
  > No one has found any machinery behind the law. No one can "explain" any more than we have just "explained." No one will give you any deeper representation of the situation. We have no idea about a more basic mechanism from which these results can be deduced. (page 10)

  are too strong. We will see in Chapters 6, 13, and 15 that Bohmian mechanics and other theories provide some explanation of the double slit experiment.

- Feynman's presentation conveys a sense of mystery and a sense of paradox about quantum mechanics. This will be a recurrent theme in this course, and one question will be whether there is any genuine, irreducible mystery or paradox in quantum mechanics.

- Feynman suggests that the mysterious character of quantum mechanics is not surprising ("perfectly reasonable") "because all of direct, human experience and of human intuition applies to large objects." This argument seems not quite on target to me. After all, the troublesome paradoxes of the double slit are not like the notions we often find hard to imagine (for example, how big the number $6 \times 10^{23}$ is, or what 4-dimensional geometry looks like, or how a big light year is) but which are clearly sensible. They sound more like Alice in Wonderland, like they are not sensible—well, like paradoxes.

Some illustrations I'm showing you, related to the double-slit experiment:

- A picture of interference of water waves
- A picture of actual results of a double-slit experiment taken from A. Tonomura et al., *American Journal of Physics* **57(2)**: 117–120 (1989). In this experiment, 70,000 electrons were detected individually after passing through a double slit.[6] Only one electron at a time went through the double slit. About 1,000 electrons per second went through the double slit, at nearly half the speed of light. Each electron needed about $10^{-8}$ seconds to travel from the double slit to the screen.
- A movie created by B. Thaller (`http://vqm.uni-graz.at/movies.html`) showing a numerical simulation of the Schrödinger equation at a double-slit.

Note that the observations in the double-slit experiment are in agreement with, and in fact follow from, the Born rule and the Schrödinger equation: The relevant system here consists of one electron, so $\psi_t$ is a function in just 3 dimensions. The potential $V$ can be taken to be $+\infty$ (or very large) at every point of the plate containing the two slits—except in the slits themselves, where $V = 0$. Away from the plate, also $V = 0$. The Schrödinger equation governs the behavior of $\psi_t$, with the initial wave function $\psi_0$ being a *wave packet*, e.g., a Gaussian wave packet as in Exercise 4 of Assignment 1,

$$\psi_0(\boldsymbol{x}) = (2\pi\sigma^2)^{-3/4} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} e^{-\frac{\boldsymbol{x}^2}{4\sigma^2}}\,, \tag{5.1}$$

moving toward the double slit. According to the Schrödinger equation, part of $\psi$ will be reflected from the wall, part of it will pass through the two slits. The two parts of the wave emanating from the two slits, $\psi_1$ and $\psi_2$ will overlap and thus interfere, $\psi = \psi_1 + \psi_2$. When we detect the electron, its probability density is given, according to the Born rule, by

$$|\psi|^2 = |\psi_1 + \psi_2|^2 = |\psi_1|^2 + |\psi_2|^2 + 2\,\mathrm{Re}(\psi_1^*\psi_2)\,. \tag{5.2}$$

What if we include a device (such as Feynman's lamp) that will detect the electron at one of the slits? Then we detect the electron twice: once at a slit and once at the backdrop screen. Thus, we either have to regard it as a many-particle problem (involving at least two particles, the electron and the photon), or we need a version of the Born rule suitable for repeated detection. We will study both approaches in later lectures.

[6] More precisely, electrons could pass right or left of a positively charged wire of diameter 1 $\mu$m. Those passing on the right get deflected to the left, and vice versa. Thus, the arrangement leads to the superposition of waves travelling in slightly different directions—just what is needed for interference.

# 6 Bohmian Mechanics

"[Bohmian mechanics] exercises the mind in a very salutary way."
J. Bell, *Speakable and Unspeakable in Quantum Mechanics*, page 171

The situation in quantum mechanics is that we have a set of rules, known as the *quantum formalism*, for computing the possible outcomes and their probabilities for (more or less) any conceivable experiment, and everybody agrees (more or less) about the formalism. What the formalism doesn't tell us, and what is controversial, is what exactly happens during these experiments, and how nature arrives at the outcomes whose probabilities the formalism predicts. There are different theories answering these questions, and Bohmian mechanics is one of them.

Let me elucidate my statements a bit. We have already learned part of the quantum formalism: the Schrödinger equation and the Born rule. These rules have allowed us to predict the possible outcomes of the double-slit experiment with a single electron (easy here: a spot anywhere on the screen) and their probability distribution (here: a probability distribution corresponding to $|\psi|^2$ featuring a sequence of maxima and minima corresponding to interference fringes). What the rules didn't tell us was what exactly happens during this experiment (e.g., how the electron moves). Bohmian mechanics fills this gap.

We have not seen all the rules of the quantum formalism yet. We will later, in Lectures 6 and 8. So far, we have formulated the Born rule only for position measurements, and we have not considered repeated detections.

## 6.1 Definition of Bohmian Mechanics

According to Bohmian mechanics, the world consists of a space, which is a 3-dimensional Euclidean space, and particles (material points) moving around in space with time. Let us suppose there are $N$ particles in the world (say, $N \approx 10^{80}$), and let us fix a Cartesian coordinate system in Euclidean space. At every time $t$, particle number $i$ ($i = 1, \ldots, N$) has a position $\boldsymbol{Q}_i(t) \in \mathbb{R}^3$. These positions are governed by *Bohm's equation of motion*

$$\frac{d\boldsymbol{Q}_i}{dt} = \frac{\hbar}{m_i}\mathrm{Im}\frac{\nabla_i \Psi}{\Psi}(t, Q(t))\,. \tag{6.1}$$

Here, $Q(t) = (\boldsymbol{Q}_1(t), \ldots, \boldsymbol{Q}_N(t))$ is the configuration at time $t$, and $\Psi$ is a wave function that is called the *wave function of the universe* and evolves according to the Schrödinger equation

$$i\hbar\frac{\partial \Psi}{\partial t} = -\sum_{i=1}^{N} \frac{\hbar^2}{2m_i}\nabla_i^2 \Psi + V\,\Psi \tag{6.2}$$

with $V$ given by (2.5). The configuration $Q(0)$ at the initial time of the universe (say, right after the big bang) is chosen randomly by nature with probability density

$$\rho_0(q) = |\Psi_0(q)|^2\,. \tag{6.3}$$

(We write capital $Q$ for the configuration of particles and little $q$ for the configuration variable in either $\rho$ or $\Psi$.) This completes the definition of Bohmian mechanics.

The central fact about Bohmian mechanics is that its predictions agree exactly with those of the quantum formalism (which so far have always been confirmed in experiment). We will understand later why this is so.

Eq. (6.1) is an ordinary differential equation of first order (specifying the velocity rather than the acceleration). Thus, the initial configuration $Q(0)$ determines $Q(t)$ for all $t$, so Bohmian mechanics is a deterministic theory. On the other hand, $Q(t)$ is random because $Q(0)$ is. Note that this randomness does not conflict with determinism. It is a theorem, the *equivariance theorem*, that the probability distribution of $Q(t)$ is given by $|\Psi_t(q)|^2$. We will prove the equivariance theorem later in this Lecture. As a consequence, it is consistent to assume the Born distribution for every $t$. Note that due to the determinism, the Born distribution can be *assumed* only for one time (say $t = 0$); for any other time $t$, then, the distribution of $Q(t)$ is fixed by (6.1). The state of the universe at any time $t$ is given by the pair $(Q(t), \Psi_t)$.

Let us have a closer look at Bohm's equation of motion (6.1). If we recall the formula (2.16) for the probability current then we can rewrite Eq. (6.1) in the form

$$\frac{d\boldsymbol{Q}_i}{dt} = \frac{\boldsymbol{j}_i}{|\Psi|^2} = \frac{\text{probability current}}{\text{probability density}}\,. \tag{6.4}$$

This is a very plausible relation because it is a mathematical fact about any particle system with deterministic velocities that

$$\text{probability current} = \text{velocity} \times \text{probability density}\,. \tag{6.5}$$

We will come back to this relation when we prove equivariance.

Here is another way of re-writing (6.1). A complex number $z$ can be charaterized by its modulus $R \geq 0$ and its phase $S \in \mathbb{R}$, $z = Re^{iS}$. It will be convenient in the following to replace $S$ by $S/\hbar$ (but we will still call $S$ the phase of $z$). Then a complex-valued function $\Psi(t, q)$ can be written in terms of the two real-valued functions $R(t, q)$ and $S(t, q)$ according to

$$\Psi(t, q) = R(t, q)\, e^{iS(t,q)/\hbar}\,. \tag{6.6}$$

Let us plug this into (6.1): Since

$$\nabla_i\Psi = \nabla_i(Re^{iS/\hbar}) \tag{6.7}$$

$$= (\nabla_i R)e^{iS/\hbar} + R\nabla_i e^{iS/\hbar} \tag{6.8}$$

$$= (\nabla_i R)e^{iS/\hbar} + R\frac{i\nabla_i S}{\hbar}e^{iS/\hbar}\,, \tag{6.9}$$

we have that

$$\frac{\hbar}{m_i}\mathrm{Im}\frac{\nabla_i\Psi}{\Psi} = \frac{\hbar}{m_i}\mathrm{Im}\left(\underbrace{\frac{\nabla_i R}{R}}_{\text{real}} + i\frac{\nabla_i S}{\hbar}\right) \tag{6.10}$$

$$= \frac{\hbar}{m_i}\frac{\nabla_i S}{\hbar} = \frac{1}{m_i}\nabla_i S\,. \tag{6.11}$$

Thus, (6.1) can be rewritten as

$$\frac{d\boldsymbol{Q}_i}{dt} = \frac{1}{m_i}\nabla_i S(t, Q(t)) \,. \tag{6.12}$$

In words, the velocity is given (up to a constant factor involving the mass) by the gradient of the phase of the wave function.

A historical note. A few years before the development of the Schrödinger equation, Louis de Broglie had suggested a quantitative rule-of-thumb for wave–particle duality: A particle with momentum $\boldsymbol{p} = m\boldsymbol{v}$ should "correspond" to a wave with wave vector $\boldsymbol{k}$ according to the *de Broglie relation*

$$\boldsymbol{p} = \hbar\boldsymbol{k} \,. \tag{6.13}$$

The wave vector is defined by the relation $\psi = e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$ (so it is defined only for plane waves); it is orthogonal to the wave fronts (surfaces of constant phase), and its magnitude is $|\boldsymbol{k}| = 2\pi/(\text{wave length})$. Now, if the wave is not a plane wave then we can still define a *local wave vector* $\boldsymbol{k}(\boldsymbol{x})$ that is orthogonal to the surface of constant phase and whose magnitude is 1/(rate of phase change). Some thought shows that $\boldsymbol{k}(\boldsymbol{x}) = \nabla S(\boldsymbol{x})/\hbar$. If we use this expression on the right hand side of (6.13) and interpret $\boldsymbol{p}$ as mass times the velocity of the particle, we obtain exactly Eq. (6.12), that is, Bohm's equation of motion.

## 6.2 Historical Overview

The idea that the wave function might determine particle trajectories as a "guiding field" was perhaps first expressed by Albert Einstein around 1923 and considered in detail by John C. Slater in 1924. Bohmian mechanics was developed by Louis de Broglie in 1927 but then abandoned. It was rediscovered independently by Nathan Rosen (known for the Einstein–Rosen bridge in general relativity and the Einstein–Podolsky–Rosen argument) in 1945 and David Bohm in 1952. Bohm was the first to realize that it actually makes the correct predictions, and the first to take it seriously as a physical theory. Several physicists mistakenly believed that Bohmian mechanics makes wrong predictions, including de Broglie, Rosen, and Einstein. Curiously, Bohm's 1952 paper provides a strange presentation of the theory, as Bohm insisted on writing the law of motion as an equation for the acceleration $d^2\boldsymbol{Q}_j/dt^2$, obtained by taking the time derivative of (6.1).

## 6.3 Equivariance

The term "equivariance" comes from the fact that the two relevant quantities, $\rho_t$ and $|\psi_t|^2$, vary equally with $t$. (Here, $\rho_t$ is the distribution arising from $\rho_0$ by transport along the Bohmian trajectories.) The equivariance theorem can be expressed by means of the following diagram:

$$\begin{array}{ccc} \Psi_0 & \longrightarrow & \rho_0 \\ {\scriptstyle U_t}\big\downarrow & & \big\downarrow \\ \Psi_t & \longrightarrow & \rho_t \end{array} \tag{6.14}$$

The horizontal arrows mean taking $|\cdot|^2$, the left vertical arrow means the Schrödinger evolution from time 0 to time $t$, and the right vertical arrow means the transport of probability along the Bohmian trajectories. The statement about this diagram is that both paths along the arrows lead to the same result.

As a preparation for the proof, we note that the equation of motion can be written in the form

$$\frac{dQ}{dt} = v_t(Q(t))\,, \tag{6.15}$$

where $v_t : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$ is the vector field on configuration space $v_t = v = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N)$ whose $i$-th component is

$$\boldsymbol{v}_i = \frac{\hbar}{m_i}\mathrm{Im}\frac{\nabla_i \Psi}{\Psi}\,. \tag{6.16}$$

We now address the following question: If $v_t$ is known for all $t$, and the initial probability distribution $\rho_0$ is known, how can we compute the probability distribution $\rho_t$ at other times? The answer is the continuity equation

$$\frac{\partial \rho_t}{\partial t} = -\mathrm{div}\Big(\rho_t v_t\Big)\,. \tag{6.17}$$

This follows from the fact that the *probability current* is given by $\rho_t v_t$. In fact, in any dimension $d$ ($d = 3N$ or otherwise) and for any density (probability density or energy density or nitrogen density or ...) it is true that

$$\text{current} = \text{density} \times \text{velocity} \tag{6.18}$$

(provided that the velocity vector field $v_t$ is not itself random).

We are now ready to prove the equivariance theorem. (This is not a rigorous proof, but this argument contains the essence of the reason why the equivariance theorem is true.) We first show that

$$\text{if } \rho_t = |\psi_t|^2 \text{ then } \frac{\partial \rho_t}{\partial t} = \frac{\partial}{\partial t}|\psi_t|^2 \tag{6.19}$$

and then conclude that if $\rho_0 = |\psi_0|^2$ then $\rho_t = |\psi_t|^2$ (which is the equivariance theorem). By the continuity equation (6.17) for $\rho_t$ and the continuity equation (2.16) for $|\psi_t|^2$, the right equation in (6.19) is equivalent to

$$-\sum_i \nabla_i \cdot \Big(\rho_t \boldsymbol{v}_i\Big) = -\sum_i \nabla_i \cdot \boldsymbol{j}_i\,. \tag{6.20}$$

As mentioned in (6.4), $\boldsymbol{v}_i = \boldsymbol{j}_i/|\psi_t|^2$. Thus, if $\rho_t = |\psi_t|^2$ then Eq. (6.20) is true, which completes the proof.

## 6.4 The Double-Slit Experiment in Bohmian Mechanics

Let us apply what we know about Bohmian mechanics to $N = 1$ and the wave function of the double-slit experiment. We assume that the particle in the experiment moves as if it was alone in the universe, with the potential $V$ representing the wall with two slits. We will justify that assumption in a later Lecture. We know already what the wave function $\psi(t, \boldsymbol{x})$ looks like (remember the movie). Here is a picture of the possible trajectories of the particle.

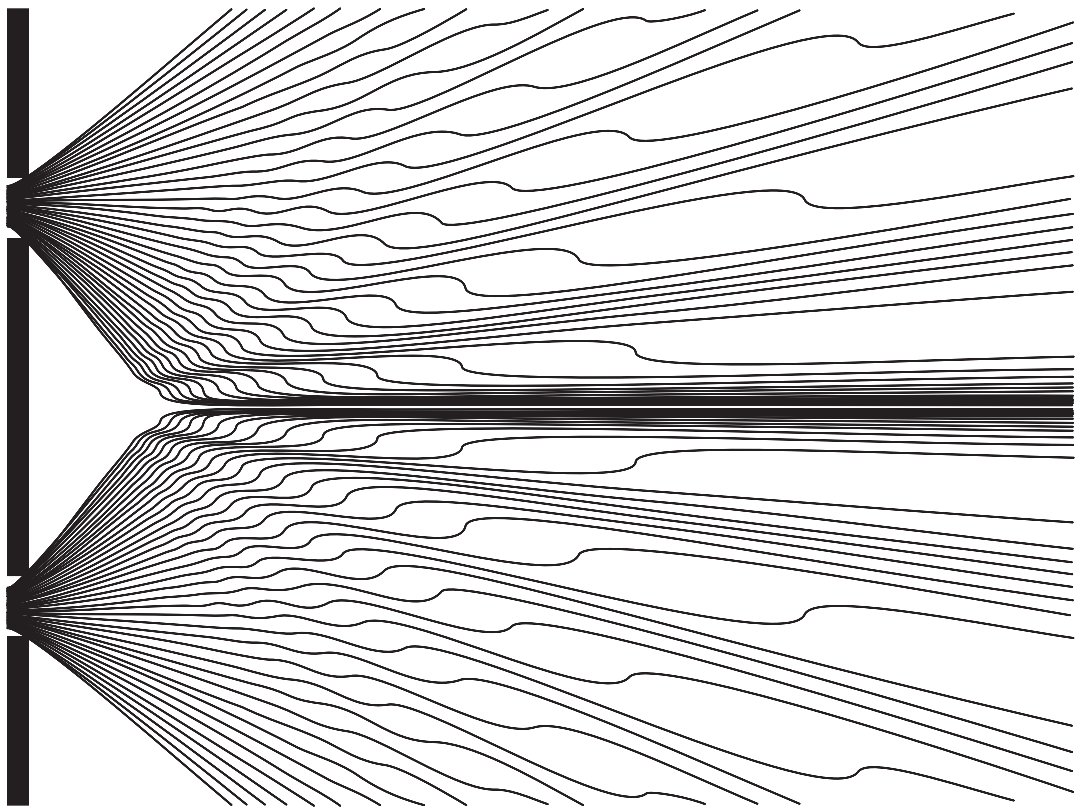

Figure 1: Several alternative Bohmian trajectories of a particle in a double-slit experiment

We know from the equivariance theorem that the position will always have probability distribution $|\psi_t|^2$. Thus, if we detect the particle at time $t$ we find its distribution in agreement with the Born rule.

Note that the particle moves not along straight lines, as it would according to classical mechanics. Note that the wave passes through both slits, while the particle passes through one only. Think about how that answers the paradoxes pointed out by Feynman. Note that the particle trajectories would be different if one slit were closed. Note that we can find out which slit the particle went through without disturbing the interference pattern: check whether the particle arrived in the upper or lower half of the detection screen.

> "Is it not clear from the smallness of the scintillation on the screen that we have to do with a particle? And is it not clear, from the diffraction and interference patterns, that the motion of the particle is directed by a wave? De Broglie showed in detail how the motion of a particle, passing through just one of two holes in screen, could be influenced by waves propagating through both holes. And so influenced that the particle does not go where the waves cancel out, but is attracted to where they cooperate. This idea seems to me so natural and simple, to resolve the wave–particle dilemma in such a clear and ordinary way, that it is a great mystery to me that it was so generally ignored." J. Bell, *Speakable and Unspeakable in Quantum Mechanics*, page 191

Coming back to Feynman's description of the double-slit experiment, we see that his statement that its outcome "cannot be explained" is not quite accurate. It is true that it cannot be explained in Newtonian mechanics, but it can in Bohmian mechanics.

## 6.5 Delayed Choice Experiments

John Archibald Wheeler proposed a variant of the double-slit experiment that may increase further the sense of paradox.[7] Since Wheeler's variant, called the *delayed-choice experiment*, uses no more than the Schrödinger equation and Born's rule, and since we know that Bohmian mechanics can account for that, it is clear that the paradox must disappear in Bohmian mechanics. Let us have a look at what Wheeler's paradox is and how Bohmian mechanics resolves it.

Wheeler considers preparing, by means of a double-slit or in some other way, two wave packets moving in different directions, so that they pass through each other. After passing through each other, they continue moving in different directions and thus get separated again. Wheeler gives the experimenter two choices: either put a screen in the overlap region or put it further away, where the two wave packets have clearly separated. If you put the screen in the overlap region, you will see an interference pattern, which is taken to indicate that the electron is a wave and went through both slits. However, if you put the screen further away, the detection occurs in one of two regions. If the detection occurs in the left (right) region, this is taken to indicate that the particle went through the right (left) slit because a wave packet passing through the right (left) will end up in the left (right) region on the screen. So, Wheeler argued, we can choose whether the electron is particle or wave: if we put the screen far away, it must be particle because we see which slit it went through; if we put the screen in the overlap, it must be wave because we see the interference pattern. Even more, we can force the electron to become wave or particle (and to go through both slits or just one) even *after* it passed through the double-slit! So it seems like there must be *retrocausation*, i.e., situations in which the cause lies in the future of the effect.

[7] J. A. Wheeler: The 'Past' and the 'Delayed-Choice Double-Slit Experiment.' Pages 9–48 in A R. Marlow (editor): *Mathematical Foundations of Quantum Theory*, Academic Press (1978)

Bohmian mechanics illustrates that these conclusions don't actually follow. Bell described that in his article; here are some key points again. To begin with, there is no retrocausation in Bohmian mechanics, as any intervention of observers will change $\psi$ only in the future, not in the past, of the intervention, and the particle trajectory will correspondingly be affected also only in the future. Another basic observation is that with the literal wave-particle dualism of Bohmian mechanics (there is a wave and there is a particle), there is nothing left of the idea that the electron is sometimes a wave and sometimes a particle, and hence even less of the idea that observers could force an electron to become a wave or to become a particle. In detail: the wave passes through both slits, the particle through one; in the overlap region, the two wave packets interfere, and the particle's $|\psi|^2$ distribution features an interference pattern; if there is no screen in the overlap region, then the particle moves on in such a way that the interference pattern disappears and two separate spots form.

After understanding the Bohmian picture of this experiment, some steps in Wheeler's reasoning appear strange. If one assumes that there are no particle trajectories in the quantum world, as one usually does in orthodox quantum mechanics (recall Feynman's chapter), then it would seem natural to say that there is no fact about which slit the electron went through, given that there was no attempt to detect the electron while passing a slit. Surprising it is, then, that Wheeler claims that the detection on the far-away screen reveals which slit it took! How can anything reveal which slit the electron took if the electron didn't take a slit?

There is another interesting aspect to the story that I will call *Wheeler's fallacy*. When you analyze the Bohmian picture in the case of far-away screen, it turns out that the trajectories passing through the left (right) slit end up in the left (right) region. (We will discuss why in the exercises.) So Wheeler makes the wrong retrodiction of which slit the electron passed through! How could this happen? Wheeler noticed that if the right (left) slit is closed, so only one packet comes out, and it comes out of the left (right) slit, then only detection events in the right (left) region occur. This is also true in Bohmian mechanics. Now Wheeler concludes then when wave packets come out of both slit, and if a detection occurs in the right region, then the particle must have passed through the left slit. This is wrong in Bohmian mechanics, and once you realize this, it is obvious that Wheeler's conclusion is a *non sequitur*—a fallacy.

Shahriar Afshar proposed and carried out a further variant of the experiment, known as Afshar's experiment.[8] In this variant, one puts the screen in the far position, but one adds obstacles (that would absorb or reflect electrons) in the overlap region, in fact in those places where the interference is destructive. If an interference pattern occurs in the overlap region, even if it is not observed, then almost no electrons arrive at the obstacles, and almost no electrons get absorbed or reflected. Thus, if all electrons arrive on the far screen in either the left or the right region, as in fact observed in the experiment, then this is indicative that there was an interference pattern in the overlap region even if it was not observed. Afshar argued that this shows that wave and particle must both have

[8]S. S. Afshar: Violation of the principle of complementarity, and its implications. *Proceedings of SPIE* **5866**: 229–244 (2005) `https://arxiv.org/abs/quant-ph/0701027`

existed. Again, Bohmian mechanics easily explains the outcome of this experiment.

# 7 Fourier Transform and Momentum

## 7.1 Fourier Transform

We know from Exercise 2 of Homework 1 that the plane wave $e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$ evolves according to the free Schrödinger equation to

$$e^{i\boldsymbol{k}\cdot\boldsymbol{x}}e^{-i\hbar\boldsymbol{k}^2 t/2m}\,. \tag{7.1}$$

Since the Schrödinger equation is linear, any linear combination of plane waves with different wave vectors $\boldsymbol{k}$,

$$\sum c_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{7.2}$$

with complex coefficients $c_{\boldsymbol{k}}$, will evolve to

$$\sum c_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}e^{-i\hbar\boldsymbol{k}^2 t/2m}\,. \tag{7.3}$$

Moreover, a "continuous linear combination"

$$\int_{\mathbb{R}^3} d^3\boldsymbol{k}\, c(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{7.4}$$

with arbitrary complex $c(\boldsymbol{k})$ will evolve to

$$\int_{\mathbb{R}^3} d^3\boldsymbol{k}\, c(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}e^{-i\hbar\boldsymbol{k}^2 t/2m}\,. \tag{7.5}$$

**Definition 7.1.** For a given function $\psi : \mathbb{R}^d \to \mathbb{C}$, the function

$$\widehat{\psi}(\boldsymbol{k}) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \psi(\boldsymbol{x})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\, d^d\boldsymbol{x} \tag{7.6}$$

is called the *Fourier transform* of $\psi$, $\widehat{\psi} = \mathscr{F}(\psi)$.

**Theorem 7.2.** *Inverse Fourier transformation:*

$$\psi(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \widehat{\psi}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\, d^d\boldsymbol{k}\,. \tag{7.7}$$

Note the different sign in the exponent (it is crucial). If we had not put the pre-factor in (7.6) we would have obtained the pre-factor squared in (7.7).

We have been sloppy in the formulation of the definition and the theorem in that we have not specified the class of functions to which these formulas apply. In fact, (7.6) can be applied whenever $\psi \in L^1$ (the space of all integrable functions, i.e., those with $\|\psi\|_{L^1} = \int d\boldsymbol{x}\, |\psi| < \infty$) and then yields $\widehat{\psi} \in L^\infty$ because $|\widehat{\psi}(\boldsymbol{k})| \leq (2\pi)^{-d/2}\, \|\psi\|_{L^1}$ by the triangle inequality. Conversely, if $\widehat{\psi} \in L^1$, then (7.7) holds, and $\psi \in L^\infty$. However, if $\psi \in L^1 \setminus L^\infty$ then $\widehat{\psi} \notin L^1$, and (7.7) is not literally applicable. A space of interest in this context is the *Schwartz space* $\mathscr{S}$ *of rapidly decaying functions*, which contains the smooth functions $\psi : \mathbb{R}^d \to \mathbb{C}$ such that for every $n \in \mathbb{N}$ and every $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ there is $C_{n,\boldsymbol{\alpha}} > 0$ such that $|\partial^{\boldsymbol{\alpha}}\psi(\boldsymbol{x})| < C_{n,\boldsymbol{\alpha}}\,|\boldsymbol{x}|^{-n}$ for all $\boldsymbol{x} \in \mathbb{R}^d$, where $\partial^{\boldsymbol{\alpha}} := \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$. For example, every Gaussian wave packet lies in $\mathscr{S}$; note that $\mathscr{S} \subset L^1 \cap L^\infty$. It turns out that Fourier transformation maps $\mathscr{S}$ bijectively to itself. Moreover, $\mathscr{S}$ is a dense subspace in $L^2$, and $\mathscr{F}$ can be extended in a unique way to a bounded operator $\mathscr{F} : L^2 \to L^2$, even though the integral (7.6) exists only for $\psi \in L^1 \cap L^2$.

Going back to Eq. (7.5) and taking $c(\boldsymbol{k}) = (2\pi)^{-3/2}\widehat{\psi}_0(\boldsymbol{k})$, we can express the solution of the free Schrödinger equation as

$$\psi_t(\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} d^3\boldsymbol{k}\, \Big(e^{-i\hbar\boldsymbol{k}^2 t/2m}\widehat{\psi}_0(\boldsymbol{k})\Big) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,. \tag{7.8}$$

In words, we can find $\psi_t$ from $\psi_0$ by taking its Fourier transform $\widehat{\psi}_0$, multiplying by a suitable function of $\boldsymbol{k}$, viz., $e^{-i\hbar\boldsymbol{k}^2 t/2m}$, and taking the inverse Fourier transform.

The same trick can be done for $N$ particles. Then $d = 3N$, $\psi = \psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, $\widehat{\psi} = \widehat{\psi}(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_N)$, and the factor to multiply by is

$$\exp\Big(-i\sum_{j=1}^N \frac{\hbar}{2m_j}\boldsymbol{k}_j^2 t\Big) \text{ instead of } \exp\Big(-i\frac{\hbar}{2m}\boldsymbol{k}^2 t\Big)\,. \tag{7.9}$$

Note that we take the Fourier transform only in the *space* variables, not in the *time* variable. There are also applications in which it is useful to consider a Fourier transform in $t$, but not here.

**Example 7.3.** The Fourier transform of a Gauss function. Let $\sigma > 0$ and

$$\psi(\boldsymbol{x}) = C\, e^{-\frac{\boldsymbol{x}^2}{4\sigma^2}} \tag{7.10}$$

with $C$ a constant. Then, using the substitution $\boldsymbol{y} = \boldsymbol{x}/(2\sigma)$,

$$\begin{aligned}
\widehat{\psi}(\boldsymbol{k}) &= \frac{C}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} e^{-\boldsymbol{x}^2/4\sigma^2} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} d^3\boldsymbol{x} && (7.11)\\
&= \underbrace{\frac{2^3 C\sigma^3}{(2\pi)^{3/2}}}_{=:C_2} \int_{\mathbb{R}^3} e^{-\boldsymbol{y}^2 - 2i\sigma\boldsymbol{k}\cdot\boldsymbol{y}} d^3\boldsymbol{y} && (7.12)\\
&= C_2 \int_{\mathbb{R}^3} e^{-(\boldsymbol{y}+i\sigma\boldsymbol{k})^2 - \sigma^2\boldsymbol{k}^2} d^3\boldsymbol{y} && (7.13)\\
&= C_2 e^{-\sigma^2\boldsymbol{k}^2} \int_{\mathbb{R}^3} e^{-(\boldsymbol{y}+i\sigma\boldsymbol{k})^2} d^3\boldsymbol{y} && (7.14)
\end{aligned}$$

The evaluation of the last integral involves the Cauchy integral theorem, varying the path of integration and estimating errors. Here, I just report that the outcome is the constant $\pi^{3/2}$, independently of $\sigma$ and $\boldsymbol{k}$. Thus,

$$\widehat{\psi}(\boldsymbol{k}) = C_3 e^{-\sigma^2\boldsymbol{k}^2} \tag{7.15}$$

with $C_3 = C_2\pi^{3/2}$. In words, the Fourier transform of a Gaussian function is another Gaussian function, but with width $1/(2\sigma)$ instead of $\sigma$. (We see here shadows of the Heisenberg uncertainty relation, which we will discuss in the next chapter.)

**Rule 7.4.** (a)

$$\widehat{\frac{\partial\psi}{\partial x_j}}(\boldsymbol{k}) = ik_j\,\widehat{\psi}(\boldsymbol{k})\,. \tag{7.16}$$

That is, differentiation of $\psi$ corresponds to multiplication of $\widehat{\psi}$ by $ik$.

(b) Conversely,

$$\widehat{-ix_j\psi} = \frac{\partial\widehat{\psi}}{\partial k_j}\,. \tag{7.17}$$

(c) If $f(\boldsymbol{x}) = e^{i\boldsymbol{k}_0\cdot\boldsymbol{x}}\,g(\boldsymbol{x})$, then $\hat{f}(\boldsymbol{k}) = \hat{g}(\boldsymbol{k}-\boldsymbol{k}_0)$.

(d) If $f(\boldsymbol{x}) = g(\boldsymbol{x}-\boldsymbol{x}_0)$, then $\hat{f}(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_0}\,\hat{g}(\boldsymbol{k})$.

*Proof.* (a) Indeed, using integration by parts (and assuming that the boundary terms vanish),

$$\begin{aligned}
\widehat{\frac{\partial\psi}{\partial x_j}}(\boldsymbol{k}) &= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} d^d\boldsymbol{x}\,\frac{\partial\psi}{\partial x_j}(\boldsymbol{x})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} && (7.18)\\
&= -\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} d^d\boldsymbol{x}\,\psi(\boldsymbol{x})\,\frac{\partial}{\partial x_j}e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} && (7.19)\\
&= -\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} d^d\boldsymbol{x}\,\psi(\boldsymbol{x})\,(-ik_j)e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} && (7.20)\\
&= ik_j\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} d^d\boldsymbol{x}\,\psi(\boldsymbol{x})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} && (7.21)\\
&= ik_j\,\widehat{\psi}(\boldsymbol{k})\,. && (7.22)
\end{aligned}$$

(This calculation is a rigorous proof in $\mathscr{S}$.)

(b) Interchanging differentiation and integration (which again is rigorously justified in $\mathscr{S}$),

$$\begin{aligned}
\frac{\partial\widehat{\psi}}{\partial k_j} &= \frac{\partial}{\partial k_j}\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\psi(\boldsymbol{x})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\,d^d\boldsymbol{x} && (7.23)\\
&= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\Bigl(-ix_j\,\psi(\boldsymbol{x})\Bigr)\,e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\,d^d\boldsymbol{x}\,. && (7.24)
\end{aligned}$$

(c) Indeed,

$$\begin{aligned}
\hat{g}(\boldsymbol{k}-\boldsymbol{k}_0) &= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} g(\boldsymbol{x})\,e^{-i(\boldsymbol{k}-\boldsymbol{k}_0)\cdot\boldsymbol{x}}\,d^d\boldsymbol{x} && (7.25)\\
&= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\Bigl(e^{i\boldsymbol{k}_0\cdot\boldsymbol{x}}\,g(\boldsymbol{x})\Bigr)\,e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\,d^d\boldsymbol{x}\,. && (7.26)
\end{aligned}$$

(d) This follows in much the same way.

□

**Example 7.5.** The general Gauss packet

$$\psi(\boldsymbol{x}) = C\, e^{i\boldsymbol{k}_0\cdot\boldsymbol{x}}\, e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_0)^2}{4\sigma^2}} \tag{7.27}$$

has Fourier transform

$$\widehat{\psi}(\boldsymbol{k}) = C_3\, e^{i\boldsymbol{k}_0\cdot\boldsymbol{x}_0}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_0}\, e^{-\sigma^2(\boldsymbol{k}-\boldsymbol{k}_0)^2}\,, \tag{7.28}$$

which is again a general Gaussian packet with center $\boldsymbol{k}_0$ and width $1/(2\sigma)$.

* * *

Fourier transformation defines a *unitary* operator $\mathscr{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, $\mathscr{F}\psi = \widehat{\psi}$. We verify that $\|\mathscr{F}\psi\|_{L^2} = \|\psi\|_{L^2}$ at least for nice $\psi$. Note first that, for $f, g \in L^1 \cap L^2$,

$$\int\Big(\int e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} f(\boldsymbol{k})\, d^d\boldsymbol{k}\Big) g(\boldsymbol{x})\, d^d\boldsymbol{x} = \int\Big(\int e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} g(\boldsymbol{x})\, d^d\boldsymbol{x}\Big) f(\boldsymbol{k})\, d^d\boldsymbol{k} \tag{7.29}$$

by changing the order of integration (which integral is done first). The theorem saying that we are allowed to change the order of integration (for an integrable integrand $fg$) is called *Fubini's theorem.* From Eq. (7.29) we can conclude $\langle g^*|\hat{f}\rangle = \langle \hat{g}^*|f\rangle$. Since

$$(\mathscr{F}f)(\boldsymbol{k})^* = (2\pi)^{-d/2}\int\Big(e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} f(\boldsymbol{x})\Big)^* d^d\boldsymbol{x} = \mathscr{F}^{-1}(f^*)(\boldsymbol{k})\,, \tag{7.30}$$

setting $g = \mathscr{F}^{-1}(f^*) = (\mathscr{F}f)^*$ yields $\langle \hat{f}|\hat{f}\rangle = \langle f|f\rangle$, which completes the proof.

## 7.2 Momentum

"Position measurements" usually consist of detecting the particle. "Momentum measurements" usually consist of letting the particle move freely for a while and then measuring its position.[9]

We now analyze this experiment using Bohmian mechanics. We define the *asymptotic velocity* $\boldsymbol{u}$ to be

$$\boldsymbol{u} = \lim_{t\to\infty} \frac{d\boldsymbol{Q}}{dt}(t) \tag{7.31}$$

if this limit exists. It can also be expressed as

$$\boldsymbol{u} = \lim_{t\to\infty} \frac{\boldsymbol{Q}(t)}{t}\,. \tag{7.32}$$

[9] Alternatively, one lets the particle collide with another particle, makes a "momentum measurement" on the latter, and makes theoretical reasoning about what the momentum of the former must have been.

To understand this, note that $(\boldsymbol{Q}(t) - \boldsymbol{Q}(0))/t$ is the average velocity during the time interval $[0, t]$; if an asymptotic velocity exists (i.e., if the velocity approaches a constant vector $\boldsymbol{u}$) then the average velocity over a long time $t$ will be close to $\boldsymbol{u}$ because for most of the time the velocity will be close to $\boldsymbol{u}$. The term $\boldsymbol{Q}(0)/t$ converges to zero as $t \to \infty$, so we obtain (7.32).

We want the momentum measurement to measure $\boldsymbol{p} := m\boldsymbol{u}$ for a free particle ($V = 0$). So we measure $\boldsymbol{Q}(t)$ for large $t$, divide by $t$, and multiply by $m$. We can and will also take this recipe as the *definition* of a momentum measurement, independently of whether we want to use Bohmian mechanics.

How large do we need $t$ to be? In practice, often not very. When thinking of a particle emitted by a radioactive atom, or coming from a particle collision in an accelerator experiment (such as the Large Hadron Collider LHC in Geneva), a millisecond is usually enough for $d\boldsymbol{Q}/dt$ to become approximately constant.

According to the Born rule, the outcome $\boldsymbol{p}$ is random, and its distribution can be characterized by saying that, for any set $B \subset \mathbb{R}^3$,

$$\mathbb{P}(\boldsymbol{u} \in B) = \lim_{t\to\infty} \mathbb{P}(Q(t)/t \in B) \tag{7.33}$$

$$= \lim_{t\to\infty} \mathbb{P}(Q(t) \in tB) \tag{7.34}$$

$$= \lim_{t\to\infty} \int_{tB} |\psi_t(\boldsymbol{x})|^2 \, d^3\boldsymbol{x}\,, \tag{7.35}$$

where

$$tB = \{t\boldsymbol{x} : \boldsymbol{x} \in B\} \tag{7.36}$$

is the scaled set $B$.

**Theorem 7.6.** *Let $\psi(t, \boldsymbol{x})$ be a solution of the free Schrödinger equation and $B \subseteq \mathbb{R}^3$. Then*

$$\lim_{t\to\infty} \int_{tB} |\psi(t, \boldsymbol{x})|^2 \, d^3\boldsymbol{x} = \int_{mB/\hbar} |\widehat{\psi}_0(\boldsymbol{k})|^2 \, d\boldsymbol{k}\,. \tag{7.37}$$

*As a consequence, the probability density of $\boldsymbol{p}$ is*

$$\frac{1}{\hbar^3}\Big|\widehat{\psi}_0\Big(\frac{\boldsymbol{p}}{\hbar}\Big)\Big|^2. \tag{7.38}$$

The theorem essentially says that when we think of $\psi_0$ as a linear combination of plane waves $e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$ as in Eq. (7.4) or (7.7), then the contribution from a particular value of $\boldsymbol{k}$ will move at a velocity of $\hbar\boldsymbol{k}/m$ (shadows of the de Broglie relation $\boldsymbol{p} = \hbar\boldsymbol{k}$!), and in the long run these contributions will tend to separate in space (i.e., overlap no longer), leaving the contribution from $\boldsymbol{k}$ in the region around $\hbar\boldsymbol{k}t/m$. We see the de Broglie relation again in (7.38) when we insert $\boldsymbol{p}/\hbar$ for $\boldsymbol{k}$ in $\widehat{\psi}$. The upshot of this analysis can be formulated as

**Born's rule for momentum.** *If we measure the momentum of a particle with wave function $\psi$ then the outcome is random with probability density*

$$\rho_{\mathrm{mom}}(\boldsymbol{p}) = \frac{1}{\hbar^3}\Big|\widehat{\psi}\Big(\frac{\boldsymbol{p}}{\hbar}\Big)\Big|^2. \tag{7.39}$$

*Likewise, if we measure the momenta of N particles with joint wave function $\psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, then the outcomes are random with joint probability density*

$$\rho_{\text{mom}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N) = \frac{1}{\hbar^{3N}} \left| \widehat{\psi}\Big(\frac{\boldsymbol{p}_1}{\hbar}, \ldots, \frac{\boldsymbol{p}_N}{\hbar}\Big) \right|^2. \tag{7.40}$$

For this reason, the Fourier transform $\widehat{\psi}$ is also called the *momentum representation of* $\psi$, while $\psi$ itself is called the *position representation* of the wave function.

**Example 7.7.** The general Gaussian wave packet (7.27), whose Born distribution in position space is a Gaussian distribution with mean $\boldsymbol{x}_0$ and width $\sigma$, has momentum distribution

$$\rho_{\text{mom}}(\boldsymbol{p}) = (\text{const.})\, e^{-2(\sigma/\hbar)^2(\boldsymbol{p}-\hbar\boldsymbol{k}_0)^2}\,, \tag{7.41}$$

that is, a Gaussian distribution with mean $\hbar\boldsymbol{k}_0$ and width

$$\sigma_P = \frac{\hbar}{2\sigma}\,. \tag{7.42}$$

In particular, if we want a momentum distribution that is sharply peaked around some value $\boldsymbol{p}_0 = \hbar\boldsymbol{k}_0$, that is, if we want $\sigma_P$ to be small, then $\sigma$ must be large, so $\psi$ must be wide, "close to a plane wave."

## 7.3 Momentum Operator

Let $p_j$, $j = 1, 2, 3$, be the component of the vector $\boldsymbol{p}$ in the direction of the $x_j$-axis. The expectation value of $p_j$ is (using Eq. (7.16) in the fourth line and unitarity of $\mathscr{F}$ in the sixth)

$$\langle p_j \rangle = \int_{\mathbb{R}^3} p_j\, \rho_{\text{mom}}(\boldsymbol{p})\, d^3\boldsymbol{p} \tag{7.43}$$

$$= \int \hbar k_j\, |\widehat{\psi}_0(\boldsymbol{k})|^2\, d^3\boldsymbol{k} \tag{7.44}$$

$$= \Big\langle \widehat{\psi}_0 \Big| \hbar k_j\, \widehat{\psi}_0 \Big\rangle \tag{7.45}$$

$$= \Big\langle \widehat{\psi}_0 \Big| (-i\hbar) \widehat{\frac{\partial \psi_0}{\partial x_j}} \Big\rangle \tag{7.46}$$

$$= -i\hbar \Big\langle \widehat{\psi}_0 \Big| \widehat{\frac{\partial \psi_0}{\partial x_j}} \Big\rangle \tag{7.47}$$

$$= -i\hbar \Big\langle \psi_0 \Big| \frac{\partial \psi_0}{\partial x_j} \Big\rangle \tag{7.48}$$

$$= \Big\langle \psi_0 \Big| \Big(-i\hbar \frac{\partial}{\partial x_j}\Big) \psi_0 \Big\rangle\,. \tag{7.49}$$

This relation motivates calling $P_j = -i\hbar\frac{\partial}{\partial x_j}$ the *momentum operator* in the $x_j$-direction, and $(P_1, P_2, P_3)$ the *vector of momentum operators*.

We note for later use that, by the same reasoning

$$\langle p_j^2 \rangle = \int (\hbar k_j)^2 |\widehat{\psi}_0(\boldsymbol{k})|^2 \, d\boldsymbol{k} = \Big\langle \psi_0 \Big| \Big( -i\hbar \frac{\partial}{\partial x_j} \Big)^2 \psi_0 \Big\rangle \,. \tag{7.50}$$

## 7.4 Tunneling

The *tunnel effect* is another quantum effect that is widely perceived as paradoxical. Consider the 1-d Schrödinger equation with a potential $V$ that has the shape of a *potential barrier* of height $V_0 > 0$. As an idealized example, suppose

$$V(x) = V_0 \, 1_{0 \leq x \leq L} \tag{7.51}$$

or a smooth approximation thereof.

Classically, the motion of a particle in the potential $V$ (or any potential in 1 dimension) can easily be deduced from energy conservation: If the initial position is $< 0$ and the initial momentum is $p_0 > 0$, then the initial energy is $E = p_0^2/2m$, and whenever the particle reaches location $x$, its momentum must be

$$p = \pm \sqrt{2m(E - V(x))} \,. \tag{7.52}$$

In particular, the particle can never reach a region in which $V(x) > E$; so, if $E < V_0$, then the particle will turn around at the barrier and move back to the left.

That is different in quantum mechanics. Consider a Gaussian wave packet, initially to the left of the barrier, with a rather sharp momentum distribution around a $p_0 > 0$ with $p_0^2/2m < V_0$. Then part of the packet will be reflected, and part of it will pass through the barrier! (And the part that passes through is much larger than just the tail of $\rho_{\text{mom}}$ with $p \geq \sqrt{V_0/2m}$.) I will show you another movie created by B. Thaller (`http://vqm.uni-graz.at/movies.html`) with a numerical simulation of the Schrödinger equation with potential (7.51). As a consequence, the Born rule predicts a substantial probability for the particle to show up on the other side of the barrier ("tunneling probability"). Figure 2 shows the Bohmian trajectories for such a situation (with only a small tunneling probability).

For computing the tunneling probability, an easy recipe is to assume that the initial $\psi$ is close to a plane wave consider only the interior part of it that actually looks like a plane wave. One solves the Schrödinger equation for a plane wave arriving, computes the amount of probability current through the barrier, and compares it to the current associated with the arriving wave.[10]

What is paradoxical about tunneling? Perhaps not so much, once we give up Newtonian mechanics and accept that the equation of motion can be non-classical, such as Bohm's. Then it is only to be expected that the trajectories are different, and not surprising that some barriers which Newton's trajectories cannot cross, Bohm's trajectories

[10] For further discussion of why that yields a reasonable result, see T. Norsen: The Pilot-Wave Perspective on Quantum Scattering and Tunneling. *American Journal of Physics* **81**: 258 (2013) `http://arxiv.org/abs/1210.7265`.

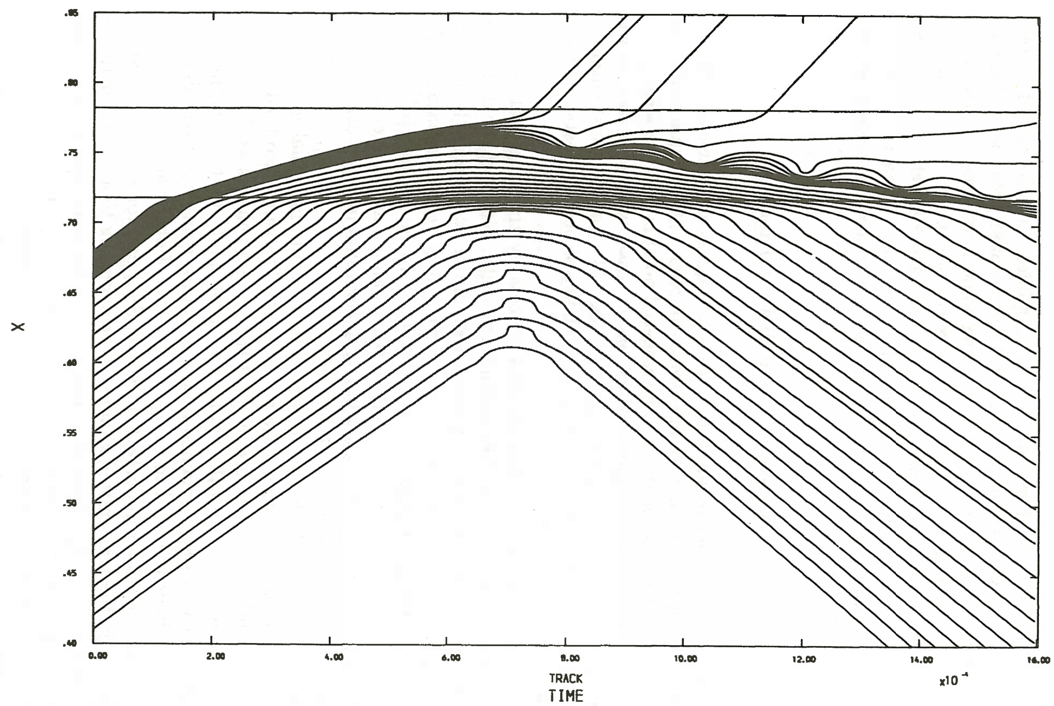


Figure 2: Bohmian trajectories in a tunneling situation. Picture taken from D. Bohm and B. J. Hiley: *The Undivided Universe*, London: Routledge (1993)

can. Part of the sense of paradox comes perhaps from a narrative that is often told when the tunnel effect is introduced: that the particle can "borrow" some energy for a short amount of time by virtue of an energy–time uncertainty relation. This narrative seems not very helpful.

The tunnel effect plays a crucial role in radioactive $\alpha$-decay (where the $\alpha$-particle leaves the nucleus by means of tunneling) and scanning tunneling electron microscopy (where the distance between a needle and a surface is measured by means of measuring the tunneling probability).

There are further related effects: *anti-tunneling* means that a particle gets reflected by a barrier so low that a classical particle with the same initial momentum would be certain to pass it; this happens because a solution of the Schrödinger equation will partly be reflected even at a low barrier. Another effect has been termed *paradoxical reflection*:[11] Consider a downward potential step as in

$$V(x) = -V_0 \, 1_{0 \leq x} \, . \tag{7.53}$$

Classically, a particle coming from the left has probability zero to be reflected back, but according to the Schrödinger equation, wave packets will be partly reflected and partly

[11] For detailed discussion, see P. L. Garrido, S. Goldstein, J. Lukkarinen, and R. Tumulka: Paradoxical Reflection in Quantum Mechanics. *American Journal of Physics* **79(12)**: 1218–1231 (2011) `http://arxiv.org/abs/0808.0610`

transmitted. Remarkably, in the limit $V_0 \to \infty$, the reflection probability converges to 1. "A quantum ball can't roll off a cliff!" On a potential plateau, surrounded by deep downward steps, a particle can be confined for a long time, although finally, in the limit $t \to \infty$, all of the wave function will leave the plateau region and propagate to spatial infinity.

# 8 Operators and Observables

## 8.1 Heisenberg's Uncertainty Relation

As before, $\langle X \rangle$ denotes the expectation of the random variable $X$. The *variance* of the momentum distribution for the initial wave function $\psi \in L^2(\mathbb{R})$ (in one dimension) is

$$\begin{aligned}
\sigma_P^2 &:= \left\langle \big(p - \langle p \rangle\big)^2 \right\rangle && (8.1)\\
&= \left\langle p^2 - 2p\langle p \rangle + \langle p \rangle^2 \right\rangle && (8.2)\\
&= \langle p^2 \rangle - 2\langle p \rangle^2 + \langle p \rangle^2 && (8.3)\\
&= \langle p^2 \rangle - \langle p \rangle^2 && (8.4)\\
&= \langle \psi | P^2 \psi \rangle - \langle \psi | P \psi \rangle^2 && (8.5)\\
&= \left\langle \psi \middle| \big(P - \langle \psi | P \psi \rangle\big)^2 \psi \right\rangle . && (8.6)
\end{aligned}$$

The position distribution $|\psi(x)|^2$ has expectation

$$\langle Q(0) \rangle = \int x |\psi(x)|^2 \, dx = \langle \psi | X \psi \rangle \tag{8.7}$$

with the *position operator* $X\psi(x) = x\psi(x)$. Moreover,

$$\langle Q(0)^2 \rangle = \int x^2 |\psi(x)|^2 \, dx = \langle \psi | X^2 \psi \rangle \, , \tag{8.8}$$

so the variance of the position distribution $|\psi(x)|^2$ is

$$\sigma_X^2 := \int (x - \langle Q(0) \rangle)^2 |\psi(x)|^2 \, dx = \left\langle \psi \middle| \big(X - \langle \psi | X \psi \rangle\big)^2 \psi \right\rangle . \tag{8.9}$$

**Theorem 8.1.** *(Heisenberg uncertainty relation) For any $\psi \in L^2(\mathbb{R})$ with $\|\psi\| = 1$,*

$$\sigma_X \, \sigma_P \geq \frac{\hbar}{2} \, . \tag{8.10}$$

This means that any wave function that is very narrow must have a wide Fourier transform.

**Example 8.2.** Consider the general Gaussian wave packet (7.27), for simplicity in 1 dimension. The standard deviation of the position distribution is $\sigma_X = \sigma$, and we computed the width of the momentum distribution in (7.42). We thus obtain for this $\psi$ that

$$\sigma_X \, \sigma_P = \frac{\hbar}{2} \, , \tag{8.11}$$

just the lowest value allowed by the Heisenberg uncertainty relation.

**Example 8.3.** Consider a wave packet passing through a slit. Let us ignore the part of the wave packet that gets reflected because it did not arrive at the slit, and focus on just the part that makes it through the slit. That is a narrow wave packet, and its standard deviation in position, $\sigma_X$, is approximately the width of the slit. If that is very small then, by the Heisenberg uncertainty relation, $\sigma_P$ must be large, so the wave packet must spread quickly after passing the slit. If the slit is wider, the spreading is weaker.

* * *

In Bohmian mechanics, the Heisenberg uncertainty relation means that whenever the wave function is such that we can know the position of a particle with (small) inaccuracy $\sigma_X$ then we are unable to know its asymptotic velocity better than with inaccuracy $\hbar/(2m\sigma_X)$; thus, we are unable to predict its future position after a large time $t$ (for $V = 0$) better than with inaccuracy $\hbar t/(2m\sigma_X)$. This is a *limitation to knowledge* in Bohmian mechanics.

The Heisenberg uncertainty relation is often understood as excluding the possibility of particle trajectories. If the particle had a trajectory, the reasoning goes, then it would have a precise position and a precise velocity (and thus a precise momentum) at any time, so the position uncertainty would be zero and the momentum uncertainty would be zero, so $\sigma_X = 0$ and $\sigma_P = 0$, in contradiction with (8.10). We know already from Bohmian mechanics that this argument cannot be right. It goes wrong by assuming that if the particle has a precise position and a precise velocity then they can also be precisely known and precisely controlled. Rather, inhabitants of a Bohmian universe, when they know a particle's wave function to be $\varphi(\boldsymbol{x})$, cannot know its position more precisely than the $|\varphi|^2$ distribution allows.

In the traditional, orthodox view of quantum mechanics, it is assumed that electrons do not have trajectories. It is assumed that the wave function is the *complete* description of the electron, in contrast to Bohmian mechanics, where the complete description is given by the pair $(\boldsymbol{Q}, \psi)$, and $\psi$ alone would only be partial information and thus an incomplete description. From these assumptions, it follows that the electron *does not have a position before we attempt to detect it.* Likewise, it does not have a momentum before we attempt to measure it. Thus, in orthodox quantum mechanics the Heisenberg uncertainty relation does *not* amount to a limitation of knowledge because there is no fact in the world that we do not know about when we do not know its position. Unfortunately, the uncertainty relation is often expressed by saying that it is impossible to measure position and momentum at the same time with arbitrary accuracy; while this would be appropriate to say in Bohmian mechanics, it is not in orthodox quantum mechanics because this formulation presumes that position and momentum have values that we could discover by measuring them.

The uncertainty relation is also involved in the double slit experiment as follows. If it did not hold, we could make the electron move exactly orthogonal to the screen after passing through the narrow slits–and arrive very near the center of the screen. Thus, the distribution on the detection screen could not have a second- or third-order maximum.

Since in orthodox quantum mechanics the double-slit experiment is understood as indicative of a paradoxical nature of reality, the uncertainty relation is then understood as "protecting" the paradox from becoming a visible contradiction. A similar argument, as pointed out by Feynman, applies to the photon colliding with the electron for detecting which slit it went through, and its effect of destroying the interference.

## 8.2 Self-adjoint Operators

The following rule is part of the quantum formalism:

> *The most relevant experiments are measurements of certain quantities called observables. Every observable is associated with a self-adjoint operator on Hilbert space.* (8.12)

It is actually a mixture of fact and opinion, as it is formulated from the traditional or orthodox point of view of quantum mechanics. I use this formulation because it is very common. We need to dissect later which part of it is fact, and which is opinion. As Bell wrote (*Speakable and Unspeakable in Quantum Mechanics*, page 215),

> On this list of bad words from good books, the worst of all is 'measurement.'

But first let us get acquainted with the mathematics of self-adjoint operators.

**Theorem 8.4.** *Every bounded operator $A : \mathscr{H} \to \mathscr{H}$ on a Hilbert space $\mathscr{H}$ possesses one and only one adjoint operator $A^\dagger$, defined by the property that for all $\psi, \phi \in \mathscr{H}$,*

$$\langle \psi | A\phi \rangle = \langle A^\dagger \psi | \phi \rangle . \tag{8.13}$$

*For an unbounded operator $A : \mathscr{D}(A) \to \mathscr{H}$ with dense domain $\mathscr{D}(A) \subset \mathscr{H}$, the adjoint operator $A^\dagger$ is uniquely defined by the property* (8.13) *for all $\psi \in \mathscr{D}(A^\dagger)$ and $\phi \in \mathscr{D}(A)$ on the domain*

$$\mathscr{D}(A^\dagger) = \Big\{ \psi \in \mathscr{H} : \exists \chi \in \mathscr{H} \forall \phi \in \mathscr{D}(A) : \langle \psi | A\phi \rangle = \langle \chi | \phi \rangle \Big\} . \tag{8.14}$$

**Definition 8.5.** An operator $A$ on a Hilbert space $\mathscr{H}$ is called *self-adjoint* or *Hermitian* iff $A = A^\dagger$. Then

$$\langle \psi | A\phi \rangle = \langle A\psi | \phi \rangle . \tag{8.15}$$

**Example 8.6.**

- Let $\mathscr{H} = \mathbb{C}^n$. Then every operator $A$ is bounded and correponds to a complex $n \times n$ matrix $A_{ij}$. The matrix of $A^\dagger$ has entries $(A^\dagger)_{ij} = (A_{ji})^*$. Indeed, if we define the matrix $B_{ij}$ by $B_{ij} = (A_{ji})^*$ then we obtain, for any $\psi = (\psi_1, \ldots, \psi_n)$

and $\phi = (\phi_1, \dots, \phi_n)$,

$$\langle\psi|A\phi\rangle = \sum_{i=1}^{n} \psi_i^* (A\phi)_i \tag{8.16}$$

$$= \sum_i \sum_j \psi_i^* A_{ij} \phi_j \tag{8.17}$$

$$= \sum_j \sum_i (A_{ij}^* \psi_i)^* \phi_j \tag{8.18}$$

$$= \sum_j \Big( \sum_i B_{ji} \psi_i \Big)^* \phi_j \tag{8.19}$$

$$= \sum_j (B\psi)_j^* \phi_j \tag{8.20}$$

$$= \langle B\psi|\phi\rangle \,. \tag{8.21}$$

As a consequence, an operator $A$ is self-adjoint iff $A_{ij} = A_{ji}^*$.

- A unitary operator is usually *not* self-adjoint.
- Let $\mathscr{H} = L^2(\mathbb{R}^d)$, and let $A$ be a multiplication operator,

$$A\psi(\boldsymbol{x}) = f(\boldsymbol{x})\,\psi(\boldsymbol{x})\,, \tag{8.22}$$

such as the potential in the Hamiltonian or the position operators. Then $A^\dagger$ is the multiplication operator that multiplies by $f^*$. Indeed,

$$\langle\psi|A\phi\rangle = \int_{\mathbb{R}^d} \psi(\boldsymbol{x})^* f(\boldsymbol{x}) \phi(\boldsymbol{x})\, d\boldsymbol{x} \tag{8.23}$$

$$= \int \big( f^*(\boldsymbol{x})\,\psi(\boldsymbol{x}) \big)^* \phi(\boldsymbol{x})\, d\boldsymbol{x} \tag{8.24}$$

$$= \langle f^*\psi|\phi\rangle \,. \tag{8.25}$$

(This calculation is rigorous if $f$ is bounded. If it is not, them some discussion of the domains of $A$ and $A^\dagger$ is needed.) Thus, $A$ is self-adjoint iff $f$ is real-valued.

- On $\mathscr{H} = L^2(\mathbb{R}^d)$, the momentum operators $P_j = -i\hbar \frac{\partial}{\partial x_j}$ are self-adjoint with the domain given by the first Sobolev space, i.e., the space of functions $\psi$ $L^2$ whose Fourier transform $\widehat{\psi}$ has the property that $\boldsymbol{k} \mapsto |\boldsymbol{k}|\,\widehat{\psi}$ is still square-integrable. The relation (8.15) can easily be verified on nice functions using integration by parts:

$$\langle\psi|P_j\phi\rangle = \int \psi^*(\boldsymbol{x})(-i\hbar)\frac{\partial\phi}{\partial x_j}(\boldsymbol{x})\, d\boldsymbol{x} \tag{8.26}$$

$$= -\int \frac{\partial\psi^*}{\partial x_j}(\boldsymbol{x})(-i\hbar)\phi(\boldsymbol{x})\, d\boldsymbol{x} \tag{8.27}$$

$$= \int \Big( -i\hbar \frac{\partial\psi}{\partial x_j}(\boldsymbol{x}) \Big)^* \phi(\boldsymbol{x})\, d\boldsymbol{x} \tag{8.28}$$

$$= \langle P_j\psi|\phi\rangle \,. \tag{8.29}$$

- In $\mathscr{H} = L^2(\mathbb{R}^d)$, the Hamiltonian is self-adjoint for suitable potentials $V$ on a suitable domain. By formal calculation (leaving aside questions of domains), since

$$H = \sum_{j=1}^{d} \frac{1}{2m} P_j^2 + V\,, \tag{8.30}$$

we have that

$$\langle\psi|H\phi\rangle = \Big\langle \psi \Big| \Big(\sum_j \frac{1}{2m} P_j^2 + V\Big)\phi \Big\rangle \tag{8.31}$$

$$= \sum_j \frac{1}{2m} \langle\psi|P_j P_j \phi\rangle + \langle\psi|V\phi\rangle \tag{8.32}$$

$$= \sum_j \frac{1}{2m} \langle P_j\psi|P_j \phi\rangle + \langle V\psi|\phi\rangle \tag{8.33}$$

$$= \sum_j \frac{1}{2m} \langle P_j P_j\psi|\phi\rangle + \langle V\psi|\phi\rangle \tag{8.34}$$

$$= \Big\langle \Big(\sum_j \frac{P_j^2}{2m} + V\Big)\psi \Big| \phi \Big\rangle \tag{8.35}$$

$$= \langle H\psi|\phi\rangle\,. \tag{8.36}$$

## 8.3 The Spectral Theorem

Before we can formulate Born's rule for arbitrary observables, we need to learn about the spectral theorem.

**Definition 8.7.** If

$$A\psi = \alpha\psi\,, \tag{8.37}$$

where $\alpha$ is a number and $\psi \in \mathscr{H}$ with $\psi \neq 0$, then $\psi$ is called an *eigenvector* (or *eigenfunction*) of $A$ with *eigenvalue* $\alpha$. The number $\alpha$ is called an eigenvalue of $A$ iff there exists $\psi \neq 0$ satisfying (8.37). The set of all eigenvalues is called the *spectrum* of $A$.

If $A$ is self-adjoint then all eigenvalues must be real. Indeed, if $\psi$ is an eigenvector of $A$ with eigenvalue $\alpha$, then

$$\alpha\langle\psi|\psi\rangle = \langle\psi|\alpha\psi\rangle = \langle\psi|A\psi\rangle = \langle A\psi|\psi\rangle = \langle\alpha\psi|\psi\rangle = \alpha^*\langle\psi|\psi\rangle\,, \tag{8.38}$$

so $\alpha = \alpha^*$ or $\alpha \in \mathbb{R}$.

**Theorem 8.8.** *(Spectral theorem) For every self-adjoint operator $A$ in a Hilbert space $\mathscr{H}$ there is a (generalized) orthonormal basis $\{\phi_{\alpha,\lambda}\}$ consisting of eigenvectors of $A$,*

$$A\phi_{\alpha,\lambda} = \alpha\phi_{\alpha,\lambda}\,. \tag{8.39}$$

*($\phi_{\alpha,\lambda}$ has two indices because for every eigenvalue $\alpha$ there may be several eigenvectors, indexed by $\lambda$.)*

An *orthonormal basis* (ONB) is a set $\{\phi_n\}$ elements of the Hilbert space $\mathscr{H}$ such that (a) $\langle \phi_m | \phi_n \rangle = \delta_{mn}$ and (b) every $\psi \in \mathscr{H}$ can be written as a linear combination of the $\phi_n$,

$$\psi = \sum_n c_n \, \phi_n \, . \tag{8.40}$$

A "generalized" orthonormal basis allows a continuous variable $k$ instead of $n$,

$$\psi = \int dk \, c_k \, \phi_k \, , \tag{8.41}$$

as we have encountered with Fourier transformation, where $k = \boldsymbol{k} \in \mathbb{R}^d$, $c_k = \widehat{\psi}(\boldsymbol{k})$, and

$$\phi_{\boldsymbol{k}}(\boldsymbol{x}) = (2\pi)^{-d/2} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \, . \tag{8.42}$$

For a generalized ONB, we don't require that the $\phi_k$ themselves be elements of $\mathscr{H}$; e.g., the $\phi_{\boldsymbol{k}}$ of Fourier transformation are not square-integrable. We will often write a $\sum$ sign even when we mean the integral over $k$. The precise definition of "generalized ONB" is a unitary isomorphism $U : \mathscr{H} \to L^2(\Omega)$ with $\Omega$ the set of possible $k$-values indexing the generalized ONB and $U\psi(k) = c_k$. For example, for the generalized ONB (8.42), $U = \mathscr{F}$. A "non-generalized" ONB then corresponds to a unitary isomorphism $U : \mathscr{H} \to \ell^2 = L^2(\mathbb{N})$.

The big payoff of the spectral theorem is that in this ONB, it is very easy to carry out the operator $A$: If

$$\psi = \sum_{\alpha,\lambda} c_{\alpha,\lambda} \, \phi_{\alpha,\lambda} \tag{8.43}$$

then

$$A\psi = \sum_{\alpha,\lambda} \alpha \, c_{\alpha,\lambda} \, \phi_{\alpha,\lambda} \, . \tag{8.44}$$

Put differently, in this ONB, $A$ is a multiplication operator, multiplying by the function $f(k) = f(\alpha, \lambda) = \alpha$. For example, in the Fourier basis (8.42), the momentum operator $P_j$ is multiplication by $\hbar k_j$.

Put differently again, the matrix associated with the operator $A$ in the ONB $\phi_{\alpha,\lambda}$ is a diagonal matrix. That is why one says that this ONB *diagonalizes* $A$.

**Born's rule for arbitrary observables.** *If we measure the observable $A$ on a system with wave function $\psi$ then the outcome is random with probability distribution*

$$\rho_A(\alpha) = \sum_\lambda \big| \langle \phi_{\alpha,\lambda} | \psi \rangle \big|^2 = \sum_\lambda \big| U\psi(\alpha, \lambda) \big|^2 \, , \tag{8.45}$$

*where $\phi_{\alpha,\lambda}$ is an orthonormal basis diagonalizing $A$; $\rho_A$ may mean either probability density or just probability, depending on whether $\alpha$ is a discrete or continuous variable.*

# 9 Spin

The phenomenon known as *spin* does not mean that the particle is spinning around its axis, though it is in some ways similar. The simplest description of the phenomenon is to say that the wave function of an electron (at time $t$) is actually not of the form $\psi : \mathbb{R}^3 \to \mathbb{C}$ but instead $\psi : \mathbb{R}^3 \to \mathbb{C}^2$. The space $\mathbb{C}^2$ is called *spin-space* and its elements *spinors* (short for spin-vectors). We will in the following write $S$ for spin-space.

## 9.1 Spinors and Pauli Matrices

Apart from being a 2-dimensional Hilbert space, spin space has the further property that with every spinor is associated a vector in physical space $\mathbb{R}^3$. This relation can be expressed as a function

$$\boldsymbol{\omega} : S \to \mathbb{R}^3 \,, \tag{9.1}$$

given explicitly by

$$\boldsymbol{\omega}(\phi) = \left( \sum_{r,s=1}^{2} \phi_r^* (\sigma_1)_{rs} \phi_s, \sum_{r,s=1}^{2} \phi_r^* (\sigma_2)_{rs} \phi_s, \sum_{r,s=1}^{2} \phi_r^* (\sigma_3)_{rs} \phi_s \right), \tag{9.2}$$

where $\sigma_i$ are the three *Pauli matrices*

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{9.3}$$

Obviously, they are self-adjoint complex $2 \times 2$ matrices. It is common to write $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ for the *vector of Pauli matrices.* With this notation, and writing

$$\phi^* \chi = \sum_{s=1}^{2} (\phi_s)^* \chi_s \tag{9.4}$$

for the inner product in spin-space, Eq. (9.2) can be expressed more succinctly as

$$\boldsymbol{\omega}(\phi) = \phi^* \boldsymbol{\sigma} \phi \,. \tag{9.5}$$

For example, the spinor $\phi = (1, 0)$ has $\boldsymbol{\omega}(\phi) = (0, 0, 1)$, which points in the $+z$-direction; $(1, 0)$ is therefore called a *spin-up spinor.* The spinor $(0, 1)$ has $\boldsymbol{\omega}(0, 1) = (0, 0, -1)$, which points in the $-z$-direction; $(0, 1)$ is therefore called a *spin-down spinor.* $\boldsymbol{\omega}$ has the properties

$$\boldsymbol{\omega}(z\phi) = |z|^2 \boldsymbol{\omega}(\phi) \tag{9.6}$$

and (homework problem)

$$|\boldsymbol{\omega}(\phi)| = \|\phi\|_S^2 = \phi^* \phi \,, \tag{9.7}$$

so unit spinors are associated with unit vectors.

Spinors have the curious property that if we rotate a spinor $\phi$ in spin-space through an angle $\theta$, with angles in Hilbert space defined by the relation

$$\cos\theta = \frac{\big|\langle\phi|\chi\rangle\big|}{\|\phi\|\,\|\chi\|}, \tag{9.8}$$

the corresponding direction $\boldsymbol{\omega}(\phi)$ in real space rotates through an angle $2\theta$. For example, $(0,1)$ can be obtained from $(1,0)$ by rotating through $90°$, while the corresponding vector is rotated from the $+z$ to the $-z$-direction, and thus through $180°$. Expressed the other way around, spinors rotate by half the angle of vectors. That is way one says that electrons have *spin one half*. As a consequence, a rotation in real space by $360°$ will correspond to one by $180°$ in spin space and carry $\phi$ to $-\phi$, whereas a rotation in real space by $720°$ will carry $\phi$ to itself.

There are also other types of spinors, other than spin-$\frac{1}{2}$: spin-1, spin-$\frac{3}{2}$, spin-2, spin-$\frac{5}{2}$, etc. The space of spin-$s$ spinors has complex dimension $2s+1$, and the analogs of the Pauli matrices are $(2s+1)\times(2s+1)$ matrices. In this context, wave functions $\psi:\mathbb{R}^3\to\mathbb{C}$ are said to have *spin 0*. Electrons, quarks, and all known species of *matter particles* have spin $\frac{1}{2}$; the photon has spin 1; all known species of *force particles* have integer spin; the only elementary particle species with spin 0 in the standard model of particle physics is the *Higgs particle* or *Higgs boson*, which was experimentally confirmed in 2012 at the Large Hadron Collider (LHC) of CERN in Geneva, Switzerland.

## 9.2 The Pauli Equation

When spin is taken into account, the Schrödinger equation reads a little differently. The appropriate version is known as the *Pauli equation*. We will not study this equation in detail; we write it down mainly for the sake of completeness:

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m}\Big(-i\hbar\nabla - \boldsymbol{A}(\boldsymbol{x})\Big)^2\psi(\boldsymbol{x}) - \frac{\hbar}{2m}\boldsymbol{\sigma}\cdot\boldsymbol{B}(\boldsymbol{x})\psi(\boldsymbol{x}) + V(\boldsymbol{x})\psi(\boldsymbol{x}) \tag{9.9}$$

with $\boldsymbol{B}$ the magnetic field, $V$ the electric and gravitational potential, $\boldsymbol{A}$ the magnetic vector potential defined by the property

$$\boldsymbol{B} = \nabla\times\boldsymbol{A}\,. \tag{9.10}$$

(In words, $\boldsymbol{B}$ is the curl of $\boldsymbol{A}$. The vector potential is, in fact, not uniquely defined by this property, but different vector potentials satisfying (9.10) for the same magnetic field can be translated into each other by *gauge transformations*, i.e., by different $\boldsymbol{x}$-dependent choices of the orthonormal basis in spin-space $S$.)

The Hilbert space of wave functions with spin is denoted $L^2(\mathbb{R}^3,\mathbb{C}^2)$ and contains the square-integrable functions $\mathbb{R}^3\to\mathbb{C}^2$. The inner product is

$$\langle\psi|\phi\rangle = \int_{\mathbb{R}^3} d^3\boldsymbol{x}\,\psi^*(\boldsymbol{x})\,\phi(\boldsymbol{x}) = \int_{\mathbb{R}^3} d^3\boldsymbol{x}\sum_{s=1}^{2}\psi_s^*(\boldsymbol{x})\,\phi_s(\boldsymbol{x})\,. \tag{9.11}$$

## 9.3 The Stern–Gerlach Experiment

Let us write

$$\psi(\boldsymbol{x}) = \begin{pmatrix} \psi_1(\boldsymbol{x}) \\ \psi_2(\boldsymbol{x}) \end{pmatrix} . \tag{9.12}$$

In the first half of a Stern–Gerlach experiment (first done in 1927 with silver atoms), a wave packet moves through a magnetic field that is carefully designed so as to deflect $\psi_1(\boldsymbol{x})$ in a different direction than $\psi_2(\boldsymbol{x})$, and thus to separate the two components in space. Put differently, if the initial wave function $\psi(t=0)$ has support in the ball $B_r(\boldsymbol{y})$ of radius $r$ around the center $\boldsymbol{y}$ then the final wave function $\psi(t=1)$ (i.e., the wave function after passing through the magnetic field) is such that $\psi_1(\boldsymbol{x}, t=1)$ has support in $B_+ := B_r\big(\boldsymbol{y} + (1,0,d)\big)$ and $\psi_2(\boldsymbol{x}, t=1)$ in $B_- := B_r\big(\boldsymbol{y} + (1,0,-d)\big)$ with deflection distance $d > r$ (so that $\psi_1$ and $\psi_2$ do not overlap). The arrangement creating this magnetic field is called a Stern–Gerlach magnet. In the second half of the Stern–Gerlach experiment, one applies detectors to the regions $B_\pm$. If the electron is found in $B_+$ then the outcome of the experiment is said to be *up*, if in $B_-$ then *down*.

A case of particular interest is that the initial wave function satisfies

$$\psi_s(\boldsymbol{x}) = \phi_s \, \chi(\boldsymbol{x}) \, , \tag{9.13}$$

where $\phi \in S$, $\|\phi\|_S = 1$, and $\chi : \mathbb{R}^3 \to \mathbb{C}$, $\|\chi\| = 1$. One says that for such a $\psi$, the spin degree of freedom is disentangled from the spatial degrees of freedom. (Before, we have considered many-particle wave functions for which some particles were disentangled from others. We may also consider a single particle and say that the $x$ variable is disentangled from the $y$ and $z$ variables iff $\psi(x,y,z) = f(x)\, g(y,z)$.)

In the case (9.13), the wave function after passing the magnet is

$$\begin{pmatrix} \phi_1 \, \chi\big(\boldsymbol{x} - (1,0,d)\big) \\ \phi_2 \, \chi\big(\boldsymbol{x} - (1,0,-d)\big) \end{pmatrix} , \tag{9.14}$$

and it follows from the Born rule for position that the probability of outcome "up" is $|\phi_1|^2$ and that of "down" is $|\phi_2|^2$.

These probabilities agree with the general Born rule (8.45) for the observable $A = \sigma_3$ on the Hilbert space $\mathscr{H} = S$. The spinors $\phi_{+1} = (1,0)$ and $\phi_{-1} = (0,1)$ form an orthonormal basis of $S$ consisting of eigenvectors of $\sigma_3$ (with eigenvalues $+1$ and $-1$, respectively); $\phi$ plays the role of $\psi$ in (8.45); its coefficients in the ONB referred to in Eq. (8.45) are $\langle \phi_{+1} | \psi \rangle = \phi_1$ and $\langle \phi_{-1} | \psi \rangle = \phi_2$. That is why the Stern–Gerlach experiment is often called a "measurement of $\sigma_3$", or a "measurement of the $z$ component of spin."

The Stern–Gerlach magnet can be rotated into any direction. For example, by rotating by 90° around the $x$-axis (a rotation that will map the $z$-axis to the $y$-axis), we obtain an arrangement that will deflect part of the initial wave packet $\psi$ in the $+y$-direction and another part in the $-y$-direction. However, these parts are not $\phi_1$ and $\phi_2$.

Instead, they are the parts along a different ONB of $S$:

$$\phi^{(+)} = \frac{1}{\sqrt{2}}(1, i) \text{ and } \phi^{(-)} = \frac{1}{\sqrt{2}}(1, -i) \text{ form an ONB of } S \text{ with } \boldsymbol{\omega}(\phi^{(\pm)}) = (0, \pm 1, 0). \tag{9.15}$$

That is, and $\psi : \mathbb{R}^3 \to S$ can be written as $\psi(\boldsymbol{x}) = c_+(\boldsymbol{x})\phi^{(+)} + c_-(\boldsymbol{x})\phi^{(-)}$, and these two terms will get spatially separated (in the $\pm y$ direction, in fact). The probabilities of outcomes "up" and "down" are then $\int d\boldsymbol{x}|c_\pm(\boldsymbol{x})|^2$. In the special case (9.13), the probabilities are just $|c_\pm|^2$, where $\phi = c_+\phi^{(+)} + c_-\phi^{(-)}$. Equivalently, the probabilities are $|\langle \phi^{(\pm)}|\phi\rangle|^2$. These values are in agreement with the general Born rule for $A = \sigma_2$ because $\phi^{(\pm)}$ are eigenvectors of $\sigma_2$ with eigenvalues $\pm 1$.

Generally, if the Stern–Gerlach magnet is rotated from the $z$-direction to direction $\boldsymbol{n}$, where $\boldsymbol{n}$ is any unit vector in $\mathbb{R}^3$, then the probabilities of its outcomes are governed by the Born rule (8.45) for $A = \boldsymbol{n} \cdot \boldsymbol{\sigma}$, which for any $\boldsymbol{n}$ is a self-adjoint $2 \times 2$ matrix with eigenvalues $\pm 1$.

## 9.4 Bohmian Mechanics with Spin

John Bell figured out in 1966 how to do Bohmian mechanics for particles with spin. It is surprisingly simple. Here is the single-particle version. Replace the Schrödinger equation by the Pauli equation and Bohm's equation of motion (6.1) by

$$\frac{d\boldsymbol{Q}}{dt} = \frac{\hbar}{m} \mathrm{Im} \frac{\psi^* \nabla \psi}{\psi^* \psi}(t, \boldsymbol{Q}(t)) \,. \tag{9.16}$$

Recall that $\psi^*\psi$ means the inner product in spin-space, so the denominator means

$$\psi^*(\boldsymbol{x})\psi(\boldsymbol{x}) = |\psi_1(\boldsymbol{x})|^2 + |\psi_2(\boldsymbol{x})|^2 \,. \tag{9.17}$$

Likewise, the numerator means

$$\psi^*(\boldsymbol{x})\nabla\psi(\boldsymbol{x}) = \psi_1^*(\boldsymbol{x})\nabla\psi_1(\boldsymbol{x}) + \psi_2^*(\boldsymbol{x})\nabla\psi_2(\boldsymbol{x}) \,. \tag{9.18}$$

The initial position $\boldsymbol{Q}(0)$ is assumed to be random with probability density

$$\rho(\boldsymbol{x}) = |\psi(\boldsymbol{x})|^2 := \|\psi(\boldsymbol{x})\|_S^2 = \psi^*(\boldsymbol{x})\psi(\boldsymbol{x}) = |\psi_1(\boldsymbol{x})|^2 + |\psi_2(\boldsymbol{x})|^2 \,. \tag{9.19}$$

It follows that $\boldsymbol{Q}(t)$ has probability density $|\psi_t|^2$ at every $t$. This version of the *equivariance theorem* can be obtained by a very similar computation as in the spinless case, involving the following variant of the continuity equation:

$$\frac{\partial |\psi(\boldsymbol{x}, t)|^2}{\partial t} = -\nabla \cdot \Big( \frac{\hbar}{m} \mathrm{Im}(\psi^* \nabla \psi) \Big) \,. \tag{9.20}$$

As a consequence of the equivariance theorem, Bohmian mechanics leads to the correct probabilities for the Stern–Gerlach experiment.

## 9.5 Is an Electron a Spinning Ball?

If it were then the following paradox would arise. According to classical electrodynamics (which of course is well confirmed for macroscopic objects), a spinning, electrically charged object behaves like a magnet in two ways: it creates its own magnetic field, and it reacts to an external magnetic field. Just as the strength of the electric charge can be expressed by a number, the charge $e$, the strength of the magnet can be expressed by a vector, the *magnetic dipole moment* or just *magnetic moment* $\boldsymbol{\mu}$. Its direction points from the south pole to the north pole, and its magnitude is the strength of the magnet. The magnetic moment of a charge $e$ spinning at angular frequency $\omega$ around the axis along the unit vector $\boldsymbol{u}$ is, according to classical electrodynamics,

$$\boldsymbol{\mu} = \gamma e \omega \boldsymbol{u}\,, \tag{9.21}$$

where the factor $\gamma$ depends on the size and shape of the object. Furthermore, if such an object flies through a Stern–Gerlach magnet oriented in direction $\boldsymbol{n}$ then, still according to classical electrodynamics, it gets deflected by an amount proportional to $\boldsymbol{\mu} \cdot \boldsymbol{n}$. Put differently, the Stern–Gerlach experiment for a classical object measures $\mu_z$, or the component of $\boldsymbol{\mu}$ in the direction of $\boldsymbol{n}$. The vector $\omega\boldsymbol{u}$ is called the *spin vector*.

Where is the paradox? It is that different choices of $\boldsymbol{n}$, when applied to objects with the same $\boldsymbol{\mu}$, would lead to a continuous interval of deflections $[-\gamma|e|\omega, +\gamma|e|\omega]$, whereas the Stern–Gerlach experiment, for whichever choice of $\boldsymbol{n}$, leads to a discrete set $\{+d, -d\}$ of two possible deflections.

The latter fact was called by Wolfgang Pauli the "non-classical two-valuedness of spin." This makes it hard to come up with a theory in which the outcome of a Stern–Gerlach experiment has anything to do with a spinning motion. While Feynman went too far when claiming that the double-slit experiment does not permit any deeper explanation, it seems safe to say that the Stern–Gerlach experiment does not permit an explanation in terms of spinning balls. Note also that Bohmian mechanics does not involve any spinning motion to account for (what has come to be called) spin.

## 9.6 Many-Particle Systems

The wave function of $N$ electrons is of the form

$$\psi_{s_1,s_2,\ldots,s_N}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)\,, \tag{9.22}$$

where each $\boldsymbol{x}_j$ varies in $\mathbb{R}^3$ and each index $s_j$ in $\{1, 2\}$. Thus, at any configuration, $\psi$ has $2^N$ complex components, or $\psi : \mathbb{R}^{3N} \to \mathbb{C}^{2^N}$. The Pauli equation then reads

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m}\sum_{k=1}^{N}\bigl(-i\hbar\nabla_k - \boldsymbol{A}(\boldsymbol{x}_k)\bigr)^2\psi - \sum_{k=1}^{N}\frac{\hbar}{2m}\boldsymbol{\sigma}_{(k)} \cdot \boldsymbol{B}(\boldsymbol{x}_k)\psi + V\psi\,, \tag{9.23}$$

where $\boldsymbol{\sigma}_{(k)}$ means $\boldsymbol{\sigma}$ acting on the index $s_k$ of $\psi$. In Bohm's equation of motion (9.16), replace $\boldsymbol{Q} \in \mathbb{R}^3$ by $Q \in \mathbb{R}^{3N}$ and sum over all spin indices $s_j$ whenever taking the spin inner product $\phi^*\psi$.

## 9.7 Representations of $SO(3)$

A deeper understanding of spinors comes from group representations.[12] Let us start easily. Consider the wave function of a single particle. Suppose it were, instead of a complex scalar field, a vector field, so $\boldsymbol{\psi} : \mathbb{R}^3 \to \mathbb{R}^3$. Well, it should be complex, so we complexify the vector field, $\boldsymbol{\psi} : \mathbb{R}^3 \to \mathbb{C}^3$. Now rotate your coordinate system according to $R \in SO(3)$. Then in the new coordinates, the same physical wave function is represented by a different mathematical function,

$$\tilde{\boldsymbol{\psi}}(\boldsymbol{x}) = R\boldsymbol{\psi}(R^{-1}\boldsymbol{x})\,. \tag{9.24}$$

Instead of real-valued potentials, the Schrödinger equation could then include matrix-valued potentials, provided the matrices are always self-adjoint:

$$i\hbar\frac{\partial \boldsymbol{\psi}}{\partial t} = -\frac{\hbar^2}{2m}\Delta\boldsymbol{\psi} + V\boldsymbol{\psi}\,. \tag{9.25}$$

Now consider another possibility: that the wave function is tensor-valued, $\psi_{ab}$ with $a, b = 1, 2, 3$. Then in a rotated coordinate system,

$$\tilde{\psi}_{ab}(\boldsymbol{x}) = \sum_{c,d=1}^{3} R_{ac}R_{bd}\psi_{cd}(R^{-1}\boldsymbol{x})\,. \tag{9.26}$$

What the two examples have in common is that the components of the wave function get transformed as well according to the scheme, for $\psi : \mathbb{R}^3 \to \mathbb{C}^d$,

$$\tilde{\psi}_r(\boldsymbol{x}) = \sum_{s=1}^{d} M_{rs}(R)\,\psi_s(R^{-1}\boldsymbol{x})\,. \tag{9.27}$$

The matrices $M(R)$ satisfy the composition law

$$M(R_1)\,M(R_2) = M(R_1R_2) \quad \text{and} \quad M(I) = I\,, \tag{9.28}$$

which means that they form a *representation* of the group $SO(3)$ of rotations—in other words, a homomorphism from $SO(3)$ to $GL(\mathbb{C}^d)$, the "general linear group" comprising all invertible operators on $\mathbb{C}^d$. Further representations of $SO(3)$ provide further possible value spaces for wave functions $\psi$.

Spin space $S$ for spin-$\frac{1}{2}$ is almost of this kind, but there is one more complication: $SO(3)$ is represented, not by linear mappings $S \to S$, but by mappings $P(S) \to P(S)$ consistent with linear mappings, where $P(S)$ is the set of all 1-dimensional subspaces of $S$ (called the *projective space* of $S$). This seems fitting as two wave functions that differ only by a phase factor, $\phi(x) = e^{i\theta}\psi(x)$, are usually regarded as representing the same physical state (they yield the same Born distribution, at all times and for all

[12]More details about the topic of this section can be found in R. U. Sexl and H. K. Urbantke: *Relativity, Groups, Particles*, Springer-Verlag (2001).

observables, and the same Bohmian trajectories for all times). That is, one can say that a wave function is really an element of $P(\mathscr{H})$ rather than $\mathscr{H}$ because every normalized element of $\mathbb{C}\psi$ is as good as $\psi$.

By a mapping $F : P(S) \to P(S)$ consistent with a linear mapping, I mean an $F$ such there is a linear mapping $M : S \to S$ with $F(\mathbb{C}\psi) = \mathbb{C}M\psi$. While $M$ determines $F$ uniquely, $F$ does not determine $M$, as $zM$ with any $z \in \mathbb{C} \setminus \{0\}$ leads to the same $F$. In particular, if we are given $F(R)$ and want an $M(R)$, then $M(R)$ is always another possible candidate. For spin-$\frac{1}{2}$, it turns out that while $F(R_1)\,F(R_2) = F(R_1R_2)$ as it should, $M(R)$ can at best be found in such a way that

$$M(R_1)\,M(R_2) = \pm M(R_1R_2)\,. \tag{9.29}$$

This sign mismatch has something to do with the halved angles. The $M$ are elements of $SU(2)$ (unitary with determinant 1), and with every element $R$ of $SO(3)$ are associated two elements of $SU(2)$ that differ by a sign.

This association can actually be regarded as a mapping

$$\varphi : SU(2) \to SO(3)\,,\; M \mapsto R\,. \tag{9.30}$$

This mapping $\varphi$ is a group homomorphism (i.e., $\varphi(M_1)\varphi(M_2) = \varphi(M_1M_2)$ and $\varphi(I) = I$), is smooth, two-to-one $[\varphi(-M) = \varphi(M)]$, and locally a diffeomorphism. The situation is similar to the group homomorphism $\chi : \mathbb{R} \to U(1)$, $\theta \mapsto e^{i\theta}$, which is also smooth, many-to-one, and locally a diffeomorphism; just like $\mathbb{R}$ is what you get from the circle $U(1)$ when you unfold it, $SU(2)$ is what you get from $SO(3)$ when you "unfold" it. (The unfolding of a manifold $\mathcal{Q}$ is called the *covering space* $\widehat{\mathcal{Q}}$, so $\widehat{SO(3)} = SU(2)$.) For every continuous curve $\gamma$ in $SO(3)$ starting in $I$, there is a unique continuous curve $\hat{\gamma}$ in $SU(2)$ with $\varphi \circ \hat{\gamma} = \gamma$, called the *lift* of $\gamma$. Thus, continuous rotations in $\mathbb{R}^3$ can be translated uniquely into continuous rotations in $S$.

The upshot of all this is that spinors are one of the various types of mathematical objects (besides vectors and tensors) that react to rotations in a well-defined way, and that is why they qualify as possible values of a wave function.

# 10 The Projection Postulate

## 10.1 Notation

In the *Dirac notation* one writes $|\psi\rangle$ for $\psi$. This may seem like a waste of symbols at first, but often it is the opposite, as it allows us to replace a notation such as $\phi_1, \phi_2, \ldots$ by $|1\rangle, |2\rangle, \ldots$. Of course, a definition is needed for what $|n\rangle$ means, just as one would be needed for $\phi_n$. It is also convenient when using long subscripts, such as replacing $\psi_{\text{left slit}}$ by $|\text{left slit}\rangle$. In spin space $S$, one commonly writes

$$|z\text{-up}\rangle = \big|\!\uparrow\big\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \quad |z\text{-down}\rangle = \big|\!\downarrow\big\rangle = \begin{pmatrix}0\\1\end{pmatrix} \tag{10.1}$$

$$|y\text{-up}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix}, \quad |y\text{-down}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix} \tag{10.2}$$

$$|x\text{-up}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \quad |x\text{-down}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} \tag{10.3}$$

(Compare to Eq. (9.15) and Exercise 11 in Assignment 4, and to Maudlin's article.)

Furthermore, in the Dirac notation one writes $\langle\phi|$ for the mapping $\mathscr{H} \to \mathbb{C}$ given by $\psi \mapsto \langle\phi|\psi\rangle$. Obviously, $\langle\phi|$ applied to $|\psi\rangle$ gives $\langle\phi|\psi\rangle$, which suggested the notation. Paul Dirac called $\langle\phi|$ a *bra* and $|\psi\rangle$ a *ket*. Obviously, $\langle\phi|A|\psi\rangle$ means the same as $\langle\phi|A\psi\rangle$. Dirac suggested that for self-adjoint $A$, the notation $\langle\phi|A|\psi\rangle$ conveys better that $A$ can be applied equally well to either $\phi$ or $\psi$. $|\phi\rangle\langle\phi|$ is an operator that maps $\psi$ to $|\phi\rangle\langle\phi|\psi\rangle = \langle\phi|\psi\rangle\phi$. If $\phi$ is a unit vector then this is the part of $\psi$ parallel to $\phi$, or the *projection of $\psi$ to $\phi$.*

Another common and useful notation is $\otimes$, called the *tensor product.* For

$$\Psi(x,y) = \psi(x)\,\phi(y) \tag{10.4}$$

one writes

$$\Psi = \psi \otimes \phi\,. \tag{10.5}$$

Likewise, for Eq. (9.13) one writes $\psi = \phi \otimes \chi$.

The symbol $\otimes$ has another meaning when applied to Hilbert spaces.

$$L^2(x,y) = L^2(x) \otimes L^2(y)\,, \tag{10.6}$$

where $L^2(x)$ means the square-integrable functions of $x$, etc. Likewise, when we replace the continuous variable $y$ by the discrete index $s$ for spin, the tensor product of the Hilbert space $\mathbb{C}^2$ of vectors $\phi_s$ and the Hilbert space $L^2(\mathbb{R}^3,\mathbb{C})$ of wave functions $\chi(\boldsymbol{x})$ is the Hilbert space $L^2(\mathbb{R}^3,\mathbb{C}^2)$ of wave functions $\psi_s(\boldsymbol{x})$:

$$\mathbb{C}^2 \otimes L^2(\mathbb{R}^3,\mathbb{C}) = L^2(\mathbb{R}^3,\mathbb{C}^2)\,. \tag{10.7}$$

Another notation we use is

$$f(t-) = \lim_{s\nearrow t} f(s)\,, \quad f(t+) = \lim_{s\searrow t} f(s) \tag{10.8}$$

for the left and right limits of a function $f$ at a jump.

## 10.2 The Projection Postulate

Here is the last rule of the quantum formalism:

**Projection postulate.** *If we measure the observable $A$ at time $t$ on a system with wave function $\psi_{t-}$ and obtain the outcome $\alpha$ then the system's wave function $\psi_{t+}$ right after the measurement is the eigenfunction of $A$ with eigenvalue $\alpha$. If there are several mutually orthogonal eigenfunctions, then*

$$\psi_{t+} = C \sum_{\lambda} |\phi_{\alpha,\lambda}\rangle\langle\phi_{\alpha,\lambda}|\psi_{t-}\rangle\,, \tag{10.9}$$

*where $C > 0$ is the normalizing constant.*

If $\lambda$ is a continuous variable, then $\sum_\lambda$ should be $\int d\lambda$. The value of $C$ is, explicitly,

$$C = \left\| \sum_{\lambda} |\phi_{\alpha,\lambda}\rangle\langle\phi_{\alpha,\lambda}|\psi_{t-}\rangle \right\|^{-1}. \tag{10.10}$$

## 10.3 Projection and Eigenspace

To get a better feeling for what the expression on the RHS of (10.9) means, consider a vector $\psi = \psi_{t-}$ and an ONB $\phi_n = \phi_{\alpha,\lambda}$, and expand $\psi$ in that basis:

$$\psi = \sum_n c_n \phi_n\,. \tag{10.11}$$

The coefficients are then given by

$$c_m = \langle\phi_m|\psi\rangle \tag{10.12}$$

because

$$\langle\phi_m|\psi\rangle = \Big\langle \phi_m \Big| \sum_n c_n\phi_n \Big\rangle = \sum_n c_n \langle\phi_m|\phi_n\rangle = \sum_n c_n \delta_{mn} = c_m\,. \tag{10.13}$$

Now change $\psi$ by replacing some of the coefficients $c_n$ by zero while retaining the others unchanged:

$$\tilde{\psi} = \sum_{n\in J} c_n\phi_n\,, \tag{10.14}$$

where $J$ is the set of those indices retained. This procedure is called *projection* to the subspace spanned by $\{\phi_n : n \in J\}$, and the projection operator is

$$P = \sum_{n\in J} |\phi_n\rangle\langle\phi_n|\,. \tag{10.15}$$

(The only projections we consider are *orthogonal projections.*) An operator $P$ is a projection iff it is self-adjoint [$P = P^\dagger$] and idempotent [$P^2 = P$]; equivalently, iff it is self-adjoint and the spectrum (set of generalized eigenvalues) is $\{0, 1\}$.

In Eq. (10.9), the index $n$ numbers the index pairs $(\alpha, \lambda)$, and the subset $J$ corresponds to those pairs that have a given $\alpha$ and arbitrary $\lambda$. Except for the factor $C$, the RHS of (10.9) is the corresponding projection of $\psi_{t-}$, which gives the projection postulate its name. The subspace of Hilbert space spanned by the $\phi_{\alpha,\lambda}$ with given $\alpha$ is the *eigenspace* of $A$ with eigenvalue $\alpha$, which is the set of all eigenvectors of $A$ with eigenvalue $\alpha$ (together with the zero vector).

For every closed subspace, there is a projection operator that projects to this subspace. For example, for any region $B \subseteq \mathbb{R}^{3N}$ in configuration space, the functions whose support lies in $B$ (i.e., which vanish outside $B$) form an $\infty$-dimensional closed subspace of $L^2(\mathbb{R}^{3N})$. The projection to this subspace is

$$(P_B\psi)(q) = \begin{cases} \psi(q) & q \in B \\ 0 & q \notin B\,, \end{cases} \tag{10.16}$$

that is, multiplication by the characteristic function $1_B$ of $B$.

## 10.4 Remarks

According to the projection postulate (also known as the measurement postulate or the collapse postulate), the wave function changes dramatically in a measurement. The change is known as the *reduction of the wave packet* or the *collapse of the wave function*.

For example, in a spin-$z$ (or $\sigma_3$-) measurement, the wave function before the measurement is an arbitrary spinor $(\phi_1, \phi_2) \in S$ with $|\phi_1|^2 + |\phi_2|^2 = 1$ (assuming Eq. (9.13) and ignoring the space dependence). With probability $|\phi_1|^2$, we obtain outcome "up" and the collapsed spinor $(\phi_1/|\phi_1|, 0)$ after the measurement. The term $\phi_1/|\phi_1|$ is just the phase of $\phi_1$. With probability $|\phi_2|^2$, we obtain "down" and the collapsed spinor $(0, \phi_2/|\phi_2|)$.

With the projection postulate, the formalism provides a prediction of probabilities for any sequence of measurements. If we prepare the initial wave function $\psi_0$ and make a measurement of $A_1$ at time $t_1$ then the Schrödinger equation determines what $\psi_{t_1-}$ is, the general Born rule (8.45) determines the probabilities of the outcome $\alpha_1$, and the projection postulate the wave function after the measurement. The latter is the initial wave function for the Schrödinger equation, which governs the evolution of $\psi$ until the time $t_2$ at which the second measurement, of observable $A_2$, occurs. The probability distribution of the outcome $\alpha_2$ is given by the Born rule again and depends on $\alpha_1$ because the initial wave function in the Schrödinger equation, $\psi_{t_1+}$, did. And so on. This scheme is the *quantum formalism*. Note that the observer can choose $t_2$ and $A_2$ after the first measurement and thus make this choice depend on the first outcome $\alpha_1$.

The projection postulate implies that if we make another measurement of $A$ right after the first one, we will with probability 1 obtain the same outcome $\alpha$.

For a position measurement, the projection postulate implies that the wave function collapses to a delta function. This is not realistic, it is over-idealized. A delta function is not a square-integrable function, and it contains in a sense an infinite amount of energy. More realistically, a position measurement has a finite inaccuracy $\varepsilon$ and could be expected to collapse the wave function to one of width $\varepsilon$, such as

$$\psi_{t+}(\boldsymbol{x}) = Ce^{-\frac{(\boldsymbol{x}-\boldsymbol{\alpha})^2}{4\varepsilon^2}}\psi_{t-}(\boldsymbol{x})\,. \tag{10.17}$$

However, this operator (multiplication by a Gaussian) is not a projection because its spectrum is more than just 0 and 1.

Another simple model of position measurement, still highly idealized but less so than collapse to $\delta(\boldsymbol{x}-\boldsymbol{\alpha})$, considers a region $B \subset \mathbb{R}^3$ and assumes that a detector either finds the particle in $B$ or not. The corresponding observable is $A = P_B$ as defined in (10.16), and the probability of outcome 1 is

$$\int_B d^3\boldsymbol{x}\, |\psi_{t-}(\boldsymbol{x})|^2\,. \tag{10.18}$$

In case of outcome 1, $\psi_{t-}$ collapses to

$$\psi_{t+} = \frac{P_B\psi_{t-}}{\|P_B\psi_{t-}\|}\,. \tag{10.19}$$

You may feel a sense of paradox about the two different laws for how $\psi$ changes with time: the unitary Schrödinger evolution and the collapse rule. Already at first sight, the two seem rather incompatible: the former is deterministic, the latter stochastic; the former is continuous, the latter not; the former is linear, the latter not. It seems strange that time evolution is governed not by a single law but by two. And even stranger that the criterion for when the collapse rule takes over is something as vague as an observer making a measurement. Upon scrutiny, the sense of paradox will persist and even deepen in the form of what is known as the *measurement problem of quantum mechanics.*

# 11 The Measurement Problem

## 11.1 What the Problem Is

This is a problem about orthodox quantum mechanics. It is solved in Bohmian mechanics and several other theories. Because of this problem, some regard the orthodox view as incoherent when it comes to analyzing the process of measurement.

Consider a "quantum measurement of the observable $A$." Realistically, there are only finitely many possible outcomes, so $A$ should have finite spectrum. Consider the system formed by the object together with the apparatus. Since the apparatus consists of electrons and quarks, too, it should itself be governed by quantum mechanics. (That is reductionism at work.) So I write $\Psi$ for the wave function of the system (object and apparatus). Suppose for simplicity that the system is isolated (i.e., there is no interaction with the rest of the universe), so $\Psi$ evolves according to the Schrödinger equation during the experiment (recall Exercise 13 of Assignment 3), which begins (say) at $t_1$ and ends at $t_2$. It is reasonable to assume that

$$\Psi(t_1) = \psi(t_1) \otimes \phi \tag{11.1}$$

with $\psi = \psi(t_1)$ the wave function of the object before the experiment and $\phi$ a wave function representing a "ready" state of the apparatus. By the spectral theorem, $\psi$ can be written as a linear combination (superposition) of eigenfunctions of $A$,

$$\psi = \sum_\alpha c_\alpha \psi_\alpha \quad \text{with} \quad A\psi_\alpha = \alpha\psi_\alpha \text{ and } \|\psi_\alpha\| = 1\,. \tag{11.2}$$

If the object's wave function is an eigenfunction $\psi_\alpha$, then, by Born's rule (8.45), the outcome is certain to be $\alpha$. Set $\Psi_\alpha(t_1) = \psi_\alpha \otimes \phi$. Then $\Psi_\alpha(t_2)$ must represent a state in which the apparatus displays the outcome $\alpha$.

Now consider again a general $\psi$ as in Eq. (11.2). Since the Schrödinger equation is linear, the wave function of object and apparatus together at $t_2$ is

$$\Psi(t_2) = \sum_\alpha c_\alpha \Psi_\alpha(t_2)\,, \tag{11.3}$$

a superposition of states corresponding to different outcomes—and not a random state corresponding to a unique outcome, as one might have expected from the projection postulate. This is the measurement problem. The upshot is that there is a conflict between the following assumptions:

- In each run of the experiment, there is a unique outcome.
- The wave function is a complete description of a system's physical state.
- The evolution of the wave function of an isolated system is always given by the Schrödinger equation.

Thus, we have to drop one of these assumptions. The first is dropped in the many-worlds picture, in which all outcomes are realized, albeit in parallel worlds. If we drop the second, we opt for additional variables as in Bohmian mechanics, where the state at time $t$ is described by the pair $(Q_t, \psi_t)$. If we drop the third, we opt for replacing the Schrödinger equation by a non-linear evolution (as in the GRW = Ghirardi–Rimini–Weber approach). Of course, a theory might also drop several of these assumptions. Orthodox quantum mechanics insists on all three assumptions, and that is why it has a problem.

We took for granted that the system was isolated and had a wave function. We may wonder whether that was asking too much. However, we could just take the system to consist of the entire universe, so it is disentangled and isolated for sure. More basically, if we cannot solve the measurement problem for an isolated system with a wave function then we have no chance of solving it for a system entangled with outside particles.

## 11.2 How Bohmian Mechanics Solves the Problem

Since it is assumed that the Schrödinger equation is valid for a closed system, the after-measurement wave function of object and apparatus together is

$$\Psi = \sum_\alpha c_\alpha \Psi_\alpha \,. \tag{11.4}$$

Since the $\Psi_\alpha$ have disjoint supports in the configuration space (of object and apparatus together), and since the particle configuration $Q$ has distribution $|\Psi|^2$, the probability that $Q$ lies in the support of $\Psi_\alpha$ is

$$\mathbb{P}\big(Q \in \mathrm{support}(\Psi_\alpha)\big) = \int\limits_{\mathrm{support}(\Psi_\alpha)} d^{3N}q\, |\Psi(q)|^2 = \int\limits_{\mathbb{R}^{3N}} d^{3N}q\, |c_\alpha \Psi_\alpha(q)|^2 = |c_\alpha|^2 \,, \tag{11.5}$$

which agrees with the prediction of the quantum formalism for the probability of the outcome $\alpha$. And indeed, when $Q \in \mathrm{support}(\Psi_\alpha)$, then the particle positions (including the particles of both the object and the apparatus!) are such that the pointer of the apparatus points to the value $\alpha$. Thus, the way out of the measurement problem is that although the wave function is a superposition of terms corresponding to different outcomes, the actual particle positions define the actual outcome.

As a consequence of the above consideration, we also see that the predictions of Bohmian mechanics for the probabilities of the outcomes of experiments agree with those of standard quantum mechanics. In particular, there is no experiment that could empirically distinguish between Bohmian mechanics and standard quantum mechanics, while there are (in principle) experiments that distinguish the two from a GRW world.

If Bohmian mechanics and standard quantum mechanics agree about all probabilities, then where do we find the collapse of the wave function in Bohmian mechanics? There are two answers, depending on which wave function we are talking about. The first answer is, if the $\Psi_\alpha$ are macroscopically different then they will never overlap again (until the time when the universe reaches thermal equilibrium, perhaps in $10^{10^{10}}$ years); this fact is called *decoherence*. If $Q$ lies in the support of one among several disjoint packets then only the packet containing $Q$ is relevant, by Bohm's law of motion (6.1), to determining $dQ/dt$. Thus, as long as the packets stay disjoint, only the packet containing $Q$ is relevant to the trajectories of the particles, and all other packets could be replaced by zero without affecting the trajectories. That is why we can replace $\Psi$ by $c_\alpha \Psi_\alpha$, with $\alpha$ the actual outcome. Furthermore, the factor $c_\alpha$ cancels out in Bohm's law of motion (6.1) and thus can be dropped as well.

The second answer is, the quantum formalism does not, in fact, talk about the wave function $\Psi$ of object and apparatus but about the wave function $\psi$ of the object alone. This leads us to the question what is meant by the wave function of a subsystem. If

$$\Psi(x, y) = \psi(x)\phi(y) \tag{11.6}$$

then it is appropriate to call $\psi$ the wave function of the $x$-system, but in general $\Psi$ does not factorize as in (11.6). In Bohmian mechanics, a natural general definition for the wave function of a subsystem is the *conditional wave function*

$$\psi(x) = \mathcal{N}\, \Psi(x, Y)\,, \tag{11.7}$$

where $Y$ is the actual configuration of the $y$-system (while $x$ is not the actual configuration $X$ but any configuration of the $x$-system) and

$$\mathcal{N} = \left( \int |\Psi(x, Y)|^2 dx \right)^{-1/2} \tag{11.8}$$

is the normalizing factor. The conditional wave function does not, in general, evolve according to a Schrödinger equation, but in a complicated way depending on $\Psi$, $Y$, and $X$. There are special situations in which the conditional wave function does evolve according to a Schrödinger equation, in particular when the $x$-system and the $y$-system do not interact *and* the wave packet in $\Psi$ containing $Q = (X, Y)$ is of a product form such as (11.6). Indeed, this is the case for the object *before*, but not *during* the measurement; as a consequence, the wave function of the object (i.e., its conditional wave function) evolves according to the Schrödinger equation before, but not during the measurement—in agreement with the quantum formalism. To determine the conditional wave function after the quantum measurement, suppose that $\Psi_\alpha$ is of the form

$$\Psi_\alpha = \psi_\alpha \otimes \phi_\alpha \tag{11.9}$$

with $\phi_\alpha$ a wave function of the apparatus with the pointer pointing to the value $\alpha$. Let $\alpha$ be the actual outcome, i.e., $Q \in \mathrm{support}(\Psi_\alpha)$. Then $Y \in \mathrm{support}(\phi_\alpha)$ and the conditional wave function is indeed

$$\psi = \psi_\alpha\,. \tag{11.10}$$

## 11.3 Schrödinger's Cat

Often referred to in the literature, this is Schrödinger's[13] 1935 formulation of the measurement problem:

> One can even set up quite ridiculous cases. A cat is penned up in a steel chamber, along with the following diabolical device (which must be secured against direct interference by the cat): in a Geiger counter there is a tiny bit of radioactive substance, so small, that *perhaps* in the course of one hour one of the atoms decays, but also, with equal probability, perhaps none; if it happens, the counter tube discharges and through a relay releases a hammer which shatters a small flask of hydrocyanic acid. If one has left this entire system to itself for an hour, one would say that the cat still lives *if* meanwhile no atom has decayed. The first atomic decay would have poisoned it. The $\psi$-function of the entire system would express this by having in it the living and dead cat (pardon the expression) mixed or smeared out in equal parts.
>
> It is typical of these cases that an indeterminacy originally restricted to the atomic domain becomes transformed into macroscopic indeterminacy, which can then be *resolved* by direct observation. That prevents us from so naively accepting as valid a "blurred model" for representing reality. In itself it would not embody anything unclear or contradictory. There is a difference between a shaky or out-of-focus photograph and a snapshot of clouds and fog banks.

## 11.4 Positivism and Realism

*Positivism* is the view that a statement which cannot be tested in experiment is meaningless or unscientific. For example, the statement in Bohmian mechanics that an electron went through the upper slit of a double-slit if and only if it arrived in the upper half of the screen, cannot be tested in experiment. After all, if you try to check which slit the electron went through by detecting every electron at the slit then the statement is no longer true in Bohmian mechanics (and in fact, no correlation with the location of arrival is found). So a positivist thinks that Bohmian mechanics is unscientific. Good statements for a positivist are *operational statements*, i.e., statements of the form "if we set up an experiment in this way, the outcome has such-and-such a probability distribution." Positivists think that the quantum formalism (thought of as a summary of all true operational statements of quantum mechanics) is the only scientific formulation of quantum mechanics. They also tend to think that $\psi$ is the complete description of a

[13]From E. Schrödinger: Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften* **23**: 807–812, 823–828, 844–849 (1935). English translation by J. D. Trimmer: The Present Situation in Quantum Mechanics, *Proceedings of the American Philosophical Society* **124**: 323–338 (1980). Reprinted in J. A. Wheeler, W. H. Zurek (ed.s): *Quantum Theory and Measurement*, Princeton University Press (1983), pages 152–167.

system, as it is the only information about the system that can be found experimentally without disturbing $\psi$. They tend not to take the measurement problem seriously.

*Realism* is the view that a fundamental physical theory is meaningless unless it provides a coherent story of what happens. Bohmian mechanics, GRW theory, and many-worlds are examples of realist theories. For a realist, the quantum formalism by itself does not qualify as a fundamental physical theory. The story provided by Bohmian mechanics, for example, is that particles have trajectories, that there is a physical object that is mathematically represented by the wave function, and that the two evolve according to certain equations. For a realist, the measurement problem is serious and can only be solved by denying one of the 3 conflicting premises.

# 12 The GRW Theory

Bohmian mechanics is not the only possible explanation of quantum mechanics. Another one is provided by the GRW theory, named after GianCarlo Ghirardi, Alberto Rimini, and Tullio Weber, who proposed it in 1986. A similar theory, CSL (for *continuous spontaneous localization*), was proposed by Philip Pearle in 1989. In both theories, $\Psi_t$ does not evolve according to the Schrödinger equation, but according to a modified evolution law. This evolution law is *stochastic*, as opposed to deterministic. That is, for any fixed $\Psi_0$, it is random what $\Psi_t$ is, and the theory provides a probability distribution over Hilbert space. A family of random variables $X_t$, with one variable for every time $t$, is called a *stochastic process*. Thus, the family $(\Psi_t)_{t>0}$ is a stochastic process in Hilbert space. We leave CSL aside and focus on the *GRW process*. In it, periods governed by the Schrödinger equation are interrupted by random jumps. Such a jump occurs, within any infinitesimal time interval $dt$, with probability $\lambda\, dt$, where $\lambda$ is a constant called the *jump rate*. Let us call the random jump times $T_1, T_2, \ldots$; the sequence $T_1, T_2, \ldots$ is known as the *Poisson process with rate* $\lambda$; it has widespread applications in probability theory. Let us have a closer look.

## 12.1 The Poisson Process

Think of $T_1, T_2, \ldots$ as the times at which a certain type of random event occurs; standard examples include the times when an earthquake (of a certain strength) occurs, or when the phone rings, or when the price of a certain share falls below a certain value. We take for granted that the ordering is chosen such that $0 < T_1 < T_2 < \ldots$.

Let us figure out the probability density function of $T_1$. The probability that $T_1$ occurs between 0 and $dt$ is $\lambda\, dt$. Thus, the probability that it does not occur is $1 - \lambda\, dt$. Suppose that it did not occur between 0 and $dt$. Then the probability that it doesn't occur between $dt$ and $2\, dt$ is again $1 - \lambda\, dt$. Thus, the total probability that no event occurs between 0 and $2\, dt$ is $(1-\lambda\, dt)^2$. Proceeding in the same way, the total probability that no event occurs between 0 and $n\, dt$ is $(1 - \lambda\, dt)^n$. Thus, the total probability that no event occurs between 0 and $t$, $\mathbb{P}(T_1 > t)$, can be approximated by setting $dt = t/n$ and letting $n \to \infty$. That is,

$$\mathbb{P}(T_1 > t) = \lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t}\,. \tag{12.1}$$

Let us write $\rho(t)$ for the probability density function of $T_1$. By definition,

$$\rho(t)\, dt = \mathbb{P}(t < T_1 < t + dt)\,. \tag{12.2}$$

To compute this quantity, we reason as follows. If $T_1$ has not occured until $t$, then the probability that it will occur within the next $dt$ is $\lambda\, dt$. Thus, (12.2) differs from (12.1) by a factor $\lambda\, dt$, or, as the factor $dt$ cancels out,

$$\rho(t) = 1_{t>0} e^{-\lambda t} \lambda\,, \tag{12.3}$$

where the expression $1_C$ is 1 whenever the condition $C$ is satisfied, and 0 otherwise. The distribution (12.3) is known as the *exponential distribution with parameter* $\lambda$, $\text{Exp}(\lambda)$. We have thus found that *the waiting time for the first event has distribution* $\text{Exp}(\lambda)$.

After $T_1$, the next $dt$ has again probability $\lambda\, dt$ for the next event to occur. The above reasoning can be repeated, with the upshot that *the waiting time* $T_2 - T_1$ *for the next event has distribution* $\text{Exp}(\lambda)$ *and is independent of what happened up to time* $T_1$. The same applies to the other waiting times $T_{n+1} - T_n$. In fact, at any time $t_0$ the waiting time until the next event has distribution $\text{Exp}(\lambda)$.

The exponential distribution has expectation value

$$\int_0^\infty t\, \rho(t)\, dt = \frac{1}{\lambda}\,. \tag{12.4}$$

This fact is very plausible if you think of it this way: If in every second the probability of an earthquake is, say, $10^{-8}$, then you would guess that an earthquake occurs on average every $10^8$ seconds. The constant $\lambda$, whose dimension is 1/time, is thus the average frequency of the earthquakes (or whichever events).

Another way of representing the Poisson process is by means of the random variables

$$X_t = \#\{i \in \mathbb{N} : T_i < t\}\,, \tag{12.5}$$

the number of earthquakes up to time $t$.

**Theorem 12.1.** *If the earthquakes in Australia are governed by a Poisson process with rate* $\lambda_1$ *and the earthquakes in Africa are governed by a Poisson process with rate* $\lambda_2$*, and the earthquakes in the two places are independent of each other, then the earthquakes in Africa and Australia together are governed by a Poisson process with rate* $\lambda_1 + \lambda_2$*.*

**Theorem 12.2.** *If we choose* $n$ *points at random in the interval* $[0, n/\lambda]$*, independently with uniform distribution, then the joint distribution of these points converges, as* $n \to \infty$*, to the Poisson process with parameter* $\lambda$*.*

## 12.2 Definition of the GRW Process

Now let us get back to the definition of the GRW process. To begin with, set the particle number $N = 1$, so that $\Psi_t : \mathbb{R}^3 \to \mathbb{C}$. The random events are, instead of earthquakes, spontaneous collapses of the wave function. That is, suppose that the random variables $T_1, T_2, T_3, \ldots$, are governed by a Poisson process with parameter $\lambda$; suppose that between $T_{k-1}$ and $T_k$, the wave function $\Psi_t$ evolves according to the Schrödinger equation (where $T_0 = 0$); at every $T_k$, the wave function changes discontinuously ("collapses") as if an outside observer made an unsharp position measurement with inaccuracy $\sigma > 0$. I will give the formula below.

The constants $\lambda$ and $\sigma$ are thought of as new constants of nature, for which GRW suggested the values

$$\lambda \approx 10^{-16}\,\text{sec}^{-1}\,, \quad \sigma \approx 10^{-7}\,\text{m}\,. \tag{12.6}$$

Alternatively, Steven Adler suggested

$$\lambda \approx 3 \times 10^{-8}\,\mathrm{sec}^{-1}\,, \quad \sigma \approx 10^{-6}\,\mathrm{m}\,. \tag{12.7}$$

This completes the definition of the GRW process for $N = 1$.

Now consider arbitrary $N \in \mathbb{N}$, and let $\Psi_0$ be (what is normally called) an $N$-particle wave function $\Psi_0 = \Psi_0(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$. Consider $N$ independent Poisson processes with rate $\lambda$, $T_{i,1}, T_{i,2}, \ldots$ for every $i \in \{1, \ldots, N\}$. Let $T_1$ be the smallest of all these random times, $T_2$ the second smallest etc., and let $I_1$ be the index associated with $T_1$ and $I_2$ the index associated with $T_2$ etc. Equivalently, $T_1, T_2, \ldots$ is a Poisson process with rate $N\lambda$, and along with every $T_k$ we choose a random index $I_k$ from $\{1, \ldots, N\}$ with uniform distribution (i.e., each $i$ has probability $1/N$), independently of each other and of the $T_k$. Equivalently, a collapse with index $i$ occurs with rate $\lambda$ for each $i \in \{1, \ldots, N\}$. Between $T_{k-1}$ and $T_k$, $\Psi_t$ evolves according to the Schrödinger equation. At $T_k$, $\Psi$ changes as if an observer outside of the system[14] made an unsharp position measurement with inaccuracy $\sigma$ on particle number $I_k$.

## 12.3 Definition of the GRW Process in Formulas

Let us begin with $N = 1$.

$$\Psi_{T_k+} = \frac{C(\boldsymbol{X}_k)\Psi_{T_k-}}{\|C(\boldsymbol{X}_k)\Psi_{T_k-}\|}\,, \tag{12.8}$$

where the *collapse operator* $C(\boldsymbol{X})$ is a multiplication operator multiplying by the square root of a 3-d Gaussian function centered at $\boldsymbol{X}$:

$$C(\boldsymbol{X})\Psi(\boldsymbol{x}) = \sqrt{g_{\boldsymbol{X},\sigma}(\boldsymbol{x})}\,\Psi(\boldsymbol{x}) \tag{12.9}$$

with

$$g_{\boldsymbol{X},\sigma}(\boldsymbol{x}) = \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-(\boldsymbol{X}-\boldsymbol{x})^2/2\sigma^2}\,. \tag{12.10}$$

The point $\boldsymbol{X}_k \in \mathbb{R}^3$ is chosen at random with probability density

$$\rho(\boldsymbol{X}_k = \boldsymbol{y}|T_1, \ldots, T_k, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k-1}) = \|C(\boldsymbol{y})\Psi_{T_k-}\|^2\,, \tag{12.11}$$

where $\rho(\cdots|\cdots)$ means the probability density, *given* the values of $T_1, \ldots, T_k, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k-1}$. The right hand side of (12.11) is indeed a probability density because it is nonnegative and

$$\int d^3\boldsymbol{y}\,\rho(\boldsymbol{X}_k = \boldsymbol{y}|\cdots) = \int d^3\boldsymbol{y}\,\|C(\boldsymbol{y})\Psi\|^2 = \int d^3\boldsymbol{y}\int d^3\boldsymbol{x}\,|C(\boldsymbol{y})\Psi(\boldsymbol{x})|^2 = \tag{12.12}$$

$$= \int d^3\boldsymbol{x}\int d^3\boldsymbol{y}\,g_{\boldsymbol{y},\sigma}(\boldsymbol{x})\,|\Psi(\boldsymbol{x})|^2 = \int d^3\boldsymbol{x}\,|\Psi(\boldsymbol{x})|^2 = 1\,. \tag{12.13}$$

[14] Or rather, outside of the universe, as the idea is that the entire universe is governed by GRW theory.

For arbitrary $N \in \mathbb{N}$ and $\Psi_t = \Psi_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$,

$$\Psi_{T_k+} = \frac{C_{I_k}(\boldsymbol{X}_k)\Psi_{T_k-}}{\|C_{I_k}(\boldsymbol{X}_k)\Psi_{T_k-}\|} \tag{12.14}$$

where the collapse operator $C_I(\boldsymbol{X})$ is the following multiplication operator:

$$C_I(\boldsymbol{X})\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sqrt{g_{\boldsymbol{X},\sigma}(\boldsymbol{x}_I)}\, \Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\,. \tag{12.15}$$

The random point $\boldsymbol{X}_k$ is chosen at random with probability density

$$\rho(\boldsymbol{X}_k = \boldsymbol{y}|T_1, \ldots, T_k, I_1, \ldots, I_k, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k-1}) = \|C_{I_k}(\boldsymbol{y})\Psi_{T_k-}\|^2\,. \tag{12.16}$$

This completes the definition of the GRW process. But not yet the definition of the GRW theory.

## 12.4 Primitive Ontology

There is a further law in GRW theory, concerning matter in 3-space. There are two different versions of this law and, accordingly, two different versions of the GRW theory, abbreviated as GRWm (m for *matter density ontology*) and GRWf (f for *flash ontology*). For comparison, in Bohmian mechanics the matter in 3-space consists of the particles (with trajectories).

In GRWm it is a law that, at every time $t$, matter is continuously distributed in space with density function $m(\boldsymbol{x}, t)$ for every location $\boldsymbol{x} \in \mathbb{R}^3$, given by

$$m(\boldsymbol{x}, t) = \sum_{i=1}^{N} m_i \int\limits_{\mathbb{R}^{3N}} d^3\boldsymbol{x}_1 \cdots d^3\boldsymbol{x}_N\, \delta^3(\boldsymbol{x}_i - \boldsymbol{x}) \left|\psi_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\right|^2. \tag{12.17}$$

In words, one starts with the $|\psi|^2$–distribution in configuration space $\mathbb{R}^{3N}$, then obtains the marginal distribution of the $i$-th degree of freedom $\boldsymbol{x}_i \in \mathbb{R}^3$ by integrating out all other variables $\boldsymbol{x}_j$, $j \neq i$, multiplies by the mass associated with $\boldsymbol{x}_i$, and sums over $i$.

In GRWf it is a law that matter consists of material points in space-time called flashes. That is, matter is neither made of particles following world lines, nor of a continuous distribution of matter such as in GRWm, but rather of discrete points in space-time. According to GRWf, the space-time locations of the flashes can be read off from the history of the wave function: every flash corresponds to one of the spontaneous collapses of the wave function, and its space-time location is just the space-time location of that collapse. The flashes form the set

$$F = \{(\boldsymbol{X}_1, T_1, I_1), \ldots, (\boldsymbol{X}_k, T_k, I_k), \ldots\}\,. \tag{12.18}$$

Note that if the number $N$ of the degrees of freedom in the wave function is large, as in the case of a macroscopic object, the number of flashes is also large (if $\lambda = 10^{-16}$

$s^{-1}$ and $N = 10^{23}$, we obtain $10^7$ flashes per second). Therefore, for a reasonable choice of the parameters of the GRWf theory, a cubic centimeter of solid matter contains more than $10^7$ flashes per second. That is to say that large numbers of flashes can form macroscopic shapes, such as tables and chairs. "A piece of matter then is a galaxy of [flashes]." (Bell, page 205) That is how we find an image of our world in GRWf.

A few remarks. The $m$ function of GRWm and the flashes of GRWf are called the *primitive ontology* of the theory. Ontology means what exists according to a theory; for example, in Bohmian mechanics $\psi$ and $Q$, in GRWm $\psi$ and $m$, in GRWf $\psi$ and $F$. The "primitive" ontology is the part of the ontology representing matter in 3-d space (or 4-d space-time): $Q$ in Bohmian mechanics, $m$ in GRWm, and $F$ in GRWf.

It may be seem that a continuous distribution of matter should conflict with the evidence for the existence of atoms, electrons and quarks, and should thus make wrong predictions. We will see below why that is not the case—why GRWm makes nearly the same predictions as the quantum formalism.

## 12.5 The GRW Solution to the Measurement Problem

We will now look at why the GRW process succeeds in solving the measurement problem, specifically in collapsing macroscopic (but not microscopic) superpositions, and why the deviations from quantum mechanics are in a sense small.

First, the collapses are supposed to occur *spontaneously*, just at random, without the intervention of an outside observer, indeed without any physical cause described by the theory; GRW is a stochastic theory. Let us look at the number of collapses. The average waiting time between two collapses is $1/N\lambda$. For a single particle, $N = 1$, this time is $\approx 10^{16}\,\text{sec} \approx 10^8\,\text{years}$. That is, for a single particle the wave function collapses only every 100 million years. So we should not expect to see any of these spontaneous collapses when doing an experiment with a single particle, or even with hundreds of particles. If, however, we consider a macroscopic system, consisting perhaps of $10^{23}$ particles, then the average waiting time is $10^{-7}\,\text{sec}$, so we have a rather dense shower of collapses.

A collapse amounts to multiplication by a Gaussian with width $\sigma \approx 10^{-7}\,\text{m}$, which is large on the atomic scale (recall that the size of an atom is about one Angstrom $= 10^{-10}\text{m}$) but small on the macroscopic scale. So, if an electron is in a superposition of being in Paris and being in Tokyo, and if the center $\boldsymbol{X}$ of the collapse lies in Paris, then the collapse operator has the effect of damping the wave function in Tokyo (which is roughly $10^7\,\text{m}$ away from Paris) by a factor of $\exp(10^{28})$. Thus, after the collapse, the wave function in Tokyo is very near zero. On the other hand, if a collapse hits an electron in a bound state in an atom, the collapse will not much affect the electron's wave function.

Let us examine the probability distribution $\rho(\boldsymbol{X} = \boldsymbol{y})$ of the center $\boldsymbol{X}$ of a collapse. For a one-particle wave function $\Psi$, it is essentially $|\Psi|^2$; more precisely, it is the quantum distribution $|\Psi|^2$ convolved with $g_\sigma$, that is, smeared out (or blurred, or coarse-grained) over a distance $\sigma$ that is smaller than the macroscopic scale. For an $N$-particle wave function $\Psi$, $\rho(\boldsymbol{X} = \boldsymbol{y})$ is essentially the *marginal* of $|\Psi|^2$ connected to the $\boldsymbol{x}_I$-variable, i.e., the distribution on 3-space obtained from the $|\Psi|^2$ distribution on $3N$-space by integrating out $3N - 3$ variables. (More precisely, smeared over width $\sigma$.) Thus, again, on the macroscopic scale, the distribution of $\boldsymbol{X}$ is the same as the quantum mechanical probability distribution for the position of the $I$-th particle.

A wave function like the one we encountered in the measurement problem,

$$\Psi = \sum_\alpha c_\alpha \Psi_\alpha \,, \tag{12.19}$$

where $\Psi_\alpha$ is a wave function corresponding to the pointer pointing to the value $\alpha$, would behave in the following way. Assuming the pointer contains $10^{23}$ particles, then every $10^{-7}$ sec a collapse would occur connected to one of the pointer particles. Since $\Psi_\alpha$ is concentrated in a region in configuration space where all of the pointer particles are at some location $\boldsymbol{y}_\alpha$, and assuming that the $\boldsymbol{y}_\alpha$ are sufficiently distant for different values of $\alpha$ (namely much more than $\sigma$), a single collapse connected to any of the pointer particles will suffice for essentially removing all contributions $\Psi_\alpha$ except one. Indeed, suppose the collapse is connected to the particle $\boldsymbol{x}_i$, which is one of the pointer particles. Then the random center $\boldsymbol{X}$ of the collapse will be distributed according to a coarse-grained version of the $i$-th marginal of $|\Psi|^2$; since the separation between the $\boldsymbol{y}_\alpha$ is greater than $\sigma$, we can neglect the coarse graining, and we can just take the $i$-th marginal of the $|\Psi|^2$ distribution. Thus, $\boldsymbol{X}$ will be close to one of the $\boldsymbol{y}_\alpha$, and the probability that $\boldsymbol{X}$ is close to $\boldsymbol{y}_{\alpha'}$ is $|c_\alpha|^2$. Then, the multiplication by a Gaussian centered at $\boldsymbol{X}$ will shrink all other packets $\Psi_\alpha$ by big factors, of the order $\exp(-(\boldsymbol{y}_\alpha - \boldsymbol{y}_{\alpha'})^2/2\sigma^2)$, effectively collapsing them away.

Thus, within a fraction of a second, a superposition such as (12.19) would decay into one of the packets $\Psi_\alpha$ (times a normalization factor), and indeed into $\Psi_{\alpha'}$ with probability $|c_\alpha|^2$, the same probability as attributed by quantum mechanics to the outcome $\alpha'$.

Let us make explicit how GRW succeeded in setting up the laws in such a way that they are effectively different laws for microscopic and macroscopic objects: (i) We realize that a few collapses (or even a single collapse) acting on a few (or one) of the pointer particles will collapse the entire wave function $\Psi$ of object and apparatus together to essentially just one of the contributions $\Psi_\alpha$. (ii) The frequency of the collapses is proportional to the number of particles (which serves as a quantitative measure of "being macroscopic"). (iii) We can't ensure that microscopic systems experience *no collapses at all*, but we can ensure the collapses are *very infrequent*. (iv) We can't ensure that macroscopic superpositions such as $\Psi = \sum c_\alpha \Psi_\alpha$ collapse *immediately*, but we can ensure they collapse *within a fraction of a second*.

## 12.6 Empirical Tests

I have pointed out why GRW theory leads to essentially the same probabilities as prescribed by the quantum formalism. Yet, it is obvious that there are some experiments

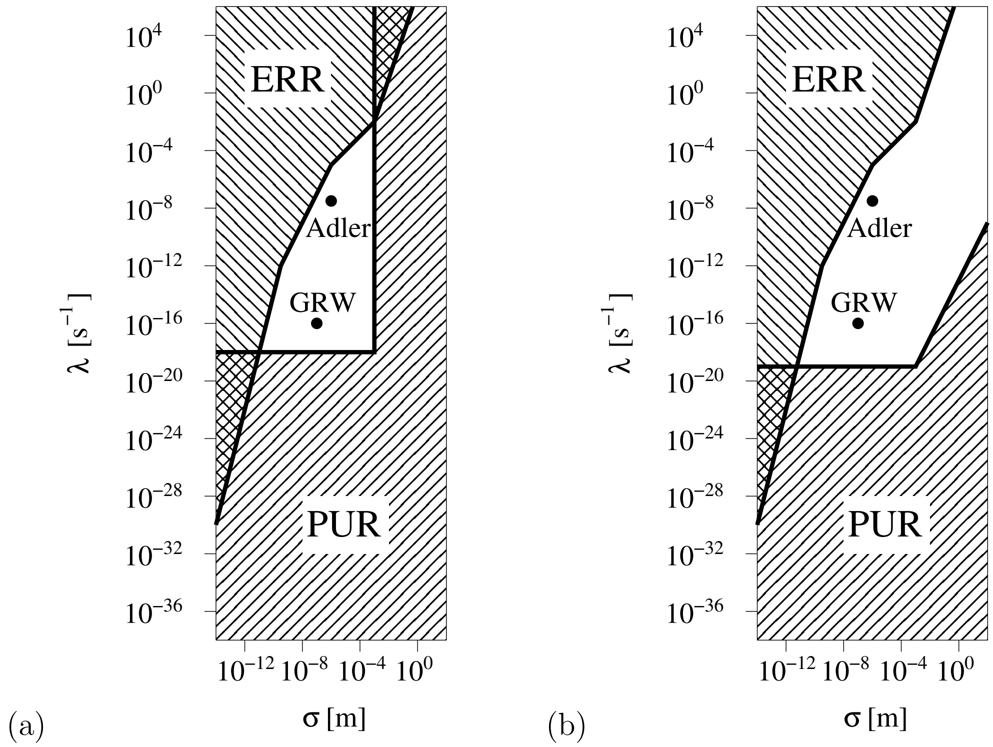


Figure 3: Parameter diagram (log-log-scale) of the GRW theory with the primitive ontology given by (a) flashes, (b) the matter density function. ERR = empirically refuted region as of 2012 (equal in (a) and (b)), PUR = philosophically unsatisfactory region. GRW's and Adler's choice of parameters are marked. Figure taken from W. Feldmann and R. Tumulka: Parameter Diagrams of the GRW and CSL Theories of Wave Function Collapse. *Journal of Physics A: Mathematical and Theoretical* **45**: 065304 (2012) http://arxiv.org/abs/1109.6579

for which GRW theory predicts different outcomes than the quantum formalism. Here is an example. GRW theory predicts that if we keep a particle isolated it will spontaneously collapse after about 100 million years, and quantum mechanics predicts it will not collapse. So let's take $10^4$ electrons, for each of them prepare its wave function to be a superposition of a packet in Paris and a packet in Tokyo; let's keep each electron isolated for 100 million years; according to GRW, a fraction of

$$\int_0^{1/\lambda} \lambda\, e^{-\lambda t}\, dt = \int_0^1 e^{-s}\, ds = 1 - e^{-1} = 63.2\% \tag{12.20}$$

of the $10^4$ wave functions will have collapsed; according to quantum mechanics, none will have collapsed; now let's bring the packets from Paris and Tokyo together, let them overlap and observe the interference pattern; according to quantum mechanics, we should observe a clear interference patterns; if all of the wave functions had collapsed, we should observe no interference pattern at all; according to GRW, we should observe only a faint interference pattern, damped (relative to the quantum prediction) by a factor of $e$. Ten thousand points should be enough to decide whether the damping factor is there or not. This example illustrates two things: that in principle GRW makes different predictions, and that in practice these differences may be difficult to observe (because of the need to wait for 100 million years, and because of the difficulty with keeping the electrons isolated for a long time, in particular avoiding decoherence).

Another testable consequence of the GRW process is *universal warming*. Since the GRW collapse usually makes wave packets narrower, their Fourier transforms (momentum representation) become wider, by the Heisenberg uncertainty relation. As a tendency, this leads to a long-run increase in energy. This effect amounts to a spontaneous warming at a rate of the order of $10^{-15}$ K per year.

No empirical test of GRW theory against the quantum formalism can presently be carried out, but experimental techniques are progressing; see Figure 3. Adler's parameters have in the meantime been empirically refuted as a byproduct of the LIGO experiment that detects gravitational waves. A test of GRW's parameters seems feasible using a planned interferometer on a satellite in outer space. Interferometers are disturbed by the presence of air, temperatures far from absolute zero, vibrations of the apparatus, and the presence of gravity; that is why being in outer space is an advantage for an interferometer and allows for heavier objects shot through the double slit and longer flight times. Such an interferometer is being considered by the European Space Agency ESA and may be up and running in 2025.

## 12.7 The Need for a Primitive Ontology

Primitive ontology is a subtle philosophical topic.

We may wonder whether, instead of GRWf or GRWm, we could assume that only $\psi$ exists, and no primitive ontology; let us call this view GRW$\emptyset$. To illustrate the difference between GRWf/GRWm and GRW$\emptyset$, let me make up a creation myth (as a metaphorical way of speaking): Suppose God wants to create a universe governed by GRW theory. He creates a wave function $\psi$ of the universe that starts out as a particular $\psi_0$ that he chose and evolves stochastically according to a particular version of the GRW time evolution law. According to GRW$\emptyset$, God is now done. According to GRWf or GRWm, however, a second act of creation is necessary, in which he creates the matter, i.e., either the flashes or continously distributed matter with density $m$, in both cases coupled to $\psi$ by the appropriate laws.

There are several motivations for considering GRW$\emptyset$. First, it seems more parsimonious than GRWm or GRWf. Second, it was part of the motivation behind GRW theory to avoid introducing an ontology in additon to $\psi$. In fact, much of the motivation came from the measurement problem, which requires that we either modify the Schrödinger equation or introduce additional ontology (such as $Q$ in Bohmian mechanics), and GRW theory was intended to choose the first option, not the second.

Furthermore, there is a sense in which GRW$\emptyset$ clearly works: The GRW wave function $\psi_t$ is, at almost all times, concentrated, except for tiny tails, on a set of configurations that are macroscopically equivalent to each other. So we can read off from the post-measurement wave function, e.g., what the actual outcome of a quantum measurement was.

On the other hand, there is a logical gap between saying

$$\text{“}\psi \text{ is the wave function of a live cat”} \tag{12.21}$$

and saying

$$\text{“there is a live cat.”} \tag{12.22}$$

After all, in Bohmian mechanics, (12.22) follows from (12.21) by virtue of a law of the theory, which asserts that the configuration $Q(t)$ is $|\psi_t|^2$ distributed at every time $t$. Thus, Bohmian mechanics suggests that (12.22) would not follow from (12.21) if there was not a law connecting the two by means of the primitive ontology. If that is so, then it does not follow in GRW∅ either. Another indication in this direction is the fact that the region "PUR" in Figure 3 depends on the primitive ontology we consider, GRWf or GRWm.

Other aspects of the question whether GRW∅ is a satisfactory theory have to do with a number of paradoxes that arise in GRW∅ but evaporate in GRWf and GRWm.[15] For the sake of simplicity, I will focus on GRWm and leave aside GRWf.

**Paradox:** Here is a reason one might think that the GRW theory fails to solve the measurement problem. Consider a quantum state like Schrödinger's cat, namely a superposition

$$\psi = c_1\psi_1 + c_2\psi_2 \tag{12.23}$$

of two macroscopically distinct states $\psi_i$ with $\|\psi_1\| = 1 = \|\psi_2\|$, such that both contributions have nonzero coefficients $c_i$. Given that there is a problem—the measurement problem—in the case in which the coefficients are equal, one should also think that there is a problem in the case in which the coefficients are not exactly equal, but roughly of the same size. One might say that the reason there is a problem is that, according to quantum mechanics, there is a superposition whereas according to our intuition there should be a definite state. But then it is hard to see how this problem should go away just because $c_2$ is much smaller than $c_1$. How small would $c_2$ have to be for the problem to disappear? No matter if $c_2 = c_1$ or $c_2 = c_1/100$ or $c_2 = 10^{-100}c_1$, in each case both contributions are there. But the only relevant effect of the GRW process replacing the unitary evolution, as far as Schrödinger's cat is concerned, is to randomly make one of the coefficients much smaller than the other (although it also affects the shape of the suppressed contribution).

**Answer:** From the point of view of GRWm, the reasoning misses the primitive ontology. Yes, the wave function is still a superposition, but the definite facts that our intuition wants can be found in the primitive ontology. The cat is made of $m$, not of

[15] The following discussion is adapted from R. Tumulka: Paradoxes and Primitive Ontology in Collapse Theories of Quantum Mechanics. Pages 139–159 in S. Gao (editor), *Collapse of the Wave Function*, Cambridge University Press (2018) `https://arxiv.org/abs/1102.5767`.

$\psi$. If $\psi$ is close to $|\text{dead}\rangle$, then $m$ equals $m_{|\text{dead}\rangle}$ up to a small perturbation, and that can reasonably be accepted as the $m$ function of a dead cat. While the wave function is a superposition of two packets $\psi_1, \psi_2$ that correspond to *two very different* kinds of (particle) configurations in ordinary QM or Bohmian mechanics, there is only *one* configuration of the matter density $m$—the definite fact that our intuition wants.

**Paradox:** As a variant of the first paradox, one might say that even after the GRW collapses have pushed $|c_1|^2$ near 1 and $|c_2|^2$ near 0 in the state vector (12.23), there is still a positive probability $|c_2|^2$ that if we make a quantum measurement of the macro-state—of whether the cat is dead or alive—we will find the state $\psi_2$, even though the GRW state vector has collapsed to a state vector near $\psi_1$, a state vector that might be taken to indicate that the cat is really dead (assuming $\psi_1 = |\text{dead}\rangle$). Thus, it seems not justified to say that, when $\psi$ is close to $|\text{dead}\rangle$, the cat is really dead.

**Answer:** In GRWm, what we mean when saying that the cat is dead is that the $m$ function looks and behaves like a dead cat. In orthodox QM, one might mean instead that a quantum measurement of the macro-state would yield $|\text{dead}\rangle$ with probability 1. These two meanings are not exactly equivalent in GRWm: that is because, if $m \approx m_{|\text{dead}\rangle}$ (so we should say that the cat is dead) and if $\psi$ is close but not exactly equal to $|\text{dead}\rangle$, then there is still a tiny but non-zero probability that within the next millisecond the collapses occur in such a way that the cat is suddenly alive! But that does not contradict the claim that a millisecond before the cat was dead; it only means that GRWm allows resurrections to occur—with tiny probability! In particular, if we observe the cat after that millisecond, there is a positive probability that we find it alive (simply because it *is* alive) even though before the millisecond it actually was dead.

**Paradox:** Let $\psi_1$ be the state "the marble is inside the box" and $\psi_2$ the state "the marble is outside the box"; these wave functions have disjoint supports $S_1, S_2$ in configuration space (i.e., wherever one is nonzero the other is zero). Let $\psi$ be given by (12.23) with $0 < |c_2|^2 \ll |c_1|^2 < 1$; finally, consider a system of $n$ (non-interacting) marbles at time $t_0$, each with wave function $\psi$, so that the wave function of the system is $\psi^{\otimes n}$. Then for each of the marbles, we would feel entitled to say that it is inside the box, but on the other hand, the probability that all marbles be found inside the box is $|c_1|^{2n}$, which can be made arbitrarily small by making $n$ sufficiently large.

**Answer:** According to the $m$ function, each of the marbles is inside the box at the initial time $t_0$. However, it is known that a superposition like (12.23) of macroscopically distinct states $\psi_i$ will approach under the GRW evolution either a wave function $\psi_1(\infty)$ concentrated in $S_1$ or another $\psi_2(\infty)$ in $S_2$ with probabilities $|c_1|^2$ and $|c_2|^2$, respectively. (Here I am assuming $H = 0$ for simplicity. Although both coefficients will still be nonzero after any finite number of collapses, one of them will tend to zero in the limit $t \to \infty$.) Thus, for large $n$ the wave function will approach one consisting of approximately $n|c_1|^2$ factors $\psi_1(\infty)$ and $n|c_2|^2$ factors $\psi_2(\infty)$, so that ultimately about $n|c_1|^2$ of the marbles will be inside and about $n|c_2|^2$ outside the box—independently of whether anybody observes them or not. The occurrence of some factors $\psi_2(\infty)$ at a later time provides another example of the resurrection-type events mentioned earlier; they are unlikely but do occur, of course, if we make $n$ large enough.

The act of observation plays no role in the argument and can be taken to merely record pre-existing macroscopic facts. To be sure, the physical interaction involved in the act of observation may have an effect on the system, such as speeding up the evolution from $\psi$ towards either $\psi_1(\infty)$ or $\psi_2(\infty)$; but GRWm provides unambiguous facts about the marbles also in the absence of observers.

# 13 The Copenhagen Interpretation

A very influential view, almost synonymous with the orthodox view of quantum mechanics, is the *Copenhagen interpretation* (CI), named after the research group headed by Niels Bohr, who was the director of the Institute for Theoretical Physics at the University of Copenhagen, Denmark. Further famous defenders of this view and members of Bohr's group (temporarily also working in Copenhagen) include Werner Heisenberg, Wolfgang Pauli, and Leon Rosenfeld. Bohr and Einstein were antagonists in a debate about the foundations of quantum mechanics that began around 1925 and continued until Einstein's death in 1955. In Feynman's text you have already seen an exposition of (parts of) the orthodox view. Here is a description of the main elements of CI.

## 13.1 Two Realms

In CI, the world is separated into two realms: macroscopic and microscopic. In the macroscopic realm, there are no superpositions. Pointers always point in definite directions. The macroscopic realm is described by the classical positions and momenta of objects. In the microscopic realm, there are no definite facts. For example, an electron does not have a definite position. The microscopic realm is described by wave functions. One could say that the primitive ontology of CI consists of the macroscopic matter (described by its classical positions and momenta). In CI terminology, the macroscopic realm is called *classical* and the microscopic realm *quantum*.[16] Instead of classical and quantum, Bell called them *speakable* and *unspeakable*. (The macroscopic realm hosts the objects with definite properties, of which one can speak. You may have gotten the sense that Bell is not a supporter of the idea of two separate realms.)

The microscopic realm, when isolated, is governed by the Schrödinger equation. The macroscopic realm, when isolated, is governed by classical mechanics. The two realms interact whenever a measurement is made; then the macro realm records the measurement outcome, and the micro realm undergoes a collapse of the wave function.

I see a number of problems with the concept of two separate realms.

- It is not precisely defined where the border between micro and macro lies. That lies in the nature of the word "macroscopic." Clearly, an atom is micro and a table is macro, but what is the exact number of particles required for an object to be "macroscopic"? The vagueness inherent in the concept of "macroscopic" is unproblematical in Bohmian mechanics, GRW theory, or classical mechanics, but it is problematical here because it is involved in the formulation of the laws of nature. Laws of nature should not be vague.

[16]This is a somewhat unfortunate terminology because the word classical suggests not only definite positions but also particular laws (say, Newton's equation of motion) which may actually not apply. The word quantum is somewhat unfortunate as well because in a reductionist view, all laws (also those governing macroscopic objects) should be consequences of the quantum laws applying to the individual electrons, quarks, etc.

- Likewise, what counts as a measurement and what does not? This ambiguity is unproblematical when we only want to compute the probabilities of outcomes of a given experiment because it will not affect the computed probabilities. But an ambiguity is problematical when it enters the laws of nature.
- The special role played by measurements in the laws according to CI is also implausible and artificial. Even if a precise definition of what counts as a measurement were given, it would not seem believable that during measurement other laws than normal are in place.
- The separation of the two realms, without the formulation of laws that apply to both, is against reductionism. If we think that macro objects are made out of micro objects, then the separation is problematical.

## 13.2 Positivism

CI leans towards positivism. In the words of Werner Heisenberg (1958):

> "We can no longer speak of the behavior of the particle independently of the process of observation."

Feynman (1959) did not like that:

> "Does this mean that my observations become real only when I observe an observer observing something as it happens? This is a horrible viewpoint. Do you seriously entertain the thought that without observer there is no reality? Which observer? Any observer? Is a fly an observer? Is a star an observer? Was there no reality before $10^9$ B.C. before life began? Or are you the observer? Then there is no reality to the world after you are dead? I know a number of otherwise respectable physicists who have bought life insurance."

## 13.3 Impossibility of Non-Paradoxical Theories

Another traditional part of CI is the claim that it is impossible to provide any coherent (non-paradoxical) realist theory of what happens in the micro realm. Heisenberg (1958) again:

> "The idea of an objective real world whose smallest parts exist objectively in the same sense as stones or trees exist, independently of whether or not we observe them [...], is impossible."

We know from Bohmian mechanics that this claim is, in fact, wrong.

## 13.4 Completeness of the Wave Function

In CI, a microscopic system is completely described by its wave function. That is, there are no further variables (such as Bohm's particle positions) whose values nature knows and we do not. For this reason, the wave function is also called the *quantum state* or the *state vector*.

## 13.5 Language of Measurement

CI introduced (and established) the words "measurement" and "observable," and emphasized the analogy suggested by these words: E.g., that the momentum operator is analogous to the momentum variable in classical mechanics, and that the spin observable $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is analogous to the spin vector of classical mechanics (which points along the axis of spinning, and whose magnitude is proportional to the angular frequency).

I have already mentioned that these two words are quite inappropriate because they suggest that there was a value of the observable $A$ that was merely discovered (i.e., made known to us) in the experiment, whereas in fact the outcome is often only created during the experiment. Think, for example, of a Stern–Gerlach experiment in Bohmian mechanics: The particle does not have a value of $z$-spin before we carry out the experiment. And in CI, since it insists that wave functions are complete, it is true in spades that $A$ does not have a pre-existing, well-defined value before the experiment. So this terminology is *even less* appropriate in CI—and yet, it is a cornerstone of CI! Well, CI leans towards paradoxes.

## 13.6 Complementarity

Another idea of CI, called *complementarity*, is that in the micro realm, reality is paradoxical (contradictory) but the contradictions can never be seen (and are therefore not problematical) because of the Heisenberg uncertainty relation. (Recall Feynman's discussion of how the uncertainty relation keeps some things invisible.) Here is Bohr's definition of complementarity:

> "Any given application of classical concepts precludes the simultaneous use of other classical concepts which in a different connection are equally necessary for the elucidation of the phenomena."

I would describe the idea as follows. In order to compute a quantity of interest (e.g., the wave length of light scattered off an electron), we use both Theory A (e.g., classical theory of billiard balls) and Theory B (e.g., classical theory of waves) although A and B contradict each other.[17] It is impossible to find one Theory C that replaces both A

[17] In fact, before 1926 many successful theoretical considerations for predicting the results of experiments proceeded in this way. For example, people made a calculation about the collision between an electron and a photon as if they were classical billiard balls, then converted the momenta into wave lengths using de Broglie's relation $p = \hbar k$, then made another calculation about waves with wave number $k$.

and B and explains the entire physical process. (Here we meet again the impossibility claim mentioned in Section 13.3.) Instead, we should leave the conflict between A and B unresolved and accept the idea that reality is paradoxical.

Bell (*Speakable and Unspeakable in Quantum Mechanics*, page 190) wrote the following about complementarity:

> "It seems to me that Bohr used this word with the reverse of its usual meaning. Consider for example the elephant. From the front she is head, trunk and two legs. From the back she is bottom, tail, and two legs. From the sides she is otherwise, and from the top and bottom different again. These various views are complementary in the usual sense of the word. They supplement one another, they are consistent with one another, and they are all entailed by the unifying concept 'elephant.' It is my impression that to suppose Bohr used the word 'complementary' in this ordinary way would have been regarded by him as missing his point and trivializing his thought. He seems to insist rather that we must use in our analysis elements which *contradict* one another, which do not add up to, or derive from, a whole. By 'complementarity' he meant, it seems to me, the reverse: contradictoriness."

Einstein (1949):

> "Despite much effort which I have expended on it, I have been unable to achieve a sharp formulation of Bohr's principle of complementarity."

Bell commented (1986):

> "What hope then for the rest of us?"

Another version of complementarity concerns *observables that cannot be simultaneously measured.* We have encountered this situation in a homework exercise. Compare two experiments, each consisting of two measurements: (a) first measure $\sigma_2$ and then $\sigma_3$, (b) first measure $\sigma_3$ and then $\sigma_2$. We have seen that the joint probability distribution of the outcomes depends on the order. Some observables, though, *can* be measured simultaneously, i.e., the joint distribution does not depend on the order. Examples: $X_2$ and $X_3$, the $y$-component of position and the $z$-component; or $\sigma_2$ of particle 1 and $\sigma_3$ of particle 2.

**Theorem 13.1.** *The observables $A$ and $B$ can be simultaneously measured (i.e., for every wave function the joint probability distribution of the outcomes is independent of the order of the two measurements) iff the operators $A$ and $B$ commute, $AB = BA$.*

**Theorem 13.2.** *(An extension of the spectral theorem) Iff $A$ and $B$ commute, then there exists an ONB $\{\phi_n\}$ whose elements are eigenvectors of both operators $A$ and $B$, $A\phi_n = \alpha_n\phi_n$ and $B\phi_n = \beta_n\phi_n$.*

**Example 13.3.**

$$\sigma_2\sigma_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \sigma_3\sigma_2 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}. \tag{13.1}$$

Any two multiplication operators commute. In particular, the position operators $X_i$, $X_j$ commute with each other. The momentum operators $P_j = -i\hbar\partial/\partial x_j$ commute with each other. $X_i$ commutes with $P_j$ for $i \neq j$, but

$$[X_j, P_j] = i\hbar I \,, \tag{13.2}$$

with $I$ the identity operator. Eq. (13.2) is called *Heisenberg's canonical commutation relation.* To verify it, it suffices to consider a function $\psi$ of a 1-dimensional variable $x$. Using the product rule,

$$\begin{aligned} [X,P]\psi(x) &= XP\psi(x) - PX\psi(x) && (13.3)\\ &= x(-i\hbar)\frac{\partial\psi}{\partial x} - (-i\hbar)\frac{\partial}{\partial x}\Big(x\psi(x)\Big) && (13.4)\\ &= -i\hbar x\frac{\partial\psi}{\partial x} + i\hbar\psi(x) + i\hbar x\frac{\partial\psi}{\partial x} && (13.5)\\ &= i\hbar\psi(x)\,. && (13.6) \end{aligned}$$

So, for two commuting observables, the quantum formalism provides a *joint probability distribution.* For non-commuting observables, it does not. That is, it provides *two* joint probability distributions, one for each order, but that means it does not provide an unambiguous joint probability distribution. Moreover,

> *two non-commuting observables typically do not both have sharp values at the same time.* (13.7)

Also this fact is often called complementarity. For example, there is no quantum state that is an eigenvector to both $\sigma_2$ and $\sigma_3$. In CI, this fact is understood as a paradoxical trait of the micro-realm that we are forced to accept. That this paradoxical trait is connected to non-commutativity fits nicely with the analogy between operators in quantum mechanics and quantities in classical mechanics (as described in Section 13.5): In classical mechanics, which is free of paradoxes, all physical quantities (e.g., positions, momenta, spin vectors) are just numbers and therefore commute.

As a further consequence of (13.7), a measurement of $B$ must disturb the value of $A$ if $AB \neq BA$. (Think of the exercise in which $|z\text{-up}\rangle$ underwent a $\sigma_2$- and then a $\sigma_3$-measurement: After the $\sigma_2$-measurement, the particle was not certain any more to yield "up" in the $\sigma_3$-measurement.) Also the Heisenberg uncertainty relation is connected to (13.7), as it expresses that position and momentum cannot both have sharp values (i.e., $\sigma_X = 0$ and $\sigma_P = 0$) at the same time. In fact, the following generalized version of Heisenberg's uncertainty relation applies to observables $A$ and $B$ instead of $X$ and $P$:

**Theorem 13.4.** *(Robertson–Schrödinger inequality)*[18] *For any bounded self-adjoint operators $A, B$ and any $\psi \in \mathscr{H}$ with $\|\psi\| = 1$,*

$$\sigma_A\,\sigma_B \geq \frac{1}{2}\Big|\langle\psi|[A,B]|\psi\rangle\Big|\,. \tag{13.8}$$

[18] H.P. Robertson: The Uncertainty Principle. *Physical Review* **34**: 163–164 (1929)
E. Schrödinger: Zum Heisenbergschen Unschärfeprinzip. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalisch-mathematische Klasse* **14**: 296–303 (1930)

Note that the inequality is so much the stronger as the commutator $[A, B]$ is bigger, and becomes vacuous when $[A, B] = 0$.

*Proof.* Recall that the distribution over the spectrum of $A$ defined by $\psi$ has expectation value $\langle A \rangle := \langle \psi | A | \psi \rangle$ and variance

$$\sigma_A^2 = \langle \psi | (A - \langle A \rangle)^2 | \psi \rangle = \|\phi_A\|^2 \tag{13.9}$$

with

$$\phi_A := (A - \langle A \rangle)\psi \,, \tag{13.10}$$

where we simply wrote $\langle A \rangle$ for $\langle A \rangle I$. By the Cauchy-Schwarz inequality,

$$\sigma_A^2 \, \sigma_B^2 = \|\phi_A\|^2 \, \|\phi_B\|^2 \geq \big| \langle \phi_A | \phi_B \rangle \big|^2 \,. \tag{13.11}$$

Since

$$\begin{aligned} \langle \phi_A | \phi_B \rangle &= \langle \psi | (A - \langle A \rangle)(B - \langle B \rangle) | \psi \rangle && (13.12) \\ &= \langle \psi | (AB - \langle A \rangle B - A \langle B \rangle + \langle A \rangle \langle B \rangle) | \psi \rangle && (13.13) \\ &= \langle AB \rangle - \langle A \rangle \langle B \rangle \,, && (13.14) \end{aligned}$$

we obtain that

$$\begin{aligned} \big| \langle \phi_A | \phi_B \rangle \big|^2 &\geq \Big( \mathrm{Im} \langle \phi_A | \phi_B \rangle \Big)^2 && (13.15) \\ &= \Big| \frac{\langle \phi_A | \phi_B \rangle - \langle \phi_B | \phi_A \rangle}{2i} \Big|^2 && (13.16) \\ &= \Big| \frac{\langle AB \rangle - \langle A \rangle \langle B \rangle - \langle BA \rangle + \langle B \rangle \langle A \rangle}{2} \Big|^2 && (13.17) \\ &= \frac{1}{4} \big| \langle \psi | [A, B] | \psi \rangle \big|^2 \,. && (13.18) \end{aligned}$$

□

## 13.7 Reactions to the Measurement Problem

While Bohmian mechanics, GRW theory, and many-worlds theories have clear answers to the measurement problem, this is not so with Copenhagen. I report some answers that I heard Copenhagenists give (with some comments in brackets); I must admit that I do not see how these answers would make the problem go away.

- Nobody can actually solve the Schrödinger equation for $10^{23}$ interacting particles. (Sure, and we do not need to. If $\Psi_\alpha$ looks like a state including a pointer pointing to $\alpha$ then we know by linearity that $\Psi_{t_1}$ evolves to $\Psi_{t_2} = \sum c_\alpha \Psi_\alpha$, a superposition of macroscopically different states.)

- Systems are never isolated. (If we cannot solve the problem for an isolated system, what hope can we have to treat a non-isolated one? The way you usually treat a non-isolated system is by regarding it as a subsystem of a bigger, isolated system, maybe the entire universe.)

- Maybe there is no wave function of the universe. (It is up to Copenhagenists to propose a formulation that applies to the entire universe. Bohm, GRW, and many-worlds can do that.)

- Who knows whether the initial wave function is really a product as in $\Psi_{t_1} = \psi \otimes \phi$. (It is not so important that it is precisely a product, but it is important that we could perform a quantum measurement on any $\psi$.)

- The collapse of the wave function is like the collapse of a probability distribution: as soon as I have more information, such as $X \in B$, I have to update my probability distribution $\rho_{t-}$ for $X$ accordingly, namely to

$$\rho_{t+}(x) = 1_{x\in B}\, \rho_{t-}(x)\,. \tag{13.19}$$

  (The parallel is indeed striking. However, if we insist that the wave function is complete, then there never is any new information, as there is nothing that we are ignorant of.)

- Decoherence makes sure that you can replace the superposition $\Psi = \sum c_\alpha \Psi_\alpha$ by a mixture [i.e., a random one of the $\Psi_\alpha$]. (A super-observer cannot distinguish between the superposition and the mixture, but we are asking whether in reality it is a superposition or a mixture.)

# 14 Many Worlds

Put very briefly, Everett's many-worlds theory is GRW$\emptyset$ with $\lambda = 0$, and Schrödinger's many-worlds theory is GRWm with $\lambda = 0$.

The motivation for the many-worlds view comes from the wave function (11.3) of object and apparatus together after a quantum measurement. It is a superposition of macroscopically different terms. If we insist that the Schrödinger equation is correct (and thus reject non-linear modifications such as GRW), and if we insist that the wave function is complete, then we must conclude that there are different parts of reality, each looking like our world but with a different measurement outcome, and without any interaction between the different parts. They are parallel worlds. This view was suggested by Hugh Everett III in 1957.[19]

Everett's is not the only many-worlds theory, though. It is less well known that also Schrödinger had a many-worlds theory in 1926, and it is useful to compare the two.[20] Schrödinger, however, did not realize that his proposal was a many-worlds theory. He thought of it as a single-world theory. He came to the conclusion that it was empirically inadequate and abandoned it. Let us first try to get a good understanding of this theory.

## 14.1 Schrödinger's Many-Worlds Theory

According to Schrödinger's 1926 theory, matter is distributed continuously in space with density

$$m(\boldsymbol{x},t) = \sum_{i=1}^{N} m_i \int\limits_{\mathbb{R}^{3N}} d^3\boldsymbol{x}_1 \cdots d^3\boldsymbol{x}_N \, \delta^3(\boldsymbol{x}_i - \boldsymbol{x}) \left|\psi_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\right|^2 , \qquad (14.1)$$

and $\psi_t$ evolves according to the Schrödinger equation. The equation for $m$ is exactly the same as in GRWm, except that $\psi$ is not the same wave function. (Actually, Schrödinger replaced the mass factor $m_i$ by the electric charge $e_i$, but this difference is not crucial. It amounts to a different choice of weights in the weighted average over $i$. In fact, Schrödinger's choice has the disadvantage that the different signs of charges will lead to partial cancellations and thus to an $m$ function that looks less plausible as the density of matter. Nevertheless, the two choices turn out to be empirically equivalent, i.e., lead to the same predictions.)

[19] H. Everett: *The Theory of the Universal Wavefunction.* Ph. D. thesis, Department of Physics, Princeton University (1955). Reprinted on page 3–140 in B. DeWitt and R.N. Graham (editors): *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton: University Press (1973)

H. Everett: Relative State Formulation of Quantum Mechanics. *Reviews of Modern Physics* **29**: 454–462 (1957)

[20] E. Schrödinger: Quantisierung als Eigenwertproblem (Vierte Mitteilung). *Annalen der Physik* **81**: 109–139 (1926). English translation by J.F. Shearer in E. Schrödinger: *Collected Papers on Wave Mechanics.* New York: Chelsea (1927).

See also V. Allori, S. Goldstein, R. Tumulka, and N. Zanghì: Many-Worlds and Schrödinger's First Quantum Theory. *British Journal for the Philosophy of Science* **62(1)**: 1–27 (2011) `http://arxiv.org/abs/0903.2211`

In analogy to GRWm, we may call this theory Sm (where S is for the Schrödinger equation). Consider a double-slit experiment in this theory. Before the arrival at the detection screen, the contribution to the $m$ function coming from the electron sent through the double slit (which is the only contribution in the region of space between the double-slit and the detection screen) is a lump of matter smeared out over rather large distances (as large as the interference pattern). This lump is not homogeneous, it has interference fringes. And the overall amount of matter in this lump is tiny: If you integrate $m(\boldsymbol{x}, t)$ over $\boldsymbol{x}$ in the region between the double-slit and the detection screen, the result is $10^{-30}$ kg, the mass of an electron. But focus now on the fact that the matter is spread out. Schrödinger incorrectly thought that this fact must lead to the wrong prediction that the entire detection screen should glow faintly instead of yielding one bright spot, and that was why he thought Sm was empirically inadequate.

To understand why this reasoning was incorrect, consider a post-measurement situation (e.g., Schrödinger's cat). The wave function is a superposition of macroscopically different terms, $\Psi = \sum_\alpha c_\alpha \Psi_\alpha$. The $\Psi_\alpha$ do not overlap; i.e., where one $\Psi_\alpha$ is significantly nonzero, the others are near zero. Thus, when we compute $|\Psi|^2$ there are no (significant) cross terms; that is, for each $q$ there is at most one $\alpha$ contributing, so

$$|\Psi(q)|^2 = |c_\alpha|^2 \, |\Psi_\alpha(q)|^2 \, . \tag{14.2}$$

Define $m_\alpha(\boldsymbol{x})$ as what $m$ would be according to (14.1) with $\psi = \Psi_\alpha$. Then we obtain (to an excellent degree of approximation)

$$m(\boldsymbol{x}) = \sum_\alpha |c_\alpha|^2 \, m_\alpha(\boldsymbol{x}) \, . \tag{14.3}$$

In words, the $m$ function is a linear combination of $m$ functions corresponding to the macroscopically different terms in $\Psi$. So, for Schrödinger's cat in Sm, there is a dead cat and there is a live cat, each with half the mass. However, they do not notice they have only half the mass, and they do not notice the presence of the other cat. That is because, if we let the time evolve, then each $\Psi_\alpha(t)$ evolves in a way that corresponds to a reasonable story of just one cat; after all, it is how the wave function would evolve according to the projection postulate after a measurement of the cat had collapsed the superposition to one of the $\Psi_\alpha$. Furthermore, $\Psi(t) = \sum_\alpha c_\alpha \Psi_\alpha(t)$ by linearity, and since the $\Psi_\alpha(t)$ remain non-overlapping, we have that (14.3) applies to every $t$ from now on, that is

$$m(\boldsymbol{x}, t) = \sum_\alpha |c_\alpha|^2 \, m_\alpha(\boldsymbol{x}, t) \, . \tag{14.4}$$

Each $m_\alpha(t)$ looks like the reasonable story of just one cat that $\Psi_\alpha(t)$ corresponds to. Thus, the two cats do not interact with each other; they are causally disconnected. After all, the two contributions $m_\alpha$ come from $\Psi_\alpha$ that are normally thought of as *alternative* outcomes of the experiment. So the two cats are like ghosts to each other: they can see and walk through each other.

And not just the cat has split in two. If a camera takes a photograph of the cat then $\Psi$ must be taken to be a wave function of the cat and the camera together (among other things). $\Psi_1$ may then correspond to a dead cat and a photo of a dead cat, $\Psi_2$ to a live cat and a photo of a live cat. If a human being interacts with the cat (say, looks at it), then $\Psi_1$ will correspond to a brain state of seeing a dead cat and $\Psi_2$ to one of seeing a live cat. That is, there are two copies of the cat, two copies of the photo, two copies of the human being, two copies of the entire world. That is why I said that Sm has a many-worlds character. In each world, though, things seem rather ordinary: Like a single cat in an ordinary (though possibly pitiful) state, and all records and memories are consistent with each other and in agreement with the state of the cat.

## 14.2 Everett's Many-Worlds Theory

Everett's many-worlds theory, which could be called S$\emptyset$ (S for the Schrödinger equation and $\emptyset$ for the empty primitive ontology) is based on the idea that the same picture would arise if we dispense with the $m$ function. Frankly, I do not see how it would; I actually cannot make sense of S$\emptyset$ as a physical theory. But I would say the more basic problem is not how to obtain probabilities, but how to obtain *things* such as cats, chairs, pointers. The primitive ontology is missing. And that problem is solved in Sm. Note, though, that for a person who believes that S$\emptyset$ makes sense, this theory would seem like the simplest possible coherent theory that would account for quantum mechanics. To such a person it would seem that the existence of many worlds is a necessary consequence of the Schrödinger equation, which, after all, leads to macroscopic superpositions such as the $\Psi = \sum_\alpha c_\alpha \Psi_\alpha$ above. In contrast, a person who believes that S$\emptyset$ does not make sense while Sm does, will not have such a sense of necessity, as the many-worlds character of the theory does not come from $\Psi$ but from $m$, and if we had postulated a different primitive ontology (say, Bohmian particles instead of (14.1)), then no many-worlds character would have arisen.

While there is disagreement in the literature about the relevance of a primitive ontology, many authors argue that S$\emptyset$ has a *preferred basis problem*: If there exists nothing more than $\Psi$, and if $\Psi$ is just a vector in Hilbert space $\mathscr{H}$, then how do we know which basis to choose in $\mathscr{H}$ to obtain the different worlds? For example, if

$$\Psi = \tfrac{1}{\sqrt{2}}|\text{dead}\rangle + \tfrac{1}{\sqrt{2}}|\text{alive}\rangle\,, \tag{14.5}$$

then we could also write

$$\Psi = \tfrac{e^{i\pi/4}}{\sqrt{2}}|+\rangle + \tfrac{e^{-i\pi/4}}{\sqrt{2}}|-\rangle\,, \tag{14.6}$$

where

$$|+\rangle = \tfrac{1}{\sqrt{2}}|\text{dead}\rangle + \tfrac{i}{\sqrt{2}}|\text{alive}\rangle\,, \qquad |-\rangle = \tfrac{1}{\sqrt{2}}|\text{dead}\rangle - \tfrac{i}{\sqrt{2}}|\text{alive}\rangle \tag{14.7}$$

form another ONB of the subspace spanned by $|\text{dead}\rangle$ and $|\text{alive}\rangle$. So how do we know that the two worlds correspond to $|\text{dead}\rangle$ and $|\text{alive}\rangle$ rather than to $|+\rangle$ and $|-\rangle$? Obviously, in Sm there is no such problem because a preferred basis (the position basis) is built into the law (14.1) for $m$.

## 14.3 Bell's First Many-Worlds Theory

Bell also made a proposal (first formulated in 1971, published[21] in 1981) adding a primitive ontology to Everett's S$\emptyset$; Bell did not seriously propose or defend the resulting theory, he just regarded it as an ontological clarification of Everett's theory. According to this theory, at every time $t$ there exists an uncountably infinite collection of universes, each of which consists of $N$ material points in Euclidean 3-space. Thus, each world has its own configuration $Q$, but some configurations are more frequent in the ensemble of worlds than others, with $|\Psi_t|^2$ distribution across the ensemble. At every other time $t'$, there is again an infinite collection of worlds, but there is no fact about which world at $t'$ is the same as which world at $t$.

## 14.4 Bell's Second Many-Worlds Theory

Another variant of this theory, considered by Bell in 1976,[22] supposes that there is really a single world at every time $t$ consisting of $N$ material points in Euclidean 3-space. The configuration $Q_t$ chosen with $|\Psi_t|^2$ distribution indepedently at every time. Although this theory has a definite $Q_t$ at every $t$, it also has a many-worlds character because in every arbitrarily short time interval, configurations from all over configuration space are realized, in fact with distribution roughly equal to $|\Psi_t|^2$ (if the interval is short enough and $\Psi_t$ depends continuously on $t$) across the ensemble of worlds existing at different times. This theory seems rather implausible compared to Bohmian mechanics, as it implies that our memories are completely wrong: after all, it implies that one minute ago the world was not at all like what we remember it to be like a minute ago. Given that all of our reasons for believing in the Schrödinger equation and the Born rule are based on memories of reported outcomes of experiments, it seems that this theory undercuts itself: if we believe it is true then we should conclude that our belief is not justified.

It is not very clear to me whether the same objection applies to Bell's first many-worlds theory. But certainly, both theories have, due to their radically unusual idea of what reality is like, a flavor of skeptical scenarios (such as the brain in the vat), in fact a stronger such flavor than Sm.

## 14.5 Probabilities in Many-World Theories

Maudlin expressed in his article on the measurement problem a rather negative opinion about many-worlds theories; I think a bit too negative. His objection was, if every outcome $\alpha$ of an experiment is realized, what could it mean to say that outcome $\alpha$ has probability $|c_\alpha|^2$ to occur? If, as in Sm and in S$\emptyset$, all the equations are deterministic,

[21] J.S. Bell: Quantum Mechanics for Cosmologists. Pages 611–637 in C. Isham, R. Penrose and D. Sciama (editors), *Quantum Gravity 2*, Oxford: Clarendon Press (1981). Reprinted as chapter 15 of J.S. Bell: *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: University Press (1987)

[22] J.S. Bell: The Measurement Theory of Everett and de Broglie's Pilot Wave. Pages 11–17 in M. Flato et al. (editors): *Quantum Mechanics, Determinism, Causality, and Particles*, Dordrecht: Reidel (1976). Reprinted as chapter 11 of J.S. Bell: *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: University Press (1987)

then there is nothing random; and in the situation of the measurement problem, there is nothing that we are ignorant of. So what could talk of probability mean?

Here is what it could mean in Sm: Suppose we have a way of *counting worlds.* And suppose we repeat a quantum experiment (say, a Stern–Gerlach experiment with $|c_{\mathrm{up}}|^2 = |c_{\mathrm{down}}|^2 = 1/2$) many times (say, a thousand times). Then we obtain in each world a sequence of 1000 ups and downs such as

$$\uparrow\downarrow\uparrow\uparrow\downarrow\uparrow\downarrow\downarrow\downarrow\dots. \tag{14.8}$$

Note that there are $2^{1000} \approx 10^{300}$ such sequences. The statement that the fraction of ups lies between 47% and 53% is true in some worlds and false in others. Now count the worlds in which the statement is true. Suppose that the statement is true in *the overwhelming majority of worlds.* Then that would explain why we find ourselves in such a world. And that, in turn, would explain why we observe a relative frequency of ups of about 50%. And *that* is what we needed to explain for justifying the use of probabilities.

Now consider $|c_{\mathrm{up}}|^2 = 1/3$, $|c_{\mathrm{down}}|^2 = 2/3$. Then the argument might seem to break down, because it is then still true that in the overwhelming majority of sequences such as (14.8) the frequency of ups is about 50%. But consider the following

**Rule for counting worlds.** *The "fraction of worlds" $f(P)$ with property $P$ in the splitting given by $\Psi = \sum_\alpha c_\alpha \Psi_\alpha$ and $m(\boldsymbol{x}) = \sum_\alpha |c_\alpha|^2 m_\alpha(\boldsymbol{x})$ is*

$$f(P) = \sum_{\alpha \in M} |c_\alpha|^2 \,, \tag{14.9}$$

*where $M$ is the set of worlds $\alpha$ with property $P$.*

Note that $f(P)$ lies between 0 and 1 because $\sum_\alpha |c_\alpha|^2 = 1$. It is not so clear whether this rule makes sense—whether there is room in physics for such a law. But let us accept it for the moment and see what follows. Consider the property $P$ that the relative frequency of ups lies between 30% and 36%. Then $f(P)$ is actually the same value as the probability of obtaining a frequency of ups between 30% an 36% in 1000 consecutive independent random tossings of a biased coin with $\mathbb{P}(\mathrm{up}) = 1/3$ and $\mathbb{P}(\mathrm{down}) = 2/3$. And in fact, this value is very close to 1. Thus, the above rule for counting worlds implies the frequency of ups lies between 30% and 36% in the overwhelming majority of worlds. This reasoning was essentially developed by Everett.

A comparison with Bohmian mechanics is useful. The initial configuration of the lab determines the precise sequence such as (14.8). If the initial configuration is chosen with $|\Psi_0|^2$ distribution, then with overwhelming probability the sequence will have a fraction of ups between 30% and 36%. That is, if we *count initial conditions* with the $|\Psi_0|^2$ distribution, that is, if we say that the fraction of initial conditions lying in a set $B \subseteq \mathbb{R}^{3N}$ is $\int_B |\Psi_0|^2$, then we can say that for *the overwhelming majority of Bohmian worlds*, the observed frequency is about 33%. Now to make the connection with many-worlds, note that the reasoning does not depend, in fact, on whether all of the worlds are realized or just one. That is, imagine many Bohmian worlds with the same initial wave function $\Psi_0$ but different initial configurations, distributed across the ensemble according to $|\Psi_0|^2$. Then there is an explanation for why inhabitants should see a frequency of about 33%.

The problem that remains is whether there is room for a rule for counting worlds. In terms of a creation myth, suppose God created the wave function $\Psi$ and made it a law that $\Psi$ evolves according to the Schrödinger equation; then he created matter in 3-space distributed with density $m(\boldsymbol{x}, t)$ and made it a law that $m$ is given by (14.1). Now what would God need to do in order to make the rule for counting worlds a law? He does not create anything further, so in which way would two universes with equal $\Psi$ and $m$ but different rules for counting worlds differ? That is a reason for thinking that ultimately, Sm fails to work (though in quite a subtle way).

Various authors have proposed other reasonings for justifying probabilities in many-worlds theories; they seem less relevant to me, but let me mention a few. David Deutsch[23] proposed that it is rational for inhabitants of a universe governed by a many-worlds theory (a "multiverse," as it is often called) to behave as if the events they perceive were random with probabilities given by the Born rule; he proposed certain principles of rational behavior from which he derived this. (Of course, this reasoning does not provide an explanation of why we observe frequencies in agreement with Born's rule.) Lev Vaidman[24] proposed that in a many-worlds scenario, I can be ignorant of which world I am in: before the measurement, I know that there will be a copy of me in each post-measurement world, and afterwards, I do not know which worlds I am in until I look at the pointer position. And I could try to express my ignorance through a probability distribution, although it is not clear why the Born distribution would be correct and other distributions would not.

For comparison, in Bell's many-worlds theories it is not hard to make sense of probabilities. In Bell's first theory, there is an ensemble of worlds at every time $t$, and clearly most of the worlds have configurations that look as if randomly chosen with $|\Psi|^2$ distribution, in particular with a frequency of ups near 33% in the example described earlier. In Bell's second theory, $Q_t$ is actually random with $|\Psi_t|^2$ distribution, and although the recorded sequence of outcomes fluctuates within every fraction of a second, the sequence in our memories and records at time $t$ has, with probability near 1, a frequency of ups near 33%.

[23]D. Deutsch: Quantum theory of probability and decisions. *Proceedings of the Royal Society of London A* **455**: 3129–3137 (1999) `http://arxiv.org/abs/quant-ph/9906015`

[24]L. Vaidman: On Schizophrenic Experiences of the Neutron or Why We should Believe in the Many-Worlds Interpretation of Quantum Theory. *International Studies in the Philosophy of Science* **12**: 245–261 (1998) `http://arxiv.org/abs/quant-ph/9609006`

# 15 The Einstein–Podolsky–Rosen Argument

In the literature, the "EPR paradox" is often mentioned. It is clear from EPR's article that they did not intend to describe a paradox (as did, e.g., Wheeler when describing the delayed-choice experiment), but rather to describe an argument. The argument supports the conclusion that there are additional variables beyond the wave function. I now explain their reasoning in my own words, partly in preparation for Bell's 1964 argument ,which builds on EPR's argument.

## 15.1 The EPR Argument

EPR consider 2 particles in 1 dimension with entangled wave function

$$\Psi(x_1, x_2) = \delta(x_1 - x_2 + x_0)\,, \tag{15.1}$$

with $x_0$ a constant. (We ignore the fact that this wave function is unphysical because it does not lie in Hilbert space; the same argument could be made with square-integrable functions but would become less transparent.) An observer, let us call her Alice, measures the position of particle 1. The outcome $X_1$ is uniformly distributed, and the wave function collapses to

$$\Psi'(x_1, x_2) = \delta(x_1 - X_1)\delta(x_2 - X_1 - x_0)\,, \tag{15.2}$$

so that another observer, Bob, measuring the position of particle 2, is certain to obtain $X_2 = X_1 + x_0$. It follows that particle 2 had a position even before Bob made his experiment. Now EPR make the assumption that

> no real change can take place in the second system in consequence of [a measurement on] the first system. (15.3)

They take it as obviously true, but it is worthy of a closer examination. We will come back to it in the next chapter. It then follows that particle 2 had a definite position even before Alice made her experiment, despite the fact that $\Psi$ is not an eigenfunction of $x_2$-position. Quod erat demonstrandum.

EPR draw further conclusions from their example by considering also momentum. Note that the Fourier transform of $\Psi$ is

$$\widehat{\Psi}(k_1, k_2) = e^{-ik_1x_0}\,\delta(k_1 + k_2)\,. \tag{15.4}$$

Alice could measure either the position or the momentum of particle 1, and Bob either the position or the momentum of particle 2. If Alice measures position then, as seen above, the outcome $X_1$ is uniformly distributed and Bob, if he chooses to measure position, finds $X_2 = X_1 + x_0$ with certainty. If, alternatively, Alice measures momentum then the outcome $K_1$ will be uniformly distributed and the wave function in momentum representation collapses from $\widehat{\Psi}$ to

$$\widehat{\Psi}''(k_1, k_2) = e^{-iK_1x_0}\,\delta(k_1 - K_1)\,\delta(k_2 + K_1) \tag{15.5}$$

so that Bob, if he chooses to measure momentum, is certain to find $K_2 = -K_1$. In the same way as above, it follows that Bob's particle had a position before any of the experiments, *and* that it had a momentum!

There even arises a way of simultaneously measuring the position and momentum of particle 2: Alice measures position $X_1$ and Bob momentum $K_2$. Since particle 2 has, as just proved, a well-defined position and a well-defined momentum, and since, by (15.3), Alice's measurement did not influence particle 2, $K_2$ must be the original momentum of particle 2. Likewise, if Bob had chosen to measure position, his result would have agreed with the original position, and since it would have obeyed $X_2 = X_1 + x_0$, we can infer from Alice's result what the original position must have been.

## 15.2 Bohm's Version of the EPR Argument Using Spin

In 1951, before he discovered Bohmian mechanics, Bohm wrote a textbook about quantum mechanics in which he followed the orthodox view. In it, he also described the following useful variant of the EPR argument.

Consider two spin-$\frac{1}{2}$ particles with joint spinor in $\mathbb{C}^4$ given by the *singlet state*

$$\phi = \frac{1}{\sqrt{2}}\Big(|z\text{-up}\rangle|z\text{-down}\rangle - |z\text{-down}\rangle|z\text{-up}\rangle\Big)\,. \tag{15.6}$$

Alice measures $\sigma_3$ on particle 1. The outcome $Z_1$ is $\pm 1$, each with probability 1/2. If $Z_1 = +1$ then the wave function collapses to

$$\phi'_+ = |z\text{-up}\rangle|z\text{-down}\rangle\,, \tag{15.7}$$

and Bob, measuring $\sigma_3$ on particle 2, is certain to obtain $Z_2 = -1$. If, however, $Z_1 = -1$ then the wave function collapses to

$$\phi'_- = |z\text{-down}\rangle|z\text{-up}\rangle\,, \tag{15.8}$$

and Bob is certain to obtain $Z_2 = +1$. Thus, always $Z_2 = -Z_1$; one speaks of *perfect anti-correlation.* As a consequence, particle 2 had a definite value of $z$-spin even before Bob's experiment. Now, from the assumption (15.3) it follows that it had that value even before Alice's experiment. Likewise, particle 1 had a definite value of $z$-spin before any attempt to measure it.

Again as in EPR's reasoning, we can consider other observables, say $\sigma_1$ and $\sigma_2$. In homework Exercise 30 of Assignment 7, we checked that the singlet state has the same form relative to the $x$-spin basis or the $y$-spin basis. It follows that if Alice and Bob both measure $x$-spin then their outcomes are also perfectly anti-correlated, and likewise for $y$-spin. It can be inferred that each spin component, for each particle, has a well-defined value before any experiment.

Moreover, Alice and Bob together can measure $\sigma_1$ *and* $\sigma_3$ of particle 2: Alice measures $\sigma_1$ of particle 1 and Bob $\sigma_3$ of particle 2. By (15.3) and the perfect anti-correlation, the negative of Alice's outcome is what Bob would have obtained had he measured $\sigma_1$; and by (15.3), Bob's outcome is not affected by Alice's experiment.

## 15.3 Einstein's Boxes Argument

We have seen that EPR's argument yields more than just the incompleteness of the wave function. It also yields that particles have well-defined positions *and* momenta. If we only want to establish the incompleteness of the wave function, which seems like a worthwhile goal for a proof, a simpler argument will do. Einstein developed such an argument already in 1927 (before the EPR paper), presented it at a conference but never published it.[25]

Consider a single particle whose wave function $\psi(\boldsymbol{x})$ is confined to a box $B$ with impermeable walls and (more or less) uniform in $B$. Now split $B$ (e.g., by inserting a partition) into two boxes $B_1$ and $B_2$, move one box to Tokyo and the other to Paris. There is some nonzero amount of the particle's wave function in Paris and some in Tokyo. Carry out a detection in Paris. Let us assume that

$$\text{no real change can take place in Tokyo in consequence of a measurement in Paris.} \tag{15.9}$$

If we believed that the wave function was a complete description of reality, then there would be no fact of the matter, before the detection experiment, about whether the particle is in Paris or Tokyo, but afterwards there would be. This contradicts (15.9), so the wave function cannot be complete.

The assumption (15.9) is intended as allowing changes in Tokyo after a while, such as the while it would take a signal to travel from Paris to Tokyo at the speed of light. That is, (15.9) (and similarly (15.3)) is particularly motivated by the theory of relativity, which strongly suggests that signals cannot propagate faster than at the speed of light. On one occasion, Einstein wrote that the faster-than-light effect entailed by insisting on completeness of the wave function was "spukhafte Fernwirkung" (spooky action-at-a-distance).

## 15.4 Too Good To Be True

EPR's argument is, in fact, correct. Nevertheless, it may strike you that its conclusion, the incompleteness of the wave function, is very strong—maybe too strong to be true. After all, it is not true in GRW or many-worlds! How can this be: that EPR proved something that is not true?

This can happen only because the assumption (15.3) is actually not true in these theories. And in Bohmian mechanics, where the wave function is in fact incomplete, it is not true that all spin observables have pre-existing actual values, as would follow from EPR's reasoning. Thus, also in Bohmian mechanics (15.3) is not true. We will see in the next chapter that (15.3) is problematical in any version of quantum mechanics. This fact was discovered 30 years after EPR's paper by John Bell.

[25]It has been reported by, e.g., L. de Broglie: *The Current Interpretation of Wave Mechanics: A Critical Study.* Elsevier (1964). A more detailed discussion is given by T. Norsen: Einstein's Boxes, *American Journal of Physics* **73(2)**: 164–176 (2005) `http://arxiv.org/abs/quant-ph/0404016`

# 16 Nonlocality

Two space-time points $x = (s, \boldsymbol{x})$ and $y = (t, \boldsymbol{y})$ are called *spacelike separated* iff no signal propagating at the speed of light can reach $x$ from $y$ or $y$ from $x$. This occurs iff

$$|\boldsymbol{x} - \boldsymbol{y}| > c|s - t|\,, \tag{16.1}$$

with $c = 3 \times 10^8$ m/s the speed of light. Einstein's theory of relativity strongly suggests that signals cannot propagate faster than at the speed of light (*superluminally*). That is, if $x$ and $y$ are spacelike separated then no signal can be sent from $x$ to $y$ or from $y$ to $x$. This in turn suggests that

> If $x$ and $y$ are spacelike separated then events at $x$ cannot influence events at $y$. (16.2)

This statement is called *locality.* It is true in relativistic versions of classical physics (mechanics, electrodynamics, and also in Einstein's relativistic theory of gravity he called the *general theory of relativity*). Bell proved in 1964 that locality is false if certain empirical predictions of the quantum formalism are correct; this analysis is often called *Bell's theorem.*[26] The relevant predictions have since been experimentally confirmed; the first convincing tests were carried out by Alain Aspect in 1982.[27] Thus, locality is false in our world; this fact is often called *quantum nonlocality.* Our main goal in this chapter is to understand Bell's proof.

Some remarks.

- Einstein believed in locality until his death in 1955. Locality is very closely related to (almost the same as) the EPR assumption (15.3): If Alice's measurement takes place at $x$ and Bob's at $y$, and if $x$ and $y$ are spacelike separated, then locality implies that Alice's measurement on particle 1 at $x$ cannot affect particle 2 at $y$. Conversely, the only situation in which we can be certain that the two particles *cannot* interact occurs if Alice's and Bob's experiments are spacelike separated and locality holds true. Ironically, EPR were wrong even though their argument was correct: The premise (15.3) is false. They took locality for granted. Likewise in Einstein's boxes argument, the assumption (15.9) is equivalent to locality: The point of talking about Tokyo and Paris is that these two places are distant, and since there clearly can be influences if we allow more time than distance/$c$, the assumption is that there cannot be an influence between spacelike separated events.

- Despite nonlocality, it is not possible to send messages faster than light, according to the appropriate relativistic version of the quantum formalism; this fact is often called the *no-signalling theorem.* We will prove it in great generality in a later

[26] J. S. Bell: On the Einstein-Podolsky-Rosen Paradox. *Physics* **1**: 195–200 (1964) Reprinted as chapter 2 of J. S. Bell: *Speakable and unspeakable in quantum mechanics.* Cambridge University Press (1987)

[27] A. Aspect, J. Dalibard, G. Roger: Experimental Test of Bell's Inequalities using Time-Varying Analyzers. *Physical Review Letters* **49**: 1804–1807 (1982)

chapter. Put differently, the superluminal influences cannot be used by agents for sending messages.

- Does nonlocality prove relativity wrong? That statement would be too strong. Nonlocality proves a *certain understanding* of relativity wrong. Much of relativity theory, however, remains untouched by nonlocality.

- If $x$ and $y$ are spacelike separated then relativistic Hamiltonians contain no interaction term between $x$ and $y$.

  Let me explain this statement. The Schrödinger equation is non-relativistic and needs to be replaced, in a relativistic theory, by a relativistic equation. The latter is different from the non-relativistic Schrödinger equation in two ways: (i) Instead of interaction potentials, interaction arises from the creation and annihilation of particles. For example, an electron can create a photon, which travels to another electron and is annihilated there. Potentials can only be used as an approximation. (ii) Even leaving interaction aside, relativity requires a modification of the Schrödinger equation. The best known such modification is the *Dirac equation* for electrons. It entails that the wave function can propagate no faster than at the speed of light $c$. Since also photon wave functions propagate no faster than at $c$, and since potentials are absent, there is no interaction term in the Hamiltonian between particles at $x$ and at $y$.

  So there are two meanings to the word "interaction": first, an interaction term in the Hamiltonian; second, any influence. Bell's proof shows that in the absence of the first type of interaction, the second type can still be present.

- Bell's proof shows for a certain experiment that *either* events at $x$ must have influenced events at $y$ *or* vice versa, but does not tell us who influenced whom.

## 16.1 Bell's Experiment

As in Bohm's version of the EPR example, consider two spin-$\frac{1}{2}$ particles in the singlet state

$$\phi = \frac{1}{\sqrt{2}}\Big(|z\text{-up}\rangle|z\text{-down}\rangle - |z\text{-down}\rangle|z\text{-up}\rangle\Big)\,. \tag{16.3}$$

While keeping their spinor constant, the two particles are brought to distant places. Alice makes an experiment on particle 1 at (or near) space-time point $x$ and Bob one on particle 2 at $y$; $x$ and $y$ are spacelike separated. Each experimenter chooses a direction in space, corresponding to a unit vector $\boldsymbol{n} \in \mathbb{R}^3$, and carries out a Stern–Gerlach experiment in that direction, i.e., a quantum measurement of $\boldsymbol{n}\cdot\boldsymbol{\sigma}$. The difference to Bohm's example is that Alice and Bob can choose different directions. I write $\boldsymbol{\alpha}$ for Alice's unit vector, $\boldsymbol{\beta}$ for Bob's, $Z^1$ for the random outcome $\pm1$ of Alice's experiment, and $Z^2$ for that of Bob's. Let us compute the joint distribution $\mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ of $Z^1$ and $Z^2$.

**Fact 1.** For any unit vector $\boldsymbol{n} \in \mathbb{R}^3$,

$$\phi \propto \Big( |\boldsymbol{n}\text{-up}\rangle |\boldsymbol{n}\text{-down}\rangle - |\boldsymbol{n}\text{-down}\rangle |\boldsymbol{n}\text{-up}\rangle \Big) . \tag{16.4}$$

Sketch of proof: Consider first the case that $\boldsymbol{n}$ is infinitesimally close to the $z$-direction, arising from $(0,0,1)$ by a rotation around the axis along the unit vector $\boldsymbol{m} = (\cos\gamma, \sin\gamma, 0)$ through an infinitesimal angle $\delta\varphi$. Then

$$\sigma_{\boldsymbol{m}} = \boldsymbol{m}\cdot\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{\boldsymbol{m}++} & \sigma_{\boldsymbol{m}+-} \\ \sigma_{\boldsymbol{m}-+} & \sigma_{\boldsymbol{m}--} \end{pmatrix} = \begin{pmatrix} 0 & \cos\gamma - i\sin\gamma \\ \cos\gamma + i\sin\gamma & 0 \end{pmatrix} \tag{16.5}$$

and

$$|\boldsymbol{n}\text{-up}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{\delta\varphi}{2}\boldsymbol{m}\cdot\boldsymbol{\sigma}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{\delta\varphi}{2}\begin{pmatrix} 0 \\ \sigma_{\boldsymbol{m}-+} \end{pmatrix} \tag{16.6}$$

$$|\boldsymbol{n}\text{-down}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{\delta\varphi}{2}\boldsymbol{m}\cdot\boldsymbol{\sigma}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{\delta\varphi}{2}\begin{pmatrix} \sigma_{\boldsymbol{m}+-} \\ 0 \end{pmatrix} \tag{16.7}$$

because spinors rotate through half the angle $\delta\varphi$. As a consequence, to first order in $\delta\varphi$,

$$|\boldsymbol{n}\text{-up}\rangle|\boldsymbol{n}\text{-down}\rangle = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \frac{\delta\varphi}{2}\begin{pmatrix} \sigma_{\boldsymbol{m}+-} & 0 \\ 0 & \sigma_{\boldsymbol{m}-+} \end{pmatrix} \tag{16.8}$$

and

$$|\boldsymbol{n}\text{-up}\rangle|\boldsymbol{n}\text{-down}\rangle - |\boldsymbol{n}\text{-down}\rangle|\boldsymbol{n}\text{-up}\rangle =$$
$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \frac{\delta\varphi}{2}\begin{pmatrix} \sigma_{\boldsymbol{m}+-} & 0 \\ 0 & \sigma_{\boldsymbol{m}-+} \end{pmatrix} - \frac{\delta\varphi}{2}\begin{pmatrix} \sigma_{\boldsymbol{m}+-} & 0 \\ 0 & \sigma_{\boldsymbol{m}-+} \end{pmatrix} . \tag{16.9}$$

This proves (16.4) for an infinitesimal change in $\boldsymbol{n}$. Now think of a finite change in $\boldsymbol{n}$ as partitioned into infinitely many infinitesimal changes. This proves (16.4) for arbitrary $\boldsymbol{n}$. $\square$

**Fact 2.** Independently of whether Alice's or Bob's experiment occurs first, the joint distribution of $Z^1, Z^2$ is

$$\mu_{\boldsymbol{\alpha},\boldsymbol{\beta}} := \begin{pmatrix} \mathbb{P}(\text{up,up}) & \mathbb{P}(\text{up,down}) \\ \mathbb{P}(\text{down,up}) & \mathbb{P}(\text{down,down}) \end{pmatrix} \tag{16.10}$$

$$= \begin{pmatrix} \frac{1}{4} - \frac{1}{4}\boldsymbol{\alpha}\cdot\boldsymbol{\beta} & \frac{1}{4} + \frac{1}{4}\boldsymbol{\alpha}\cdot\boldsymbol{\beta} \\ \frac{1}{4} + \frac{1}{4}\boldsymbol{\alpha}\cdot\boldsymbol{\beta} & \frac{1}{4} - \frac{1}{4}\boldsymbol{\alpha}\cdot\boldsymbol{\beta} \end{pmatrix} \tag{16.11}$$

$$= \begin{pmatrix} \frac{1}{2}\sin^2(\theta/2) & \frac{1}{2}\cos^2(\theta/2) \\ \frac{1}{2}\cos^2(\theta/2) & \frac{1}{2}\sin^2(\theta/2) \end{pmatrix} , \tag{16.12}$$

with $\theta$ the angle between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Proof: Assume that Alice's experiment occurs first and write the initial spinor as

$$\phi = c|\boldsymbol{\alpha}\text{-up}\rangle|\boldsymbol{\alpha}\text{-down}\rangle - c|\boldsymbol{\alpha}\text{-down}\rangle|\boldsymbol{\alpha}\text{-up}\rangle \tag{16.13}$$

with $c$ a complex constant with $|c| = 1/\sqrt{2}$. According to Born's rule, Alice obtains $+1$ or $-1$, each with probability $1/2$. In case $Z^1 = +1$, $\phi$ collapses to

$$\phi'_+ = |\boldsymbol{\alpha}\text{-up}\rangle|\boldsymbol{\alpha}\text{-down}\rangle\,. \tag{16.14}$$

According to Born's rule, the probability that Bob obtains $Z^2 = +1$ is

$$\mathbb{P}(Z^2 = +1|Z^1 = +1) = \Big|\langle\boldsymbol{\beta}\text{-up}|\boldsymbol{\alpha}\text{-down}\rangle\Big|^2 = 1 - \Big|\langle\boldsymbol{\beta}\text{-up}|\boldsymbol{\alpha}\text{-up}\rangle\Big|^2\,. \tag{16.15}$$

Since the angle in Hilbert space between $|\boldsymbol{\beta}\text{-up}\rangle$ and $|\boldsymbol{\alpha}\text{-up}\rangle$ is half the angle between $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, and since they are unit vectors in Hilbert space, we have that

$$\Big|\langle\boldsymbol{\beta}\text{-up}|\boldsymbol{\alpha}\text{-up}\rangle\Big| = \cos(\theta/2) \tag{16.16}$$

and thus

$$\mathbb{P}(Z^2 = +1|Z^1 = +1) = 1 - \cos^2(\theta/2) = \sin^2(\theta/2) \tag{16.17}$$

and

$$\mathbb{P}(Z^1 = +1, Z^2 = +1) = \frac{1}{2}\sin^2(\theta/2)\,. \tag{16.18}$$

Since $\cos^2 x = \frac{1}{2} + \frac{1}{2}\cos(2x)$, this value can be rewritten as

$$\mathbb{P}(Z^1 = +1, Z^2 = +1) = \frac{1}{2} - \frac{1}{2}\cos^2(\theta/2) = \frac{1}{2} - \frac{1}{4} - \frac{1}{4}\cos\theta = \frac{1}{4} - \frac{1}{4}\boldsymbol{\alpha}\cdot\boldsymbol{\beta}\,. \tag{16.19}$$

The other three matrix elements can be computed in the same way. Assuming that Bob's experiment occurs first leads to the same matrix. □

Remarks.

- Note that the four entries in $\mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ are nonnegative and add up to 1, as they should.
- In the case $\boldsymbol{\alpha} = \boldsymbol{\beta}$ corresponding to Bohm's version of the EPR example,

  $$\mu_{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \tag{16.20}$$

  implying the perfect anti-correlation $Z^2 = -Z^1$.
- The *marginal distribution* is the distribution of $Z^1$ alone, irrespective of $Z^2$. It is $1/2, 1/2$. Likewise for $Z^2$. Let us assume that Alice's experiment occurs first. Then the fact that the marginal distribution for $Z^2$ is $1/2, 1/2$ amounts to a *no-signalling theorem* for Bell's experiment: Bob cannot infer from $Z^2$ any information about Alice's choice $\boldsymbol{\alpha}$ because the distribution of $Z^2$ does not depend on $\boldsymbol{\alpha}$. (The general no-signalling theorem that we will prove later covers all possible experiments.)
- The fact that the joint distribution of the outcomes does not depend on the order of experiments means that the observables measured by Alice and Bob can be simultaneously measured. What are these observables, actually? Alice's is the matrix $\sigma_{\boldsymbol{\alpha}} \otimes I$ with components $\sigma_{\boldsymbol{\alpha} s_1 s_1'}\delta_{s_2 s_2'}$, and Bob's is $I \otimes \sigma_{\boldsymbol{\beta}}$ with components $\delta_{s_1 s_1'}\sigma_{\boldsymbol{\beta} s_2 s_2'}$.

## 16.2 Bell's 1964 Proof of Nonlocality

Let us recapitulate what needs to be shown in Bell's theorem. The claim is that the joint distribution $\mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ of $Z^1$ and $Z^2$, as a function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, is such that it cannot be created in a local way (i.e., in the absence of influences) if no information about $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is available beforehand. We can also put it this way: it is impossible for two computers $A$ and $B$ to be set up in such a way that, upon input of $\boldsymbol{\alpha}$ into $A$ and $\boldsymbol{\beta}$ into $B$, $A$ produces a random number $Z^1$ and $B$ $Z^2$ with joint distribution $\mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ if $A$ and $B$ cannot communicate (while they can use prepared random bits that both have copies of).[28] To put this yet differently, two suspects interrogated separately by police cannot provide answers $Z^1$ and $Z^2$ with distribution $\mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ when asked the questions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, no matter which prior agreement they took.

Bell's proof involves two parts. The first part is the EPR argument (in Bohm's version), applied to all directions $\boldsymbol{\alpha}$; it shows that if locality is true then the values of $Z^1$ and $Z^2$ must have been determined in advance. Thus, in every run of the experiment, there exist well-defined values $Z^1_{\boldsymbol{\alpha}}$ *for every* $\boldsymbol{\alpha}$ and $Z^2_{\boldsymbol{\alpha}} = -Z^1_{\boldsymbol{\alpha}}$ even before any measurement. Moreover, Alice's outcome will be $Z^1_{\boldsymbol{\alpha}}$ for the $\boldsymbol{\alpha}$ she chooses; also Bob's outcome will be $Z^2_{\boldsymbol{\beta}} = -Z^1_{\boldsymbol{\beta}}$ for the $\boldsymbol{\beta}$ he chooses, *also if* $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$ *and independently of whether Alice's or Bob's experiment occurs first.* (Put differently, the two suspects must have agreed in advance on the answer to every possible question.)

In other words, locality implies the existence of random variables $Z^i_{\boldsymbol{\alpha}}$, $i = 1, 2$ and $|\boldsymbol{\alpha}| = 1$, such that Alice's outcome is $Z^1_{\boldsymbol{\alpha}}$ and Bob's is $Z^2_{\boldsymbol{\beta}}$. In particular, focusing on components in only 3 directions $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, locality implies the existence of 6 random variables $Z^i_{\boldsymbol{\alpha}}$, $i = 1, 2$, $\boldsymbol{\alpha} = \mathbf{a},\ \mathbf{b},\ \mathbf{c}$ such that

$$Z^i_{\boldsymbol{\alpha}} = \pm 1 \tag{16.21}$$

$$Z^1_{\boldsymbol{\alpha}} = -Z^2_{\boldsymbol{\alpha}} \tag{16.22}$$

and, more generally,

$$\mathbb{P}(Z^1_{\boldsymbol{\alpha}} \neq Z^2_{\boldsymbol{\beta}}) = q_{\boldsymbol{\alpha}\boldsymbol{\beta}}, \tag{16.23}$$

where the $q_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}(+-) + \mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}(-+) = (1+\boldsymbol{\alpha}\cdot\boldsymbol{\beta})/2 = \cos^2(\theta/2)$ are the corresponding quantum mechanical probabilities.

The second part of the proof involves only very elementary mathematics. Clearly,

$$\mathbb{P}\left(\{Z^1_{\mathbf{a}} = Z^1_{\mathbf{b}}\} \cup \{Z^1_{\mathbf{b}} = Z^1_{\mathbf{c}}\} \cup \{Z^1_{\mathbf{c}} = Z^1_{\mathbf{a}}\}\right) = 1 \ , \tag{16.24}$$

since at least two of the three (2-valued) variables $Z^1_{\boldsymbol{\alpha}}$ must have the same value. Hence, by elementary probability theory,

$$\mathbb{P}\left(Z^1_{\mathbf{a}} = Z^1_{\mathbf{b}}\right) + \mathbb{P}\left(Z^1_{\mathbf{b}} = Z^1_{\mathbf{c}}\right) + \mathbb{P}\left(Z^1_{\mathbf{c}} = Z^1_{\mathbf{a}}\right) \geq 1, \tag{16.25}$$

and using the perfect anti-correlations (16.22) we have that

$$\mathbb{P}\left(Z^1_{\mathbf{a}} = -Z^2_{\mathbf{b}}\right) + \mathbb{P}\left(Z^1_{\mathbf{b}} = -Z^2_{\mathbf{c}}\right) + \mathbb{P}\left(Z^1_{\mathbf{c}} = -Z^2_{\mathbf{a}}\right) \geq 1. \tag{16.26}$$

[28] This statement is perhaps a bit less general than Bell's theorem because computers always work in either a deterministic or a stochastic way, while Bell's theorem would apply even to a theory, if it exists, that is neither deterministic nor stochastic.

(16.26) is equivalent to the celebrated *Bell inequality.* It is incompatible with (16.23). For example, when the angles between **a**, **b** and **c** are 120°, the 3 relevant $q_{\boldsymbol{\alpha\beta}}$ are all 1/4, implying a value of 3/4 for the left hand side of (16.26).

## 16.3 Bell's 1976 Proof of Nonlocality

Here is a different proof of nonlocality, first published by Bell in 1976;[29] it is also described in Bell's article "Bertlmann's socks." It was designed for the purpose of allowing small experimental errors in all probabilities, so that the perfect anti-correlation in the case $\theta = 0$ becomes merely a *near-perfect* anti-correlation, and the conclusion of pre-existing values cannot be drawn.[30]

Suppose that two computers produce outcomes $Z^1, Z^2$, each either $+1$ or $-1$, with joint distribution $\mathbb{P}(Z^1, Z^2|\boldsymbol{\alpha}, \boldsymbol{\beta})$ when given the input $\boldsymbol{\alpha}$ respectively $\boldsymbol{\beta}$. Let $\lambda$ be the information given in advance to both computers, such as an algorithm and random bits, and let $\rho$ be the probability distribution of $\lambda$. Then

$$\mathbb{P}(Z^1, Z^2|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \int d\lambda\, \rho(\lambda)\, \mathbb{P}(Z^1, Z^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)\,, \tag{16.27}$$

where the last factor is the conditional distribution of the outcomes, given $\lambda$.

What is the condition on $\mathbb{P}$ that characterizes the absence of communication? Suppose computer 1 makes its decision about $Z^1$ first. In the absence of communication, it has only $\lambda$ and $\boldsymbol{\alpha}$ as the basis of its decision (which may still be random); thus, the (marginal) distribution of $Z^1$ does not depend on $\boldsymbol{\beta}$:

$$\mathbb{P}(Z^1|\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) = \mathbb{P}(Z^1|\boldsymbol{\alpha}, \lambda)\,. \tag{16.28}$$

Computer 2 has only $\lambda$ and $\boldsymbol{\beta}$ as the basis of its decision; thus, the (conditional) distribution of $Z^2$ does not depend on $\boldsymbol{\alpha}$ or $Z^1$:

$$\mathbb{P}(Z^2|Z^1, \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) = \mathbb{P}(Z^2|\boldsymbol{\beta}, \lambda)\,. \tag{16.29}$$

From these two equations together, we obtain

$$\mathbb{P}(Z^1, Z^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) = \mathbb{P}(Z^1|\boldsymbol{\alpha}, \lambda)\, \mathbb{P}(Z^2|\boldsymbol{\beta}, \lambda) \tag{16.30}$$

as the characterization of locality (i.e., the absence of communication). Note that $Z^1$ and $Z^2$ can very well be *dependent* (correlated), like Bertlmann's socks or the glove left at home and the glove in my pocket, if the mutual dependence is based on their dependence on the common cause $\lambda$.

[29] J. S. Bell: The theory of local beables. *Epistemological Letters* **9**: 11 (1976)

[30] The advantage of robustness of the argument under small errors comes at the price that the argument needs to assume that the true theory of quantum mechanics is either deterministic or stochastic. I am unable to provide an example of a theory that is neither, but some authors (e.g., John H. Conway and Simon Kochen) have conjectured that the true laws of nature be neither; and Bell's original nonlocality proof, presented in Section 16.2 above, would apply also in that case.

Now we want to know how the locality condition (16.30) restricts the possibility of functions to occur as $\mathbb{P}(Z^1, Z^2|\boldsymbol{\alpha}, \boldsymbol{\beta})$. To this end, we introduce the *correlation coefficient* defined by

$$\kappa(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{z_1=\pm 1} \sum_{z_2=\pm 1} z_1 z_2 \mathbb{P}(Z^1 = z_1, Z^2 = z_2|\boldsymbol{\alpha}, \boldsymbol{\beta})\,. \tag{16.31}$$

**Proposition 16.1.** *Locality implies the following version of Bell's inequality known as the CHSH inequality*[31]*:*

$$\Big|\kappa(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \kappa(\boldsymbol{\alpha}, \boldsymbol{\beta}') + \kappa(\boldsymbol{\alpha}', \boldsymbol{\beta}) - \kappa(\boldsymbol{\alpha}', \boldsymbol{\beta}')\Big| \leq 2\,. \tag{16.32}$$

*Proof.* Locality (16.30) implies that

$$\begin{aligned}
\kappa(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \int d\lambda\, \rho(\lambda) \sum_{z_1=\pm 1} \sum_{z_2=\pm 1} z_1 z_2 \mathbb{P}(Z^1 = z_1, Z^2 = z_2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) && (16.33)\\
&= \int d\lambda\, \rho(\lambda) \sum_{z_1=\pm 1} \sum_{z_2=\pm 1} z_1 z_2 \mathbb{P}(Z^1|\boldsymbol{\alpha}, \lambda)\mathbb{P}(Z^2|\boldsymbol{\beta}, \lambda) && (16.34)\\
&= \int d\lambda\, \rho(\lambda)\, \mathbb{E}(Z^1|\boldsymbol{\alpha}, \lambda)\, \mathbb{E}(Z^2|\boldsymbol{\beta}, \lambda) && (16.35)
\end{aligned}$$

Since $Z^i \in \{1, -1\}$, we have that

$$\Big|\mathbb{E}(Z^1|\boldsymbol{\alpha}, \lambda)\Big| \leq 1 \quad \text{and} \quad \Big|\mathbb{E}(Z^2|\boldsymbol{\beta}, \lambda)\Big| \leq 1\,. \tag{16.36}$$

So,

$$\begin{aligned}
\Big|\kappa(\boldsymbol{\alpha}, \boldsymbol{\beta}) \pm \kappa(\boldsymbol{\alpha}, \boldsymbol{\beta}')\Big| &= \Big|\int d\lambda\, \rho(\lambda)\, \mathbb{E}(Z^1|\boldsymbol{\alpha}, \lambda)\Big(\mathbb{E}(Z^2|\boldsymbol{\beta}, \lambda) \pm \mathbb{E}(Z^2|\boldsymbol{\beta}', \lambda)\Big)\Big| && (16.37)\\
&\leq \int d\lambda\, \rho(\lambda)\Big|\mathbb{E}(Z^2|\boldsymbol{\beta}, \lambda) \pm \mathbb{E}(Z^2|\boldsymbol{\beta}', \lambda)\Big|\,. && (16.38)
\end{aligned}$$

Now for any $u, v \in [-1, 1]$,

$$|u + v| + |u - v| \leq 2 \tag{16.39}$$

because

$$\begin{aligned}
(u + v) + (u - v) &= 2u \leq 2 & (-u - v) + (u - v) &= -2v \leq 2 && (16.40)\\
(u + v) + (v - u) &= 2v \leq 2 & (-u - v) + (v - u) &= -2u \leq 2\,. && (16.41)
\end{aligned}$$

[31] This version (though with a different derivation making stronger assumptions) first appeared in J. F. Clauser, R. A. Holt, M. A. Horne, A. Shimony: Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters* **23**: 880–884 (1969)

Hence, setting $u = \mathbb{E}(Z^2|\boldsymbol{\beta}, \lambda)$ and $v = \mathbb{E}(Z^2|\boldsymbol{\beta}', \lambda)$,

$$\begin{aligned}
&\Big|\kappa(\boldsymbol{\alpha},\boldsymbol{\beta}) + \kappa(\boldsymbol{\alpha},\boldsymbol{\beta}') + \kappa(\boldsymbol{\alpha}',\boldsymbol{\beta}) - \kappa(\boldsymbol{\alpha}',\boldsymbol{\beta}')\Big| \\
&\qquad \leq \Big|\kappa(\boldsymbol{\alpha},\boldsymbol{\beta}) + \kappa(\boldsymbol{\alpha},\boldsymbol{\beta}')\Big| + \Big|\kappa(\boldsymbol{\alpha}',\boldsymbol{\beta}) - \kappa(\boldsymbol{\alpha}',\boldsymbol{\beta}')\Big| && (16.42)\\
&\qquad \overset{(16.38)}{\leq} \int d\lambda\, \rho(\lambda) \left(\Big|\mathbb{E}(Z^2|\boldsymbol{\beta},\lambda) + \mathbb{E}(Z^2|\boldsymbol{\beta}',\lambda)\Big| + \Big|\mathbb{E}(Z^2|\boldsymbol{\beta},\lambda) - \mathbb{E}(Z^2|\boldsymbol{\beta}',\lambda)\Big|\right) && (16.43)\\
&\qquad \overset{(16.39)}{\leq} 2\,. && (16.44)
\end{aligned}$$

□

Since the quantum mechanical prediction $\mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ for the Bell experiment has

$$\kappa(\boldsymbol{\alpha},\boldsymbol{\beta}) = \mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}(++) - \mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}(+-) - \mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}(-+) + \mu_{\boldsymbol{\alpha},\boldsymbol{\beta}}(--) = -\boldsymbol{\alpha}\cdot\boldsymbol{\beta} = -\cos\theta\,, \quad (16.45)$$

setting (in some plane)

$$\boldsymbol{\alpha} = 0^\circ\,, \quad \boldsymbol{\alpha}' = 90^\circ\,, \quad \boldsymbol{\beta} = 45^\circ\,, \quad \boldsymbol{\beta}' = -45^\circ \qquad (16.46)$$

leads to

$$\kappa(\boldsymbol{\alpha},\boldsymbol{\beta}) + \kappa(\boldsymbol{\alpha},\boldsymbol{\beta}') + \kappa(\boldsymbol{\alpha}',\boldsymbol{\beta}) - \kappa(\boldsymbol{\alpha}',\boldsymbol{\beta}') = -2\sqrt{2}\,, \qquad (16.47)$$

violating (16.32).

Now if the values of $\mathbb{P}(Z^1 = z_1, Z^2 = z_2|\boldsymbol{\alpha},\boldsymbol{\beta})$ are known only with some inaccuracy (because they were obtained experimentally, not from the quantum formalism) then also the $\kappa(\boldsymbol{\alpha},\boldsymbol{\beta})$ are subject to some inaccuracy. But if (16.32) is violated by more than the inaccuracy, then locality is refuted.

## 16.4 Photons

Experimental tests of Bell's inequality are usually done with photons instead of electrons. For photons, spin is usually called *polarization*, and the Stern–Gerlach magnets are replaced with *polarization analyzers* (also known as *polarizers*), i.e., crystals that are transparent to the $|z\text{-up}\rangle$ part of the wave but reflect (or absorb) the $|z\text{-down}\rangle$ part. Like the Stern–Gerlach magnets, the analyzers can be rotated into any direction. Since photons have spin 1, $\theta/2$ needs to be replaced by $\theta$.

# 17 Further Discussion of Nonlocality

## 17.1 Nonlocality in Bohmian Mechanics, GRW, Copenhagen, Many-Worlds

Since we have considered only non-relativistic formulations of these theories, we cannot directly analyze spacelike separated events, but instead we can analyze the case of two systems (e.g., Alice's lab and Bob's lab) without interaction (i.e., without an interaction term between them in the Hamiltonian).

- **Bohmian mechanics** is explicitly nonlocal, as the velocity of particle 2 depends on the position of particle 1, *no matter how distant and no matter whether there is interaction*. That is where the superluminal influence occurs. (Historically, Bell's nonlocality analysis was inspired by the examination of Bohmian mechanics.)

  This influence depends on entanglement: In the absence of entanglement, the velocity of particle 2 is independent of the position of particle 1. The fact that Bohmian mechanics is local for disentangled wave functions shows that it was necessary for proving non-locality to consider at least two particles and an entangled wave function (such as the singlet state). It can be shown that any entangled wave function violates Bell's inequality for some observables.

  Furthermore, the position of particle 1 will depend on the external fields at work near particle 1. That is, for any given initial position of particle 1, its later position will depend on the external fields. An example of an external field is the field of the Stern–Gerlach magnet. To a large extent, we can control external fields at our whim; e.g., we can rotate the Stern–Gerlach magnet. Bohm's equation of motion implies that these fields have an instantaneous influence on the motion of particle 2.

- In **GRW theory**, nonlocality comes in at the point when the wave function collapses, as then it does so *instantaneously over arbitrary distances*.

  At least, this trait of the theory *suggests* that GRW is nonlocal, and in fact that is the ultimate source of the nonlocality. Strictly speaking, however, the definition of nonlocality, i.e., the negation of (16.2), requires that *events* at $x$ and at $y$ influence each other, and the value of the wave function $\psi_t(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is linked to *several* space-time points, $(t, \boldsymbol{x}_1)$ and $(t, \boldsymbol{x}_2)$, and thus is not an example of an "event at $x$." So we need to formulate the proof that GRW theory is nonlocal more carefully; of course, Bell's proof achieves this, but we can give a more direct proof. Since the "events at $x$" are not given by the wave function itself but by the primitive ontology, we need to consider GRWf and GRWm separately.

  In GRWf, consider Einstein's boxes example. The wave function of a particle is half in a box in Paris and half in a box in Tokyo. Let us apply detectors to both boxes at time $t$, and consider the macroscopic superposition of the detectors arising from the Schrödinger equation. It is random whether the first flash (in any detector) after $t$ occurs in Paris or in Tokyo. Suppose it occurs in Tokyo, and suppose it can occur in one of two places in Tokyo, corresponding to the outcomes 0 or 1. If it was 1, then after the collapse the wave function of the particle is 100% in Tokyo, and later flashes in Paris are certain to occur in a place where they indicate the outcome 0—that is a nonlocal influence of a flash in Tokyo on the flashes in Paris.

Likewise in GRWm: If, after the first collapse, the pointer of the detector in Tokyo, according to the $m$ function, points to 1 then the pointer in Paris immediately points to 0. (You might object that the Tokyo pointer position according to the $m$ function was not the *cause* of the Paris pointer position, but rather both pointer positions were caused by the collapse of the wave function. However, this distinction is not relevant to whether the theory is nonlocal.)

Note that while Bell's proof shows that *any* version of quantum mechanics must be nonlocal, for proving that *GRWf and GRWm* are nonlocal it is sufficient to consider a simpler situation, that of Einstein's boxes.

Both GRWf and GRWm are already nonlocal when governing a universe containing only one particle; thus, their nonlocality does not depend on the existence of a macroscopic number of particles, and they are even nonlocal in a case (one particle) in which Bohmian mechanics is local. For example, consider a particle with wave function

$$\psi = \frac{1}{\sqrt{2}}\big(|\text{here}\rangle + |\text{there}\rangle\big) \tag{17.1}$$

at time $t$, as in Einstein's boxes example. Suppose that $|\text{here}\rangle$ and $|\text{there}\rangle$ are two narrow wave packets separated by a distance of 500 million light years. The distance is so large that the first collapse is likely to occur before a light signal can travel between the two places. For GRWf, a flash here precludes a flash there—that is a nonlocal influence. For GRWm, if the wave function collapses to $|\text{here}\rangle$ then $m(\text{here})$ doubles and $m(\text{there})$ instantaneously goes to zero—that is a nonlocal influence. (There is a relativistic version of GRWm[32] in which $m(\text{there})$ goes to zero only after a delay of distance/$c$, or when a collapse centered "there" occurs. Nevertheless, also this theory is nonlocal even for one particle because when a collapse centered "there" occurs, which can happen any time, then $m(\text{there})$ cannot double (as it could in a local theory) but must go to zero.)

- That **orthodox quantum mechanics** (OQM) is nonlocal can also be seen from Einstein's boxes argument: OQM says the outcomes of the detectors are not predetermined. (That is, there is no fact about where the particle really is before any detectors are applied.) Thus, the outcome of the Tokyo detector must have influenced the Paris detector, or vice versa.

[32] D. Bedingham, D. Dürr, G.C. Ghirardi, S. Goldstein, R. Tumulka, and N. Zanghì: Matter Density and Relativistic Models of Wave Function Collapse. *Journal of Statistical Physics* **154**: 623–631 (2014) `http://arxiv.org/abs/1111.1425`

This, of course, was the point of Einstein's boxes argument: He objected to OQM because it is nonlocal.

- **Many-worlds** is nonlocal, too. This is not obvious from Bell's argument because the latter is formulated in a single-world framework. Here is why Sm is nonlocal.[33] After Alice carries out her Stern–Gerlach experiment, there are two pointers in her lab, one pointing to $+1$ and the other to $-1$. Then Bob carries out his experiment, and there are two pointers in his lab. Suppose Bob chose the same direction as Alice. Then the world in which Alice's pointer points to $+1$ is the same world as the one in which Bob's pointer points to $-1$, and this nonlocal fact was created in a nonlocal way by Bob's experiment. The same kind of nonlocality occurs in Sm already in Einstein's boxes experiment: The world in which a particle was detected in Paris is the same as the one in which no particle was detected in Tokyo—a nonlocal fact that arises as soon as both experiments are completed, without the need to wait for the time it takes light to travel from Paris to Tokyo.

  How about Bell's many-worlds theories? The second theory, involving a random configuration selected independently at every time, is very clearly nonlocal, for example in Einstein's boxes experiment: At every time $t$, nature makes a random decision about whether the particle is in Paris, and if it is, nature ensures immediately that there is no particle in Tokyo. A local theory would require that the particle has a continuous history of traveling, at a speed less than that of light, to either Paris or Tokyo, and this history is missing in Bell's second many-worlds theory. Bell's first many-worlds theory is even more radical, in fact in such a way that the concept of locality is not even applicable. The concept of locality requires that at every point in space, there are local variables whose changes propagate at most at the speed of light. Since in Bell's first many-worlds theory, no association is made between worlds at different times, one cannot even ask how any local variables would change with time. Thus, this theory is nonlocal as well.

Another remark concerns the connection between Bell's 1976 nonlocality proof and the theories mentioned above. In physical theories, $\lambda$ represents the information located at all space-time points from which light signals can reach both $x$ and $y$. In orthodox quantum mechanics and GRW theory, $\lambda$ is the wave function $\psi$; in Bohmian mechanics, $\lambda$ is $\psi$ together with the initial configuration of the two particles.

## 17.2 Popular Myths About Bell's Proof

Let P be the hypothesis that, prior to any experiment, there exist values $Z^i_{\boldsymbol{n}}$ (for all $i = 1, 2$ and $\boldsymbol{n} \in \mathbb{R}^3$ with $|\boldsymbol{n}| = 1$) such that Alice and Bob obtain as outcomes $Z^1_{\boldsymbol{\alpha}}$ and $Z^2_{\boldsymbol{\beta}}$. These values are often called *hidden variables.* Then Bell's nonlocality argument,

[33] The argument is taken from V. Allori, S. Goldstein, R. Tumulka, and N. Zanghì: Many-Worlds and Schrödinger's First Quantum Theory. *British Journal for the Philosophy of Science* **62(1)**: 1–27 (2011) `http://arxiv.org/abs/0903.2211`

described in Section 16.2, has the following structure:

$$\begin{array}{rccl} \text{Part 1:} & \text{quantum mechanics} + \text{locality} & \Rightarrow & \text{P} \end{array} \tag{17.2}$$

$$\begin{array}{rccl} \text{Part 2:} & \text{quantum mechanics} & \Rightarrow & \text{not P} \end{array} \tag{17.3}$$

$$\begin{array}{rccl} \text{Conclusion:} & \text{quantum mechanics} & \Rightarrow & \text{not locality} \end{array} \tag{17.4}$$

For this argument what is relevant about "quantum mechanics" is merely the predictions concerning experimental outcomes corresponding to (16.21)–(16.23) (with part 1 using in fact only (16.22)).

Certain popular myths about Bell's proof arise from missing part 1 and noticing only part 2 of the argument. (In Bell's 1964 paper, part 1 is formulated in 3 lines, part 2 in 2.5 pages.) Bell, *Speakable and unspeakable*, p. 143:

> It is important to note that to the limited degree to which *determinism* plays a role in the EPR argument, it is not assumed but *inferred.* What is held sacred is the principle of 'local causality' – or 'no action at a distance'. [...] It is remarkably difficult to get this point across, that determinism is not a *presupposition* of the analysis.

Here, "determinism" means P. What Bell writes about the EPR argument is true in spades about his own nonlocality argument: P plays a "limited role" because it is only an auxiliary statement, and non-P is not the upshot of the argument.

The mistake of missing part 1 leads to the impression that Bell proved that

$$\text{hidden variables are impossible,} \tag{17.5}$$

or that

$$\text{hidden variables, while perhaps possible, must be nonlocal.} \tag{17.6}$$

These claims are still widespread, and were even more common in the 20th century.[34] They are convenient for Copenhagenists, who tend to think that coherent theories of the microscopic realm are impossible (see Section 13.3). Let me explain what is wrong about (17.5) and (17.6).

Statement (17.5) is plainly wrong, since a deterministic hidden-variables theory exists and works, namely Bohmian mechanics. The hidden variables that Bohmian mechanics provides[35] for the Bell experiment are of the form $Z^i_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, as the outcome according to Bohmian mechanics depends on *both* parameter choices (at least for one $i$, namely for the second Stern–Gerlach experiment). Considering the three directions relevant to Bell's inequality, the $Z^i_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ are 18 random variables instead of 6 $Z^i_{\boldsymbol{\alpha}}$, and the dependence on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ reflects the nonlocality of Bohmian mechanics. Bell did not establish the impossibility of a deterministic reformulation of quantum theory, nor did he ever claim to have done so.

[34]For example, recall the title of Clauser et al.'s paper: Proposed Experiment to Test Local Hidden-Variable Theories. Other authors claimed that Bell's argument excludes "local realism."

[35]We assume a fixed temporal order of the two spin measurements, and that each is carried out as a Stern–Gerlach experiment.

Statement (17.6) is true and non-trivial but nonetheless rather misleading. It follows from (17.2) and (17.3) that *any* (single-world) account of quantum phenomena must be nonlocal, not just any hidden-variables account. Bell's argument shows that nonlocality is implied by the predictions of standard quantum theory itself. Thus, if nature is governed by these predictions (as has been confirmed in experiment), then *nature is nonlocal*.

# 18 POVMs: Generalized Observables

## 18.1 Definition

An observable is mathematically represented by a self-adjoint operator. A *generalized observable* is mathematically represented by a *positive-operator-valued measure (POVM)*.

**Definition 18.1.** An operator is called *positive* iff it is self-adjoint and all (generalized) eigenvalues are greater than or equal to zero. (In linear algebra, a positive operator is commonly called "positive semi-definite.") Equivalently, a bounded operator $A : \mathscr{H} \to \mathscr{H}$ is positive iff

$$\langle\psi|A|\psi\rangle \geq 0 \quad \text{for every } \psi \in \mathscr{H}\,. \tag{18.1}$$

The sum of two positive operators is again a positive operator, whereas the product of two positive operators is in general not even self-adjoint. Note that every projection is a positive operator.

As a first, rough definition, we can say the following: A *POVM* is a family of positive operators $E_z$ such that

$$\sum_z E_z = I\,. \tag{18.2}$$

(Refined definition later.)

**Example 18.2.**

1. $E_1 = \begin{pmatrix} 1/2 & \\ & 1/3 \end{pmatrix}, E_2 = \begin{pmatrix} 1/2 & \\ & 2/3 \end{pmatrix}$.

   In fact, all (generalized) eigenvalues of $E_z$ must lie in $[0,1]$ because if $E_\zeta\psi = \eta\psi$, then

   $$\langle\psi|\psi\rangle = \langle\psi|I|\psi\rangle \overset{(18.2)}{=} \langle\psi|E_\zeta|\psi\rangle + \langle\psi|\sum_{z\neq\zeta} E_z|\psi\rangle \geq \langle\psi|E_\zeta|\psi\rangle = \eta\langle\psi|\psi\rangle\,, \tag{18.3}$$

   so $\eta \leq 1$.

2. $E_1 = \begin{pmatrix} 1 & \\ & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 0 & \\ & 1 \end{pmatrix}$. In the special case in which all operators $E_z$ are projection operators, $E$ is called a *projection-valued measure (PVM)*. In this case, the subspaces to which $E_z$ and $E_{z'}$ ($z \neq z'$) project must be mutually orthogonal (homework problem).

3. Every self-adjoint matrix defines a PVM: Let $z = \alpha$ run through the eigenvalues of $A$ and let $E_\alpha$ be the projection to the eigenspace of $A$ with eigenvalue $\alpha$,

   $$E_\alpha = \sum_\lambda |\phi_{\alpha,\lambda}\rangle\langle\phi_{\alpha,\lambda}|\,. \tag{18.4}$$

   Then their sum is $I$, as easily seen from the point of view of an orthonormal basis of eigenvectors of $A$. So $E$ is a PVM, the *spectral PVM* of $A$. Example 2 above is of this form for $A = \sigma_3$.

4. A POVM $E$ and a vector $\psi \in \mathscr{H}$ with $\|\psi\| = 1$ together define a probability distribution over $z$ as follows:

$$\mathbb{P}_\psi(z) = \langle \psi | E_z | \psi \rangle \,. \tag{18.5}$$

   To see this, note that $\langle \psi | E_z | \psi \rangle$ is a nonnegative real number since $E_z$ is a positive operator, and

$$\sum_z \mathbb{P}_\psi(z) = \sum_z \langle \psi | E_z | \psi \rangle = \langle \psi | I | \psi \rangle = \|\psi\|^2 = 1. \tag{18.6}$$

5. Fuzzy position observable:

$$E_z \psi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}} \psi(x) \,. \tag{18.7}$$

   Each $E_z$ is a positive operator (but not a projection) because

$$\langle \psi | E_z | \psi \rangle = \int dx \, \psi^*(x) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}} \psi(x) \geq 0 \,. \tag{18.8}$$

   The $E_z$ add to unity in the continuous sense:

$$\int E_z \, dz = I \,. \tag{18.9}$$

   Indeed,

$$\int dz \, E_z \psi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \psi(x) \int dz \, e^{-\frac{(x-z)^2}{2\sigma^2}} = \psi(x) \,. \tag{18.10}$$

The case of a continuous variable $z$ brings us to the general definition of a POVM, which I will formulate rigorously although we do not aim at rigor in general. The definition is, in fact, quite analogous to the rigorous definition of a probability distribution in measure theory: A *measure* associates a value (i.e., a number or an operator) not with a point but with a set: $E(B)$ instead of $E_z$, where $B \subseteq \mathscr{Z}$ and $\mathscr{Z}$ is the set of all $z$'s. More precisely, let $\mathscr{Z}$ be a set and $\mathscr{B}$ a $\sigma$-algebra of subsets of $\mathscr{Z}$,[36] the family of the "measurable sets." A *probability measure* is a mapping $\mu : \mathscr{B} \to [0,1]$ such that for any $B_1, B_2, \ldots \in \mathscr{B}$ with $B_i \cap B_j = \emptyset$ for $i \neq j$,

$$\mu\Big(\bigcup_{n=1}^{\infty} B_n\Big) = \sum_{n=1}^{\infty} \mu(B_n) \,. \tag{18.11}$$

[36] A *$\sigma$-algebra* is a family $\mathscr{B}$ of subsets of $\mathscr{Z}$ such that $\emptyset \in \mathscr{B}$ and, for every $B_1, B_2, B_3, \ldots$ in $\mathscr{A}$ also $B_1^c := \mathscr{Z} \setminus B_1 \in \mathscr{B}$ and $B_1 \cup B_2 \cup \ldots \in \mathscr{B}$. It follows that $\mathscr{Z} \in \mathscr{B}$ and $B_1 \cap B_2 \cap \ldots \in \mathscr{B}$. A set $\mathscr{Z}$ equipped with a $\sigma$-algebra is also called a *measurable space*. The $\sigma$-algebra usually considered on $\mathbb{R}^n$ consists of the "Borel sets" and is called the "Borel $\sigma$-algebra."

**Definition 18.3.** A *POVM* on the measurable space $(\mathscr{Z},\mathscr{B})$ acting on the Hilbert space $\mathscr{H}$ is a mapping $E$ from $\mathscr{B}$ to the set of bounded operators on $\mathscr{H}$ such that each $E(B)$ is positive, $E(\mathscr{Z}) = I$, and for any $B_1, B_2, \ldots \in \mathscr{B}$ with $B_i \cap B_j = \emptyset$ for $i \neq j$,

$$E\Bigl(\bigcup_{n=1}^{\infty} B_n\Bigr) = \sum_{n=1}^{\infty} E(B_n)\,, \tag{18.12}$$

where the series on the right-hand side converges in the operator norm.[37]

It follows that a POVM $E$ and a vector $\psi \in \mathscr{H}$ with $\|\psi\| = 1$ together define a probability measure on $\mathscr{Z}$ as follows:

$$\mu_\psi(B) = \langle\psi|E(B)|\psi\rangle\,. \tag{18.13}$$

(Verify the definition of a probability measure.) Again, one defines a PVM to be a POVM such that every $E(B)$ is a projection. In the special case in which $\mathscr{Z}$ is a countable set and $\mathscr{B}$ consists of all subsets, any POVM satisfies

$$E(B) = \sum_{z\in B} E_z \tag{18.14}$$

with $E_z = E(\{z\})$, so in that case Definition 18.3 boils down to the earlier definition around (18.2). The fuzzy position observable of Example 5 corresponds to $\mathscr{Z} = \mathbb{R}$, $\mathscr{B}$ the Borel sets, and $E(B)$ the multiplication operator

$$E(B)\psi(x) = \int_B dz\, \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}}\, \psi(x)\,, \tag{18.15}$$

which multiplies by the function $1_B * g$, where $1_B$ is the characteristic function of $B$, $g$ is the Gaussian density function, and $*$ means convolution.

It turns out that every observable is a generalized observable; that is, every self-adjoint operator $A$ defines a PVM $E$ with $E(B)$ the projection to the so-called spectral subspace of $B$. If there is an ONB of eigenvectors of $A$, then the spectral subspace of $B$ is the closed span of all eigenspaces with eigenvalues in $B$; that is, in that case $E(\{z\})$ is the projection to the eigenspace of eigenvalue $z$ (and 0 if $z$ is not an eigenvalue). In the case of a general self-adjoint operator $A$, the following is a reformulation of the spectral theorem:

**Theorem 18.4.** *For every self-adjoint operator $A$ there is a uniquely defined PVM $E$ on the real line with the Borel $\sigma$-algebra (the "spectral PVM" of $A$) such that*

$$A = \int_{\mathbb{R}} \alpha\, E(d\alpha)\,. \tag{18.16}$$

[37] It is equivalent to merely demand that the series on the right-hand side converges weakly, i.e., that $\sum_n \langle\psi|E(B_n)|\psi\rangle$ converges for every $\psi \in \mathscr{H}$.

To explain the last equation: In the same way as one can define the integral $\int_{\mathscr{Z}} f(z)\,\mu(dz)$ of a measurable function $f:\mathscr{Z}\to\mathbb{R}$ relative to a measure $\mu$, one can define an operator-valued integral $\int_{\mathscr{Z}} f(z)\,E(dz)$ relative to a POVM $E$. Eq. (18.16) is a generalization of the relation

$$A=\sum_{\alpha}\alpha\,E_\alpha \tag{18.17}$$

for self-adjoint matrices $A$. If several self-adjoint operators $A_1,\ldots,A_n$ commute pairwise, then they can be diagonalized simultaneously, i.e., there is a PVM $E$ on $\mathbb{R}^n$ such that for every $k\in\{1,\ldots,n\}$,

$$A_k=\int_{\mathbb{R}^n}\alpha_k\,E(d\boldsymbol{\alpha})\,. \tag{18.18}$$

**Example 18.5.** The PVM diagonalizing the three position operators $X_1,X_2,X_3$ on $L^2(\mathbb{R}^3)$ is

$$E(B)\psi(\boldsymbol{x})=\begin{cases}\psi(\boldsymbol{x}) & \text{if } \boldsymbol{x}\in B\\ 0 & \text{if } \boldsymbol{x}\notin B\,,\end{cases} \tag{18.19}$$

mentioned before in (10.16). Equivalently, $E(B)$ is the multiplication by the characteristic function of $B$.

**Example 18.6.** It follows from the quantum formalism that if we make consecutive ideal quantum measurements of observables $A_1,\ldots,A_n$ (which need not commute with each other) at times $0<t_1<\ldots<t_n$ respectively on a system with initial wave function $\psi_0\in\mathscr{H}$ with $\|\psi_0\|=1$, then the joint distribution of the outcomes $Z_1,\ldots,Z_n$ is of the form

$$\mathbb{P}\Big((Z_1,\ldots,Z_n)\in B\Big)=\langle\psi_0|E(B)|\psi_0\rangle \tag{18.20}$$

for all (Borel) subsets $B\subseteq\mathbb{R}^n$, where $E$ is a POVM on $\mathbb{R}^n$. The precise version of this statement requires that each $A_k$ has purely discrete spectrum (or, equivalently, an ONB of eigenvectors in $\mathscr{H}$).

*Derivation*: In that case, the spectrum is at most countable, and the spectral decomposition can be written in the form

$$A_k=\sum_{\alpha_k}\alpha_k\,P_{k,\alpha_k}\,. \tag{18.21}$$

The probability of $Z_1=\alpha_1$ is $\|P_{1,\alpha_1}e^{-iHt_1}\psi_0\|^2$; the conditional probability of $Z_2=\alpha_2$, given that $Z_1=\alpha_1$, is $\|P_{2,\alpha_2}e^{-iH(t_2-t_1)}\psi_{t_1+}\|^2$ with $\psi_{t_1+}=P_{1,\alpha_1}e^{-iHt_1}\psi_0/\|P_{1,\alpha_1}e^{-iHt_1}\psi_0\|$. Putting these formulas together (and extending to $n$ measurements), we obtain that

$$\begin{aligned}&\mathbb{P}\Big((Z_1,\ldots,Z_n)=(\alpha_1,\ldots,\alpha_n)\Big)\\ &\quad=\Big\|P_{n,\alpha_n}e^{-iH(t_n-t_{n-1})}\cdots P_{1,\alpha_1}e^{-iH(t_1-t_0)}\psi_0\Big\|^2\end{aligned} \tag{18.22}$$

with $t_0 = 0$ (and units of measurement chosen so that $\hbar = 1$), so (18.20) holds with

$$E\big(\{(\alpha_1, \ldots, \alpha_n)\}\big) = \\ e^{iH(t_1-t_0)} P_{1,\alpha_1} \cdots e^{iH(t_n-t_{n-1})} P_{n,\alpha_n} P_{n,\alpha_n} e^{-iH(t_n-t_{n-1}} \cdots P_{1,\alpha_1} e^{-iH(t_1-t_0)} . \quad (18.23)$$

It becomes clear that $E(B)$ is, in general not a projection but still a positive operator. One easily verifies that $E$ is a POVM.

**Example 18.7.** In GRWf, the joint distribution of all flashes is of the form

$$\mathbb{P}(F \in B) = \langle \Psi_0 | G(B) | \Psi_0 \rangle \quad (18.24)$$

for all sets $B \subseteq \mathscr{Z}$, with $\Psi_0$ the initial wave function and $G$ a POVM on the history space $\mathscr{Z}$ of flashes,

$$\mathscr{Z} = \Big\{ \big( (t_1, \boldsymbol{x}_1), (t_2, \boldsymbol{x}_2), \ldots \big) \in (\mathbb{R}^4)^\infty : 0 < t_1 < t_2 < \ldots \Big\}^N . \quad (18.25)$$

*Derivation*: Consider first the joint distribution of the first two flashes for $N = 1$ particle: The probability of $T_1 \in [t_1, t_1 + dt_1]$ is $1_{t_1>0}\, e^{-\lambda t_1}\, \lambda\, dt_1$; given $T_1$, the probability of $\boldsymbol{X}_1 \in d^3\boldsymbol{x}_1$ is, according to (12.11), $\|C(\boldsymbol{x}_1)\Psi_{T_1-}\|^2$ with $\Psi_{T_1-} = e^{-iHT_1}\Psi_0$ and $C(\boldsymbol{x}_1)$ the collapse operator defined in (12.9). Given $T_1$ and $\boldsymbol{X}_1$, the probability of $T_2 \in [t_2, t_2 + dt_2]$ is $1_{t_2>t_1}\, e^{-\lambda(t_2-t_1)}\, \lambda\, dt_2$; given $T_1, \boldsymbol{X}_1$, and $T_2$, the probability of $\boldsymbol{X}_2 \in d^3\boldsymbol{x}_2$ is $\|C(\boldsymbol{x}_2)e^{-iH(T_2-T_1)}\Psi_{T_1+}\|^2$ with $\Psi_{T_1+} = C(\boldsymbol{X}_1)\Psi_{T_1-}$. Putting these formulas together, the joint distribution of $T_1, \boldsymbol{x}_1, T_2$, and $\boldsymbol{X}_2$ is given by

$$\mathbb{P}\Big(T_1 \in [t_1, t_1 + dt_1], \boldsymbol{X}_1 \in d^3\boldsymbol{x}_1, T_2 \in [t_2, t_2 + dt_2], \boldsymbol{X}_2 \in d^3\boldsymbol{x}_2\Big)$$

$$= 1_{0<t_1<t_2}\, e^{-\lambda t_2}\, \lambda^2 \Big\| C(\boldsymbol{x}_2) e^{-iH(t_2-t_1)} C(\boldsymbol{x}_1) e^{-iHt_1} \Psi_0 \Big\|^2 dt_1\, d^3\boldsymbol{x}_1\, dt_2\, d^3\boldsymbol{x}_2 \quad (18.26)$$

$$= \langle \Psi_0 | G(dt_1 \times d^3\boldsymbol{x}_1 \times dt_2 \times d^3\boldsymbol{x}_2) | \Psi_0 \rangle \quad (18.27)$$

with

$$G(dt_1 \times d^3\boldsymbol{x}_1 \times dt_2 \times d^3\boldsymbol{x}_2) = 1_{0<t_1<t_2}\, e^{-\lambda t_2}\, \lambda^2 \times \\ \times\; e^{iHt_1} C(\boldsymbol{x}_1) e^{iH(t_2-t_1)} C(\boldsymbol{x}_2)^2 e^{-iH(t_2-t_1)} C(\boldsymbol{x}_1) e^{-iHt_1}\, dt_1\, d^3\boldsymbol{x}_1\, dt_2\, d^3\boldsymbol{x}_2 , \quad (18.28)$$

which is self-adjoint and positive because (18.27) is always real and $\geq 0$. It follows that also $G(B)$, obtained by summing (that is, integrating) over all infinitesimal volume elements in $B$, is self-adjoint and positive. Additivity holds by construction, and $G(\mathscr{Z}) = I$ because (18.27) is a probability distribution (so $\langle \Psi_0 | G(\mathscr{Z}) | \Psi_0 \rangle = 1$ for every $\Psi_0$ with $\|\Psi_0\| = 1$). Thus, $G$ is a POVM. For the joint distribution of more than two flashes or more than one particle, the reasoning proceeds in a similar way. For the joint distribution of *all* (infinitely many) flashes, the rigorous proof requires some more technical steps[38] but bears no surprises.

[38] carried out in R. Tumulka: A Kolmogorov Extension Theorem for POVMs. *Letters in Mathematical Physics* **84**: 41–46 (2008) `http://arxiv.org/abs/0710.3605`

## 18.2 The Main Theorem about POVMs

It says: *For every quantum physical experiment $\mathscr{E}$ on a quantum system $S$ whose possible outcomes lie in a space $\mathscr{Z}$, there exists a POVM $E$ on $\mathscr{Z}$ such that, whenever $S$ has wave function $\psi$ at the beginning of $\mathscr{E}$, the random outcome $Z$ has probability distribution given by*

$$\mathbb{P}(Z \in B) = \langle\psi|E(B)|\psi\rangle \,. \tag{18.29}$$

We will prove this statement in Bohmian mechanics and GRWf. It plays the role of Born's rule for POVMs. The experiment $\mathscr{E}$ consists of coupling $S$ to an apparatus $A$ at some initial time $t_i$, letting $S \cup A$ evolve up to some final time $t_f$, and then reading off the result $Z$ from $A$. It is assumed that $S$ and $A$ are not entangled at the beginning of $\mathscr{E}$:

$$\Psi_{S\cup A}(t_i) = \psi_S(t_i) \otimes \phi_A(t_i) \tag{18.30}$$

with $\phi_A$ the ready state of $A$. (The main theorem of POVMs can also be proven for the case in which $t_f$ is itself chosen by the experiment; e.g., the experiment might wait for a detector to click, and the outcome $Z$ may be the time of the click. I give the proof only for the simpler case in which $t_f$ is fixed in advance.) I will further assume that $\mathscr{E}$ has only finitely many possible outcomes $Z$; actually, this assumption is not needed for the proof, but it simplifies the consideration a bit and is satisfied in every realistic scenario.

**Proof from Bohmian mechanics.** Since the outcome is read off from the pointer position,

$$Z = \zeta\big(Q(t_f)\big) \,, \tag{18.31}$$

where $Q$ is the Bohmian configuration and $\zeta$ is called the *calibration function*. (In practice, the function $\zeta$ depends only on the configuration *of the apparatus*, in fact only on its macroscopic features, not on microscopic details. However, the arguments that follow apply to arbitrary calibration functions.) Let

$$U = e^{-iH_{S\cup A}(t_f - t_i)} \tag{18.32}$$

and

$$B_z = \{q \in \mathbb{R}^{3N} : \zeta(q) = z\} \,. \tag{18.33}$$

Then, using the projection operator $P_B$ defined in (10.16),

$$\begin{aligned}
\mathbb{P}(Z = z) &= \mathbb{P}\big(Q(t_f) \in B_z\big) &(18.34)\\
&= \int_{B_z} |\Psi(q,t_f)|^2 \, dq &(18.35)\\
&= \langle\Psi(t_f)|P_{B_z}|\Psi(t_f)\rangle &(18.36)\\
&= \langle\psi\otimes\phi|U^\dagger P_{B_z} U|\psi\otimes\phi\rangle &(18.37)\\
&= \langle\psi|E_z|\psi\rangle_S \,, &(18.38)
\end{aligned}$$

where $\langle\cdot|\cdot\rangle_S$ denotes the inner product in the Hilbert space of the system $S$ alone (as opposed to the Hilbert space of $S \cup A$), and $E_z$ is defined as follows: For given $\psi$, form $\psi \otimes \phi$, then apply the operator $U^\dagger P_{B_z} U$, and finally take the *partial inner product* with $\phi$. The partial inner product of a function $\Psi(x, y)$ with the function $\phi(y)$ is a function of $x$ defined as

$$\langle\phi|\Psi\rangle_y(x) = \int dy\, \phi^*(y)\, \Psi(x, y)\,. \tag{18.39}$$

Thus,

$$E_z\psi = \langle\phi|U^\dagger P_{B_z} U(\psi \otimes \phi)\rangle_y\,. \tag{18.40}$$

We now verify that $E$ is a POVM. First, $E_z$ is a positive operator because

$$\langle\psi|E_z|\psi\rangle = \langle\Psi(t_f)|P_{B_z}|\Psi(t_f)\rangle \geq 0 \tag{18.41}$$

for every $\psi$. Second, $\sum_z E_z = I$ because

$$\sum_z E_z\psi = \sum_z \langle\phi|U^\dagger P_{B_z} U(\psi \otimes \phi)\rangle_y \tag{18.42}$$

$$= \langle\phi|U^\dagger \sum_z P_{B_z} U(\psi \otimes \phi)\rangle_y \tag{18.43}$$

$$= \langle\phi|U^\dagger I U(\psi \otimes \phi)\rangle_y \tag{18.44}$$

$$= \langle\phi|I(\psi \otimes \phi)\rangle_y = \psi\,. \tag{18.45}$$

Here, we have used that

$$\sum_z P_{B_z} = I\,, \tag{18.46}$$

that $U^\dagger U = I$, and that the partial inner product of $\psi \otimes \phi$ with $\phi$ returns $\psi$. Eq. (18.46) follows from the fact that the sets $B_z$ form a partition of configuration space $\mathbb{R}^{3N}$ (i.e., they are mutually disjoint and together cover the entire configuration space, $\cup_z B_z = \mathbb{R}^{3N}$). This, in turn, follows from the assumption that the calibration function $\zeta$ is defined everywhere in $\mathbb{R}^{3N}$.[39] Thus, the proof is complete. □

**Proof from GRWf.** Let $F = \{(T_1, \boldsymbol{X}_1), (T_2, \boldsymbol{X}_2), \ldots\}$ be the set of flashes (of both $S$ and $A$) from $t_i$ onwards. We know from Example 18.7 that the distribution of $F$ (i.e., the joint distribution of all flashes after $t_i$) is given by $\Psi(t_i)$ and some POVM $G$:

$$\mathbb{P}(F \in B) = \langle\Psi(t_i)|G(B)|\Psi(t_i)\rangle\,. \tag{18.47}$$

Since the outcome $Z$ of the experiment is read off from $A$ after $t_i$, it is a function of $F$,

$$Z = \zeta(F)\,. \tag{18.48}$$

[39] The physical meaning of this asumption is that the experiment always has *some* outcome. You may worry about the possibility that the experiment could not be completed as planned due to power outage, meteorite impact, or whatever. This possibility can be taken into account by introducing a further element $f$ for "failed" into the set $\mathscr{Z}$ of possible outcomes.

($Z$ is a function of $F$ because the flashes define where the pointers point, and what the shape of the ink on a sheet of paper is. It would even be realistic to assume that $Z$ depends only on the flashes of the apparatus, but this restriction is not needed for the further argument.)

Let $B_z = \{f : \zeta(f) = z\}$, the set of flash patterns having outcome $z$. Then,

$$\mathbb{P}(Z = z) = \mathbb{P}\big(F \in B_z\big) \tag{18.49}$$
$$= \langle \Psi(t_i)|G(B_z)|\Psi(t_i)\rangle \tag{18.50}$$
$$= \langle \psi|E_z^{\mathrm{GRW}}|\psi\rangle \tag{18.51}$$

with

$$E_z^{\mathrm{GRW}}\psi = \langle \phi|G(B_z)|\psi \otimes \phi\rangle_y \,. \tag{18.52}$$

In fact, $E_z^{\mathrm{GRW}}$ may be different from $E_z$ obtained from Bohmian mechanics as in (18.40), in agreement with the fact that the same experiment (using the same initial wave function of the apparatus, etc.) may yield different outcomes in GRW than in Bohmian mechanics. (However, since we know the two theories make very very similar predictions, $E_z^{\mathrm{GRW}}$ will usually be very very close to $E_z$.) To see that $E_z^{\mathrm{GRW}}$ is a POVM, we note that

$$\langle \psi|E_z^{\mathrm{GRW}}|\psi\rangle = \langle \Psi(t_1)|G(B_z)|\Psi(t_1)\rangle \geq 0 \tag{18.53}$$

and

$$\sum_z E_z^{\mathrm{GRW}}\psi = \langle \phi|\sum_z G(B_z)|\psi \otimes \phi\rangle_y \tag{18.54}$$
$$= \langle \phi|G(\cup_z B_z)|\psi \otimes \phi\rangle_y \tag{18.55}$$
$$= \langle \phi|I|\psi \otimes \phi\rangle_y = \psi \tag{18.56}$$

using $\cup_z B_z = \mathscr{X}$. This completes the proof. □

The main theorem about POVMs is equally valid in orthodox quantum mechanics (OQM). However, since OQM does not permit a coherent analysis of measurement processes (as it suffers from the measurement problem), we cannot give a complete proof of the main theorem from OQM, but the same reasoning as given in the proof from Bohmian mechanics would be regarded as compelling in OQM. At the same time, the main theorem undercuts the spirit of OQM, which is to leave the measurement process unanalyzed and to introduce observables by postulate. Put differently, the main theorem about POVMs makes it harder to ignore the measurement problem.

## 18.3 Limitations to Knowledge

**Corollary 18.8.** *There is no experiment with $Z = \psi$ or $Z = \mathbb{C}\psi$. That is, one cannot measure the wave function of a given system, not even up to a global phase.*

*Proof.* Suppose there was an experiment with $Z = \psi$. Then, for any given $\psi$, $Z$ is deterministic, i.e., its probability distribution is concentrated on a single point, $\mathbb{P}(Z = \phi) = \delta(\phi - \psi)$. The dependence of this distribution on $\psi$ is not quadratic, and thus not of the form $\langle\psi|E_\phi|\psi\rangle$ for any POVM $E$. The argument remains valid when we replace $\psi$ by $\mathbb{C}\psi$. □

This fact amounts to a limitation of knowledge in any version of quantum mechanics in which wave functions are part of the ontology, which includes all interpretations of quantum mechanics that we have talked about: Suppose Alice chooses a direction in space $\boldsymbol{n}$, prepares a spin-$\frac{1}{2}$ particle in the state $|\boldsymbol{n}\text{-up}\rangle$, and hands that particle over to Bob. Then, by Corollary 18.8, Bob has no way of discovering $\boldsymbol{n}$ if Alice does not give the information away. The best thing Bob can do is, in fact, a Stern–Gerlach experiment in any direction he likes, say in the $z$-direction; then he obtains one bit of information, up or down; if the result was "up" then it is more likely that $\boldsymbol{n}$ lies on the upper hemisphere than on the lower.

**Corollary 18.9.** *There is no experiment in Bohmian mechanics that can measure the instantaneous velocity of a particle with unknown wave function.*

*Proof.* Again, the distribution of the velocity $\mathrm{Im}\nabla\psi/\psi(Q)$ with $Q \sim |\psi|^2$ is not quadratic in $\psi$. □

In contrast, the *asymptotic velocity* can be measured, and its probability distribution is in fact quadratic in $\psi$: Recall from (7.39) that it is given by $(m/\hbar)^3\,|\widehat{\psi}(m\boldsymbol{u}/\hbar)|^2$.

The impossibility of measuring instantaneous velocity goes along with the impossibility to measure the entire trajectory without disturbing it. If we wanted to measure the trajectory, for example by repeatedly measuring the positions every $\Delta t$ with inaccuracy $\Delta x$, then the measurements will collapse the wave function, with the consequence that the observed trajectory is very different from what the trajectory would have been had we not intervened. Some authors regard this as an argument against Bohmian mechanics. Bell disagreed (*Speakable and unspeakable in quantum mechanics*, page 202):

> To admit things not visible to the gross creatures that we are is, in my opinion, to show a decent humility, and not just a lamentable addiction to metaphysics.

So, Bell criticized the positivistic idea that anything real can always be measured. Indeed, this idea seems rather dubious in view of Corollary 18.8. We will sharpen this consideration in Section 20.3.

## 18.4 The Concept of Observable

The main theorem about POVMs suggests that POVMs form the natural generalization of the notion of observables. It also allows us to explain what an observable ultimately is. Here is the natural general definition:

**Definition 18.10.** Two experiments (that can be carried out on arbitrary wave functions $\psi \in \mathscr{H}$ with norm 1) are *equivalent in law* iff for every $\psi \in \mathscr{H}$ with $\|\psi\| = 1$, they have the same distribution of the outcome. (Thus, they are equivalent in law iff they have the same POVM.) A corresponding equivalence class of experiments is called an *observable.*

If $\mathscr{E}_1$ and $\mathscr{E}_2$ are equivalent in law and a particular run of $\mathscr{E}_1$ has yielded the outcome $z_1$, it *cannot* be concluded that $\mathscr{E}_2$ would have yielded $z_1$ as well. The counterfactual question, "what would $z_2$ have been if we had run $\mathscr{E}_2$?" cannot be tested empirically, but it can be analyzed in Bohmian mechanics; there, one sometimes finds $z_2 \neq z_1$ (for the same $Q_S$ and $\psi$ in both experiments, but different $Q_A$ and $\phi$). For example, let $\mathscr{E}_1$ be a Stern–Gerlach experiment in the $z$ direction and $\mathscr{E}_2$ a Stern–Gerlach experiment in the $-z$ direction with the outcome called $+1$ if the particle is detected in the *down* channel and $-1$ if the particle is detected in the *up* channel. Then $\mathscr{E}_1$ and $\mathscr{E}_2$ are equivalent in law, although in Bohmian mechanics, the two experiments will often yield different results when applied to the same 1-particle wave function and position.

This situation illustrates why the term "observable" can be rather misleading: It is intended to suggest "observable quantity," but an observable is not even a well-defined quantity to begin with (as the outcome $Z$ depends on $Q_A$ and $\phi$), it is *a class of experiments with equal probability distributions.*

This point is connected to Wheeler's fallacy. Recall the delayed choice experiment, but now consider detecting the particle either directly at the slits or far away, ignoring the interference region. As $\mathscr{E}_1$, we put detectors directly at the slits and say that the outcome is $Z_1 = +1$ if the particle was detected in the left slit and $Z_1 = -1$ if in the right. This is a kind of position measurement that can be represented in the 2d Hilbert space formed by wave functions of the form

$$\psi = c_1|\text{left slit}\rangle + c_2|\text{right slit}\rangle\,, \tag{18.57}$$

so $\mathbb{P}(Z_1 = +1) = |c_1|^2$. Relative to the basis $\{|\text{left slit}\rangle, |\text{right slit}\rangle\}$, the POVM is the spectral PVM of $\sigma_3$. As $\mathscr{E}_2$, we put the detectors far away and say that $Z_2 = +1$ if the particle was detected in the far right and $Z_2 = -1$ if in the far left. $\psi$ evolves to

$$\psi' = c_1|\text{far right}\rangle + c_2|\text{far left}\rangle\,, \tag{18.58}$$

so $\mathbb{P}(Z_2 = +1) = |c_1|^2$. So, $Z_1$ and $Z_2$ have the same distribution, $\mathscr{E}_1$ and $\mathscr{E}_2$ have the same POVM, and the two experiments are equivalent in law, although we know that the Bohmian particle often passes through the right slit and still ends up on the far right.

Now comes the point that has confused a number of authors[40]: Since $\mathscr{E}_1$ measures the "position observable," and since $\mathscr{E}_1$ and $\mathscr{E}_2$ "measure" the same observable, it is clear that $\mathscr{E}_2$ also measures the position observable. People concluded that $\mathscr{E}_2$ "measures through which slit the particle went"—Wheeler's fallacy! People concluded further that since the Bohmian trajectory may pass through the left slit while $Z_2 = -1$, Bohmian

[40]For example (using a different but similar setup), B.-G. Englert, M.O. Scully, G. Süssmann, and H. Walther: Surrealistic Bohm Trajectories. *Zeitschrift für Naturforschung A* **47**: 1175–1186 (1992)

mechanics must somehow disagree with measured facts about which slit the particle went through. Bad, bad Bohm!

# 19 Time of Detection

## 19.1 The Problem

Suppose we set up a detector, wait for the arrival of the particle at the detector, and measure the time $T$ at which the detector clicks. What is the probability distribution of $T$? This is a natural question not covered by the usual quantum formalism because there is no self-adjoint operator for time. But from the main theorem about POVMs it is clear that there must be a POVM $E$ such that

$$\mathbb{P}(T \in B) = \langle \psi_0 | E(B) | \psi_0 \rangle \,. \tag{19.1}$$

That is, time of detection is a generalized observable. In this section we take a look at this POVM $E$.

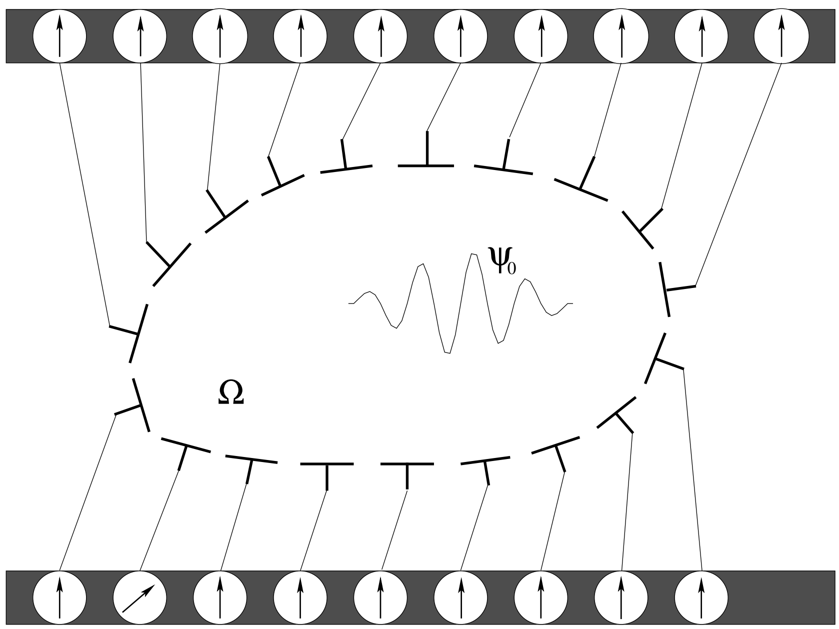


Figure 4: A quantum particle in a region $\Omega$ surrounded by a surface $\Sigma = \partial\Omega$ made out of detectors (symbolized by $\perp$'s), each of which is connected to a pointer. In part, the figure depicts the situation before the experiment, as the initial wave function $\psi_0$ is symbolized by a wave, and in part the situation after the experiment, as the location of detection is indicated by one pointer in the triggered position. Figure adapted from page 347 of D. Dürr and S. Teufel: *Bohmian mechanics*, Springer-Verlag (2009)

Suppose that we form a surface $\Sigma \subset \mathbb{R}^3$ out of little detectors so we can measure the time and the location at which the quantum particle first crosses $\Sigma$. Suppose further that, as depicted in Figure 4, $\Sigma$ divides physical space $\mathbb{R}^3$ into two regions, $\Omega$ and its complement, and the particle's initial wave function $\psi_0$ is concentrated in $\Omega$. The outcome of the experiment is the pair $Z = (T, \boldsymbol{X})$ of the time $T \in [0, \infty)$ of detection and the location $\boldsymbol{X} \in \Sigma$ of detection; should no detection ever occur, then we write $Z = \infty$. So the value space of $E$ is $\mathscr{Z} = [0, \infty) \times \Sigma \cup \{\infty\}$. We want to compute the distribution of $Z$ from $\psi_0$.

Let us compare the problem to Born's rule. In Born's rule, we choose a time $t_0$ and measure the three position coordinates at time $t_0$; here, if we take $\Omega$ to be the half space $\{(x, y, z) : x > x_0\}$ and $\Sigma$ its boundary plane $\{(x, y, z) : x = x_0\}$, then we choose the value of one position coordinate $(x_0)$ and measure the time as well as the other two position coordinates when the particle reaches that value. Put differently in terms of space-time $\mathbb{R}^4 = \{(t, x, y, z)\}$, Born's rule concerns measuring where the particle intersects the spacelike hypersurface $\{t = t_0\}$, and our problem concerns measuring where the particle intersects the timelike hypersurface $\{x = x_0\}$. We could say that we need a Born rule for timelike hypersurfaces.

I should make three caveats, though.

- I have used language such as "particle arriving at a surface" that presupposes the existence of trajectories although we know that some theories of quantum mechanics (GRWm and GRWf) claim that there are no trajectories, and still these theories are approximately empirically equivalent to Bohmian mechanics, so the time and location of the detector click would have approximately the same distribution as in Bohmian mechanics. Our problem really concerns the distribution of the detection events, and we should keep in mind that in some theories the trajectory language cannot be taken seriously.

- Even in Bohmian mechanics, there is a crucial difference between the case with the spacelike hypersurface and the one with the timelike hypersurface: The point where the particle arrives on the timelike hypersurface $\{x = x_0\}$ may depend on whether or not detectors are present on that hypersurface. A detector that does not click may still affect $\psi$ and thus the future particle trajectory. That is why I avoid the expression "time of arrival" (which is often used in the literature) in favor of "time of detection." In contrast, the point where the particle arrives at the spacelike hypersurface $\{t = t_0\}$ does not depend on whether or not detectors are placed along $\{t = t_0\}$.

- The exact POVM $E$ is given by (18.40) (with $t_f$ some late time at which we read off the values of $T$ and $\boldsymbol{X}$ recorded by the apparatus) and will depend on the exact wave function of the detectors, so different detectors will lead to slightly different POVMs. Of course, we expect that these differences are negligible. What we want is a simple rule defining the POVM for an *ideal* detector, $E_{\mathrm{ideal}}$. That, of course, involves making a definition of what counts as an ideal detector. So the formula for $E_{\mathrm{ideal}}$ is in part a matter of definition, as long as it fits well with the POVMs $E$ of real detectors.

## 19.2 The Absorbing Boundary Rule

The question of what $E_{\mathrm{ideal}}$ is is not fully settled; I will describe the most plausible proposal, the *absorbing boundary rule*.[41] Such a rule was for a long time believed to be impossible because of the quantum Zeno effect and Allcock's paradox (see homework exercises). Henceforth I will write $E$ instead of $E_{\mathrm{ideal}}$. Let $\Sigma = \partial\Omega$, $\psi_0$ be concentrated in $\Omega$, $\|\psi_0\| = 1$, and let $\kappa > 0$ be a constant of dimension 1/length (it will be a parameter of the detector). Here is the rule:

Solve the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + V\psi \tag{19.2}$$

in $\Omega$ with potential $V : \Omega \to \mathbb{R}$ and boundary condition

$$\frac{\partial\psi}{\partial n}(\boldsymbol{x}) = i\kappa\psi(\boldsymbol{x}) \tag{19.3}$$

at every $\boldsymbol{x} \in \Sigma$, with $\partial/\partial n$ the outward normal derivative on the surface, $\partial\psi/\partial n := \boldsymbol{n}(\boldsymbol{x}) \cdot \nabla\psi(\boldsymbol{x})$ with $\boldsymbol{n}(\boldsymbol{x})$ the outward unit normal vector to $\Sigma$ at $\boldsymbol{x} \in \Sigma$. Then, the rule asserts,

$$\mathbb{P}_{\psi_0}\Big(t_1 \leq T < t_2, \boldsymbol{X} \in B\Big) = \int_{t_1}^{t_2} dt \int_B d^2\boldsymbol{x}\; \boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{j}^{\psi_t}(\boldsymbol{x}) \tag{19.4}$$

for any $0 \leq t_1 < t_2$ and any set $B \subseteq \Sigma$, with $d^2\boldsymbol{x}$ the surface area element and $\boldsymbol{j}^\psi$ the probability current vector field (2.16). In other words, the joint probability density of $T$ and $\boldsymbol{X}$ relative to $dt\, d^2\boldsymbol{x}$ is the normal component of the current across the boundary, $j_n^{\psi_t}(\boldsymbol{x}) = \boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{j}^{\psi_t}(\boldsymbol{x})$. Furthermore,

$$\mathbb{P}_{\psi_0}(Z = \infty) = 1 - \int_0^\infty dt \int_\Omega d^2\boldsymbol{x}\; \boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{j}^{\psi_t}(\boldsymbol{x})\,. \tag{19.5}$$

This completes the statement of the rule.

Let us study the properties of the rule. To begin with, the boundary condition (19.3) implies that the current vector $\boldsymbol{j}$ at the boundary is always outward-pointing: For every $\boldsymbol{x} \in \Sigma$,

$$\boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{j}(\boldsymbol{x}) = \tfrac{\hbar}{m}\mathrm{Im}\Big(\psi(\boldsymbol{x})^*\frac{\partial\psi}{\partial n}(\boldsymbol{x})\Big) = \tfrac{\hbar}{m}\mathrm{Im}\Big(\psi(\boldsymbol{x})^* i\kappa\psi(\boldsymbol{x})\Big) = \frac{\hbar\kappa}{m}|\psi(\boldsymbol{x})|^2 \geq 0\,. \tag{19.6}$$

[41] R. Werner: Arrival time observables in quantum mechanics. *Annales de l'Institut Henri Poincaré, section A* **47**: 429–449 (1987)
R. Tumulka: Distribution of the Time at Which an Ideal Detector Clicks. (2016) `http://arxiv.org/abs/1601.03715`

For this reason, (19.3) is called an *absorbing boundary condition*: It implies that there is never any current coming out of the boundary. In particular, the right-hand side of (19.4) is non-negative.

So the rule invokes a new kind of time evolution for a 1-particle wave function as an effective treatment of the whole system formed by the 1 particle and the detectors together. It is useful to picture the Bohmian trajectories for this time evolution. Eq. (19.6) implies that the Bohmian velocity field $\boldsymbol{v}(\boldsymbol{x})$ is always outward-pointing at the boundary, $\boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{v}(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \Sigma$; in fact, the normal velocity is prescribed, $\boldsymbol{n}(\boldsymbol{x})\cdot\boldsymbol{v}(\boldsymbol{x}) = \hbar\kappa/m$. In particular, Bohmian trajectories can cross $\Sigma$ only in the outward direction; when they do, they end on $\Sigma$, as $\psi$ is not defined behind $\Sigma$. Put differently, no Bohmian trajectories begin on $\Sigma$, they all begin at $t = 0$ in $\Omega$ with $|\psi_0|^2$ distribution. In fact, the right-hand side of (19.4) is exactly the probability distribution of the space-time point at which the Bohmian trajectory reaches the boundary. That is not surprising, as in a Bohmian world we would expect the detector to click when and where the particle reaches the detecting surface. As a further consequence, the right-hand side of (19.5) is exactly the probability that the Bohmian trajectory never reaches $\Sigma$. In particular, (19.4) and (19.5) together define a probability distribution on $\mathscr{Z}$. Had we evolved $\psi_0$ with the Schrödinger equation on $\mathbb{R}^3$ without boundary condition on $\Sigma$, then some Bohmian trajectories may cross $\Sigma$ several times in both directions; this illustrates that the trajectory in the presence of detectors can be different from what it would have been in the absence of detectors.

Since probability can only be lost at the boundary, never gained,

$$\|\psi_t\|^2 = \int_\Omega d^2\boldsymbol{x}\, |\psi_t(\boldsymbol{x})|^2 \tag{19.7}$$

can only decrease with $t$, never increase. So here we are dealing with a new kind of Schrödinger equation whose time evolution is not unitary as the norm of $\psi$ is not conserved. The time evolution operators $W_t$, defined by the property $W_t\psi_0 = \psi_t$, have the following properties: First, they are not unitary but satisfy $\|W_t\psi\| \leq \|\psi\|$; such operators are called *contractions*. Second, $W_sW_t = W_{s+t}$ and $W_0 = I$; a family $(W_t)_{t\geq 0}$ with this property is called a *semigroup*. Thus, the $W_t$ form a *contraction semigroup*.

In fact, $\|\psi_t\|^2$ is the probability that the Bohmian particle is still somewhere in $\Omega$ at time $t$, that is, has not reached the boundary yet. In particular, as an alternative to (19.5) we can write

$$\mathbb{P}(Z = \infty) = \lim_{t\to\infty} \|\psi_t\|^2 \,. \tag{19.8}$$

The conclusions from our considerations about Bohmian trajectories can also be obtained from the Ostrogradskii–Gauss integral theorem (divergence theorem) in 4 dimensions: The 4-vector field $j = (\rho, \boldsymbol{j})$ has vanishing 4-divergence, as that is exactly what the continuity equation (2.16) expresses. Integrating the divergence over $[0,t] \times \Omega$

yields

$$0 = \int_0^t dt' \int_\Omega d^3\boldsymbol{x}\, \mathrm{div}\, j(t', \boldsymbol{x}) \tag{19.9}$$

$$= \int_\Omega d^3\boldsymbol{x}\, \rho(t, \boldsymbol{x}) - \int_\Omega d^3\boldsymbol{x}\, \rho(0, \boldsymbol{x}) + \int_0^t dt' \int_\Sigma d^2\boldsymbol{x}\, \boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{j}(t', \boldsymbol{x}) \tag{19.10}$$

$$= \|\psi_t\|^2 - 1 + \int_0^t dt' \int_\Sigma d^2\boldsymbol{x}\, \boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{j}(t', \boldsymbol{x})\,. \tag{19.11}$$

Since the last integrand is non-negative, $\|\psi_t\|^2$ is decreasing with time and equals $1-$ the flux of $\boldsymbol{j}$ into the boundary during $[0, t]$. In particular,

$$\lim_{t\to\infty} \|\psi_t\|^2 = 1 - \int\limits_0^\infty dt' \int\limits_\Sigma d^2\boldsymbol{x}\, \boldsymbol{n}(\boldsymbol{x}) \cdot \boldsymbol{j}(t', \boldsymbol{x})\,, \tag{19.12}$$

so (19.5) is non-negative, and (19.4) and (19.5) together define a probability distribution.

So what is the POVM $E$? It is given by

$$E\big(dt \times d^2\boldsymbol{x}\big) = \frac{\hbar\kappa}{m}\, W_t^\dagger\, |\boldsymbol{x}\rangle\langle\boldsymbol{x}|\, W_t\, dt\, d^2\boldsymbol{x} \tag{19.13}$$

$$E(\{\infty\}) = \lim_{t\to\infty} W_t^\dagger W_t\,. \tag{19.14}$$

Since the $E(dt)$ are not projections, there are in general no eigenstates of detection time.

Variants of the absorbing boundary rule have been developed for moving surfaces, systems of several detectable particles, and particles with spin.[42]

[42] R. Tumulka: Detection Time Distribution for Several Quantum Particles. `http://arxiv.org/abs/1601.03871`

# 20 Density Matrix and Mixed State

In this chapter we prove a limitation to knowledge in quantum mechanics that follows from the main theorem about POVMs. Let

$$\mathbb{S}(\mathscr{H}) = \{\psi \in \mathscr{H} : \|\psi\| = 1\} \tag{20.1}$$

denote the unit sphere in Hilbert space. Suppose that we have a mechanism that generates random wave functions $\Psi \in \mathbb{S}(\mathscr{H})$ with probability distribution $\mu$ on $\mathbb{S}(\mathscr{H})$. Then *it is impossible to determine $\mu$ empirically.* In fact, there exist different distributions $\mu_1 \neq \mu_2$ that are *empirically indistinguishable*, i.e., they lead to the same distribution of outcomes $Z$ for any experiment. We call such distributions *empirically equivalent* (which is an equivalence relation) and show that the equivalence classes are in one-to-one correspondence with certain operators known as *density matrices* or *density operators*.

To describe these matters, we need the mathematical concept of *trace*.

## 20.1 Trace

**Definition 20.1.** The trace of a matrix $A = (A_{mn})$ is the sum of its diagonal elements. The trace of an operator $T$ is defined to be the sum of the diagonal elements of its matrix representation $T_{nm} = \langle n|T|m\rangle$ relative to an arbitrary ONB $\{|n\rangle\}$,

$$\operatorname{tr} T = \sum_{n=1}^{\infty} \langle n|T|n\rangle \,. \tag{20.2}$$

Every positive operator either has finite trace or has trace $+\infty$, and the value of the trace does not depend on the choice of ONB. The *trace class* is the set of those operators $T$ for which the positive operator $\sqrt{T^\dagger T}$ has finite trace. For every operator from the trace class, the trace is finite and does not depend on the ONB.

The trace has the following properties for all operators $A, B, \ldots$ from the trace class:

(i) The trace is linear:

$$\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B\,, \quad \operatorname{tr}(\lambda A) = \lambda \operatorname{tr} A \tag{20.3}$$

for all $\lambda \in \mathbb{C}$.

(ii) The trace is invariant under cyclic permutation of factors:

$$\operatorname{tr}(AB\cdots YZ) = \operatorname{tr}(ZAB\cdots Y)\,. \tag{20.4}$$

In particular $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ and $\operatorname{tr}(ABC) = \operatorname{tr}(CAB)$, which is, however, not always the same as $\operatorname{tr}(CBA)$.

(iii) If an operator $T$ can be diagonalized, i.e., if there exists an orthonormal basis of eigenvectors, then $\operatorname{tr}(T)$ is the sum of the eigenvalues, counted with multiplicity (= degree of degeneracy).

(iv) The trace of the adjoint operator $T^\dagger$ is the complex-conjugate of the trace of $T$: $\mathrm{tr}(T^\dagger) = \mathrm{tr}(T)^*$.

(v) The trace of a self-adjoint operator $T$ is real.

(vi) If $T$ is a positive operator then $\mathrm{tr}(T) \geq 0$.

## 20.2 The Trace Formula in Quantum Mechanics

Suppose that (by whatever mechanism) we have generated a random wave function $\Psi \in \mathbb{S}(\mathscr{H})$ with probability distribution $\mu$ on $\mathbb{S}(\mathscr{H})$. Then for any experiment $\mathscr{E}$ with POVM $E$, the probability distribution of the outcome $Z$ is

$$\mathbb{P}(Z \in B) = \mathbb{E}\langle\Psi|E(B)|\Psi\rangle = \int\limits_{\mathbb{S}(\mathscr{H})} \mu(d\psi)\, \langle\psi|E(B)|\psi\rangle = \mathrm{tr}(\rho_\mu\, E(B))\,, \tag{20.5}$$

where $\mathbb{E}$ means expectation, and

$$\rho_\mu = \mathbb{E}|\Psi\rangle\langle\Psi| = \int\limits_{\mathbb{S}(\mathscr{H})} \mu(d\psi)\, |\psi\rangle\langle\psi| \tag{20.6}$$

is called the *density operator* or *density matrix* (rarely: *statistical operator*) of the distribution $\mu$. Eq. (20.5) is called the *trace formula*. It was discovered by John von Neumann in 1927,[43] except that von Neumann did not know POVMs and considered only PVMs. In case the distribution $\mu$ is concentrated on discrete points on $\mathbb{S}(\mathscr{H})$, (20.6) becomes

$$\rho_\mu = \mathbb{E}|\Psi\rangle\langle\Psi| = \sum_\psi \mu(\psi)\, |\psi\rangle\langle\psi|\,. \tag{20.7}$$

In order to verify (20.5), note first that

$$\mathrm{tr}\Big(|\psi\rangle\langle\psi|\, E\Big) = \langle\psi|E|\psi\rangle \tag{20.8}$$

because, if we choose the basis $\{|n\rangle\}$ in (20.2) such that $|1\rangle = \psi$, then the summands in (20.2) are $\langle n|\psi\rangle\langle\psi|E|n\rangle$, which for $n = 1$ is $\langle\psi|E|\psi\rangle$ and for $n > 1$ is zero because $\langle n|1\rangle = 0$. By linearity, we also have that

$$\mathrm{tr}\Big(\sum_j \mu(\psi_j)|\psi_j\rangle\langle\psi_j|\, E\Big) = \sum_j \mu(\psi_j)\, \langle\psi_j|E|\psi_j\rangle\,, \tag{20.9}$$

which yields (20.5) for any $\mu$ that is concentrated on finitely many points $\psi_j$ on $\mathbb{S}(\mathscr{H})$. One can prove (20.5) for arbitrary probability distribution $\mu$ by considering limits.

[43] J. von Neumann: Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik. *Göttinger Nachrichten* **1(10)**: 245–272 (1927). Reprinted in *John von Neumann: Collected Works Vol. I*, A.H. Taub (editor), Oxford: Pergamon Press (1961)

Now let us draw conclusions from the formula (20.5). It implies that the distribution of the outcome $Z$ depends on $\mu$ only through $\rho_\mu$. Different distributions $\mu_a, \mu_b$ can have the same $\rho = \rho_{\mu_a} = \rho_{\mu_b}$; for example, if $\mathscr{H} = \mathbb{C}^2$ then the uniform distribution over $\mathbb{S}(\mathscr{H})$ has $\rho = \frac{1}{2}I$, and for every orthonormal basis $|\phi_1\rangle, |\phi_2\rangle$ of $\mathbb{C}^2$ the probability distribution

$$\tfrac{1}{2}\delta_{\phi_1} + \tfrac{1}{2}\delta_{\phi_2} \tag{20.10}$$

also has $\rho = \frac{1}{2}I$. Such two distributions $\mu_a, \mu_b$ will lead to the same distribution of outcomes for any experiment, and are therefore *empirically indistinguishable.*

## 20.3 Limitations to Knowledge

We can turn this result into an argument showing that there must be facts we cannot find out by experiment: Suppose I choose between two options, I choose $\mu$ to be either $\mu_a$ or $\mu_b$. Suppose that each $\mu$ is of the form (20.10), $\mu_a$ for the eigenbasis of $\sigma_3$ and $\mu_b$ for that of $\sigma_1$. Then I choose $n = 10,000$ points $\psi_i$ on $\mathbb{S}(\mathscr{H})$ at random independently with $\mu$, then I prepare $n$ systems with wave functions $\psi_i$, and then I hand these systems to you with the challenge to determine whether $\mu = \mu_a$ or $\mu = \mu_b$. As a consequence of (20.5), you cannot determine that by means of experiments on the $n$ systems. On the other hand, nature knows the right answer, as I will argue now. I have kept records of each $\psi_i$, so I can make a list of the $m \approx 5,000$ systems that I prepared in $\phi_1$. I tell you that I did choose $\mu_b$, I give you the list and predict that for all $m$ systems on the list, a quantum measurement of $\sigma_1$ will yield $+1$, while for all others it will yield $-1$. By the laws of quantum mechanics, you will find my prediction confirmed. But had I prepared half of all systems in $|z\text{-up}\rangle$ and the other half in $|z\text{-down}\rangle$, then all outcomes of $\sigma_1$-measurements would have had to be random with equal probability for $+1$ and $-1$, so my predictions would have been wrong in about half of the cases. Thus, nature must remember at least whether it was a mixture of $\sigma_1$-eigenvectors or of $\sigma_3$-eigenvectors. (In fact, nature must remember much more, viz., which systems exactly must yield $+1$ upon measurement of $\sigma_1$.) There is a fact in nature (viz., whether $\mu = \mu_a$ or $\mu = \mu_b$) that we cannot discover empirically. Nature can keep a secret. Limitations to knowledge are a fact of quantum mechanics, regardless of which interpretation we prefer.

## 20.4 Density Matrix and Dynamics

If the random vector $\Psi$ evolves according to the Schrödinger equation, $\Psi_t = e^{-iHt/\hbar}\Psi$, the distribution changes into $\mu_t$ and the density matrix into

$$\rho_t = e^{-iHt/\hbar}\rho e^{iHt/\hbar}. \tag{20.11}$$

In analogy to the Schrödinger equation, this can be written as a differential equation,

$$\frac{d\rho_t}{dt} = -\tfrac{i}{\hbar}[H, \rho_t], \tag{20.12}$$

known as the *von Neumann equation.* The step from (20.11) to (20.12) is based on the fact that

$$\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A\,. \tag{20.13}$$

A density matrix is also often called a *quantum state.* If $\rho = |\psi\rangle\langle\psi|$ with $\|\psi\| = 1$, then $\rho$ is usually called a *pure quantum state*, otherwise a *mixed quantum state.* A probability distribution $\mu$ has $\rho_\mu = |\psi\rangle\langle\psi|$ if and only if $\mu$ is concentrated on $\mathbb{C}\psi$, i.e., $\Psi = e^{i\Theta}\psi$ with a random global phase factor.

As we have seen, a density matrix $\rho$ is always a positive operator with $\operatorname{tr}\rho = 1$. Conversely, every positive operator $\rho$ with $\operatorname{tr}\rho = 1$ is a density matrix, i.e., $\rho = \rho_\mu$ for some probability distribution $\mu$ on $\mathbb{S}(\mathscr{H})$. Here is one such $\mu$: find an orthonormal basis $\{|\phi_n\rangle : n \in \mathbb{N}\}$ of eigenvectors of $\rho$ with eigenvalues $p_n \in [0,\infty)$. Then

$$\sum_n p_n = \operatorname{tr}\rho = 1\,. \tag{20.14}$$

Now let $\mu$ be the distribution that gives probability $p_n$ to $\phi_n$; its density matrix is just the $\rho$ we started with.

# 21 Reduced Density Matrix and Partial Trace

There is another way in which density matrices arise, leading to what is called the *reduced density matrix*. Suppose that the system under consideration consists of two parts, system $a$ and system $b$, so that its Hilbert space is $\mathscr{H} = \mathscr{H}_a \otimes \mathscr{H}_b$.

**Theorem 21.1.** *In Bohmian mechanics, an experiment in which the apparatus interacts only with system $a$ but not with system $b$ has a POVM of the form*

$$E(B) = E_a(B) \otimes I_b \,, \tag{21.1}$$

*where $I_b$ is the identity on $\mathscr{H}_b$.*

Proof: homework exercise. A similar theorem holds in GRW theory.

In the case (21.1), the distribution of the outcome is

$$\mathbb{P}(Z \in B) = \langle\psi|E(B)|\psi\rangle = \mathrm{tr}\big(\rho_\psi\, E_a(B)\big) \tag{21.2}$$

with the *reduced density matrix of system $a$*

$$\rho_\psi = \mathrm{tr}_b|\psi\rangle\langle\psi| \,, \tag{21.3}$$

where $\mathrm{tr}_b$ means the *partial trace* over $\mathscr{H}_b$. The reduced density matrix and the trace formula for it were discovered by Lev Landau in 1927.[44]

## 21.1 Partial Trace

This means the following. Let $\{\phi_n^a\}$ be an orthonormal basis of $\mathscr{H}_a$ and $\{\phi_n^b\}$ an orthonormal basis of $\mathscr{H}_b$. Then $\{\phi_n^a \otimes \phi_m^b\}$ is an orthonormal basis of $\mathscr{H} = \mathscr{H}_a \otimes \mathscr{H}_b$. If $T$ is an operator on $\mathscr{H}$ then the operator $S = \mathrm{tr}_b\, T$ on $\mathscr{H}_a$ is characterized by its matrix elements

$$\langle\phi_n^a|S|\phi_k^a\rangle = \sum_{m=1}^{\infty} \langle\phi_n^a \otimes \phi_m^b|T|\phi_k^a \otimes \phi_m^b\rangle \,, \tag{21.4}$$

where the inner products on the right hand side are inner products in $\mathscr{H}_a \otimes \mathscr{H}_b$. We will sometimes write

$$S = \sum_{m=1}^{\infty} \langle\phi_m^b|T|\phi_m^b\rangle \,, \tag{21.5}$$

where the inner products are partial inner products.

The partial trace has the following properties:

(i) It is linear:

$$\mathrm{tr}_b(S+T) = \mathrm{tr}_b(S) + \mathrm{tr}_b(T) \,, \quad \mathrm{tr}_b(\lambda T) = \lambda\, \mathrm{tr}_b(T) \tag{21.6}$$

[44] L. Landau: Das Dämpfungsproblem in der Wellenmechanik. *Zeitschrift für Physik* **45**: 430–441 (1927)

(ii) $\operatorname{tr}(\operatorname{tr}_b(T)) = \operatorname{tr}(T)$. Here, the first tr symbol means the trace in $\mathscr{H}_a$, the second one the partial trace, and the last one the trace in $\mathscr{H}_a \otimes \mathscr{H}_b$. This property follows from (21.4) by setting $k = n$ and summing over $n$.

(iii) $\operatorname{tr}_b(T^\dagger) = (\operatorname{tr}_b T)^\dagger$. The adjoint of the partial trace is the partial trace of the adjoint. In particular, if $T$ is self-adjoint then so is $\operatorname{tr}_b T$.

(iv) $\operatorname{tr}_b(T_a \otimes T_b) = (\operatorname{tr} T_b) T_a$.

(v) If $T$ is a positive operator then so is $\operatorname{tr}_b T$.

(vi) $\operatorname{tr}_b\bigl[S(T_a \otimes I_b)\bigr] = (\operatorname{tr}_b S) T_a$.

(vii) $\operatorname{tr}_b\bigl[S(I_a \otimes T_b)\bigr] = \operatorname{tr}_b\bigl[(I_a \otimes T_b) S\bigr]$.

## 21.2 The Trace Formula (21.2)

From properties (vi) and (ii) we obtain that

$$\operatorname{tr}\bigl[S(T_a \otimes I_b)\bigr] = \operatorname{tr}\bigl[(\operatorname{tr}_b S) T_a\bigr] . \tag{21.7}$$

Setting $S = |\psi\rangle\langle\psi|$ and $T_a = E_a(B)$, we find that $\operatorname{tr}_b S = \rho_\psi$ and

$$\langle\psi|E_a(B) \otimes I_b|\psi\rangle = \operatorname{tr}\bigl[|\psi\rangle\langle\psi|(E_a(B) \otimes I_b)\bigr] = \operatorname{tr}\bigl[\rho_\psi E_a(B)\bigr] , \tag{21.8}$$

which proves (21.2).

From properties (ii) and (v) it follows also that $\rho_\psi$ is a positive operator with trace 1. Conversely, every positive operator $\rho$ on $\mathscr{H}_a$ with $\operatorname{tr}\rho = 1$ arises as a reduced density matrix. Indeed, if $\rho = \sum_n p_n |\phi_n\rangle\langle\phi_n|$ with $p_n \geq 0$, $\sum_n p_n = 1$ and orthonormal $\phi_n$, then choose any ONB $\{\chi_m\}$ of $\mathscr{H}_b$ and set $\psi = \sum_n \sqrt{p_n} \phi_n \otimes \chi_n$. Then $\psi \in \mathscr{H}_a \otimes \mathscr{H}_b$, $\|\psi\| = 1$, and $\operatorname{tr}_b |\psi\rangle\langle\psi| = \rho$.

## 21.3 Statistical Reduced Density Matrix

Statistical density matrices as in (20.6) and reduced density matrices can be combined: If $\Psi \in \mathscr{H}_a \otimes \mathscr{H}_b$ is random then set

$$\rho = \mathbb{E} \operatorname{tr}_b |\Psi\rangle\langle\Psi| = \operatorname{tr}_b \mathbb{E}\, |\Psi\rangle\langle\Psi| . \tag{21.9}$$

## 21.4 The Measurement Problem Again

Statistical and reduced density matrices sometimes get confused; here is an example. Consider again the wave function of the measurement problem,

$$\Psi = \sum_\alpha \Psi_\alpha , \tag{21.10}$$

the wave function of an object and an apparatus after a quantum measurement of the observable $A = \sum \alpha P_\alpha$. Suppose that $\Psi_\alpha$, the contribution corresponding to the outcome $\alpha$, is of the form

$$\Psi_\alpha = c_\alpha \, \psi_\alpha \otimes \phi_\alpha \,, \tag{21.11}$$

where $c_\alpha = \|P_\alpha \psi\|$, $\psi$ is the initial object wave function $\psi$, $\psi_\alpha = P_\alpha\psi/\|P_\alpha\psi\|$, and $\phi_\alpha$ with $\|\phi_\alpha\| = 1$ is a wave function of the apparatus after having measured $\alpha$. Since the $\phi_\alpha$ have disjoint supports in configuration space, they are mutually orthogonal; thus, they are a subset of some orthonormal basis $\{\phi_n\}$. The reduced density matrix of the object is

$$\rho_\Psi = \mathrm{tr}_b \, |\Psi\rangle\langle\Psi| = \sum_n \langle\phi_n|\Psi\rangle\langle\Psi|\phi_n\rangle = \sum_\alpha |c_\alpha|^2 \, |\psi_\alpha\rangle\langle\psi_\alpha| \,. \tag{21.12}$$

This is the same density matrix as the statistical density matrix associated with the probability distribution $\mu$ of the collapsed wave function $\psi'$,

$$\mu = \sum_\alpha |c_\alpha|^2 \, \delta_{\psi_\alpha} \,, \tag{21.13}$$

since

$$\rho_\mu = \sum_\alpha |c_\alpha|^2 |\psi_\alpha\rangle\langle\psi_\alpha| \,. \tag{21.14}$$

It is sometimes claimed that this fact solves the measurement problem. The argument is this: From (21.10) we obtain (21.12), which is the same as (21.14), which means that the system's wave function has distribution (21.13), so we have a random outcome $\alpha$. This argument is incorrect, as the mere fact that two situations—one with $\Psi$ as in (21.10), the other with random $\psi'$—define the same density matrix for the object does not mean the two situations are physically equivalent. And obviously from (21.10), the situation after a quantum measurement involves neither a random outcome nor a random wave function. As John Bell once put it, "and is not or."

It is sometimes taken as the definition of *decoherence* that the reduced density matrix is (approximately) diagonal in the eigenbasis of the relevant operator $A$. In a previous lecture I had defined decoherence as the situation that two or more wave packets $\Psi_\alpha$ are macroscopically disjoint in configuration space (and thus remain disjoint for the relevant future). The connection between the two definitions is that the latter implies the former if $\Psi_\alpha$ is of the form (21.11).

It is common to call a density matrix that is a 1-dimensional projection a *pure state* and otherwise a *mixed state*, even if it is a reduced density matrix and thus does not arise from a mixture (i.e., from a probability distribution $\mu$). A reduced density matrix $\rho_\psi$ is pure if and only if $\psi$ is a tensor product, i.e., there are $\chi_a \in \mathscr{H}_a$ and $\chi_b \in \mathscr{H}_b$ such that $\psi = \chi_a \otimes \chi_b$.

## 21.5 The No-Signaling Theorem

The no-signaling theorem is a consequence of the quantum formalism: If system $a$ is located in Alice's lab and system $b$ in Bob's, and if the two labs do not interact, then the statistical reduced density matrix of system $a$ is (i) not affected by any measurement Bob performs, and (ii) does not depend on the Hamiltonian of system $b$.

To verify (i), suppose that systems $a$ and $b$ together have wave function $\psi \in \mathscr{H}_a \otimes \mathscr{H}_b$, and that Bob measures the observable $B$, which is a self-adjoint operator on $\mathscr{H}_b$. Let $\beta$ denote the eigenvalues of $B$ and $P_\beta$ the projection to the eigenspace of eigenvalue $\beta$. The probability that Bob obtains the outcome $\beta$ is

$$\mathbb{P}(Z = \beta) = \langle \psi | I_a \otimes P_\beta | \psi \rangle \,. \tag{21.15}$$

If Bob obtains $\beta$ then $\psi$ collapses to $\psi'/Z$, where $\psi' = (I_a \otimes P_\beta)\psi$ and the normalization factor is given by $Z = \|\psi'\| = \langle \psi | I_a \otimes P_\beta | \psi \rangle^{1/2}$. Thus, the statistical reduced density matrix of system $a$ is

$$\begin{aligned}
\rho' &= \mathrm{tr}_b \sum_\beta \mathbb{P}(Z=\beta) \frac{|\psi'\rangle\langle\psi'|}{Z^2} && (21.16)\\
&= \sum_\beta \mathrm{tr}_b \Big[ (I_a \otimes P_\beta) |\psi\rangle\langle\psi| (I_a \otimes P_\beta) \Big] && (21.17)\\
&\overset{\text{(vii)}}{=} \sum_\beta \mathrm{tr}_b \Big[ |\psi\rangle\langle\psi| (I_a \otimes P_\beta) \Big] && (21.18)\\
&= \mathrm{tr}_b \Big[ |\psi\rangle\langle\psi| (I_a \otimes \sum_\beta P_\beta) \Big] && (21.19)\\
&= \mathrm{tr}_b |\psi\rangle\langle\psi| = \rho_\psi \,. && (21.20)
\end{aligned}$$

To verify (ii), note that in the absence of interaction the unitary time evolution operator is $U_t = U_{a,t} \otimes U_{b,t}$. Thus, the reduced density matrix evolves according to

$$\begin{aligned}
\rho_t &= \mathrm{tr}_b |U_t\psi\rangle\langle U_t\psi| && (21.21)\\
&= \mathrm{tr}_b \Big[ U_t |\psi\rangle\langle\psi| U_t^\dagger \Big] && (21.22)\\
&= \mathrm{tr}_b \Big[ (U_{a,t} \otimes U_{b,t}) |\psi\rangle\langle\psi| (U_{a,t}^\dagger \otimes U_{b,t}^\dagger) \Big] && (21.23)\\
&= \mathrm{tr}_b \Big[ (U_{a,t} \otimes I_b) |\psi\rangle\langle\psi| (U_{a,t}^\dagger \otimes (U_{b,t}^\dagger U_{b,t})) \Big] && (21.24)\\
&= \mathrm{tr}_b \Big[ (U_{a,t} \otimes I_b) |\psi\rangle\langle\psi| (U_{a,t}^\dagger \otimes I_b) \Big] && (21.25)\\
&= U_{a,t} [\mathrm{tr}_b \, |\psi\rangle\langle\psi|] U_{a,t}^\dagger = U_{a,t} \rho_\psi U_{a,t}^\dagger \,, && (21.26)
\end{aligned}$$

which does not depend on $U_{b,t}$. The argument extends without difficulty to statistical reduced density matrices.

## 21.6 Canonical Typicality

This is an application of reduced density matrices in quantum statistical mechanics. The main goal of quantum statistical mechanics is to derive facts of thermodynamics from a quantum mechanical analysis of systems with a macroscopic number of particles (say, $N > 10^{20}$). One of the rules of quantum statistical mechanics asserts that if a quantum system $S$ is in thermal equilibrium at absolute temperature $T \geq 0$, then it has density matrix

$$\rho_{\mathrm{can}} = \frac{1}{Z} e^{-\beta H_S}\,, \tag{21.27}$$

where $H_s$ is the system's Hamiltonian, $\beta = 1/kT$ with $k = 1.38 \cdot 10^{-23}$ J/K the Boltzmann constant, and $Z = \mathrm{tr} e^{-\beta H}$ the normalizing factor; $\rho_{\mathrm{can}}$ is called the *canonical density matrix* with inverse temperature $\beta$.

While this rule has long been used, its justification is rather recent (2006) and goes as follows. Suppose that $S$ is coupled to another system $B$ (the "heat bath"), and suppose that $S$ and $B$ together have wave function $\psi \in \mathscr{H}_S \otimes \mathscr{H}_B$ and Hamiltonian $H$ with pure point spectrum (this comes out for systems confined to finite volume). Let $I_{\mathrm{mc}} = [E, E + \Delta E]$ be an energy interval whose length $\Delta E$ is small on the macroscopic scale but large enough for $I_{\mathrm{mc}}$ to contain very many eigenvalues of $H$; $I_{\mathrm{mc}}$ is called a *micro-canonical energy shell*. Let $\mathscr{H}_{\mathrm{mc}}$ be the corresponding spectral subspace, i.e., the range of $1_{I_{\mathrm{mc}}}(H)$, and $u_{\mathrm{mc}}$ the uniform probability distribution over $\mathbb{S}(\mathscr{H}_{\mathrm{mc}})$.

**Theorem 21.2.** *(canonical typicality, informal statement) If $B$ is sufficiently "large," and if the interaction between $S$ and $B$ is negligible,*

$$H \approx H_S \otimes I_B + I_S \otimes H_B\,, \tag{21.28}$$

*then for most $\psi$ relative to $u_{\mathrm{mc}}$, the reduced density matrix of $S$ is approximately canonical for some value of $\beta$, i.e.,*

$$\mathrm{tr}_B |\psi\rangle\langle\psi| \approx \rho_{\mathrm{can}}\,. \tag{21.29}$$

In order to arrive at a typical $\psi \in \mathbb{S}(\mathscr{H}_{\mathrm{mc}})$ (and thus at thermal equilibrium between $S$ and $B$), it will be relevant to have some interaction between $S$ and $B$. Large interaction terms in $H$, however, will lead to deviations from the form (21.27). It is relevant for (21.29) that $S$ and $B$ are entangled: If they were not, then the reduced density matrix of $S$ would be pure, whereas $\rho_{\mathrm{can}}$ is usually highly mixed (i.e., has many eigenvalues that are significantly nonzero).

Canonical typicality explains why we see canonical density matrices: Because "most" wave functions of $S \cup B$ lead to a canonical density matrix for $S$.

# 22 Quantum Logic

The expression "quantum logic" is used in the literature for (at least) three different things:

- a certain piece of mathematics that is rather pretty;
- a certain analogy between two formalisms that is rather limited;
- a certain philosophical idea that is rather silly.

*Logic* is the collection of those statements and rules that are valid in every conceivable universe and every conceivable situation. Some people have suggested that logic simply consists of the rules for the connectives "and", "or," and "not", with "$\forall x \in M$" an extension of "and" and "$\exists x \in M$" an extension of "or" to (possibly infinite) ranges $M$. I would say that viewpoint is not completely right (because of Gödel's theorem[45]) and not completely wrong. Be that as it may, let us focus for a moment on the operations "and" (conjunction $A \wedge B$), "or" (disjunction $A \vee B$), and "not" (negation $\neg A$), and let us ignore infinite conjunctions or disjunctions.

A *Boolean algebra* is a set $\mathscr{A}$ of elements $A, B, C, \ldots$ of which we can form $A \wedge B$, $A \vee B$, and $\neg A$, such that the following rules hold:

- $\wedge$ and $\vee$ are associative, commutative, and idempotent ($A \wedge A = A$ and $A \vee A = A$).
- Absorption laws: $A \wedge (A \vee B) = A$ and $A \vee (A \wedge B) = A$.
- There are elements $0 \in \mathscr{A}$ ("false") and $1 \in \mathscr{A}$ ("true") such that for all $A \in \mathscr{A}$, $A \wedge 0 = 0$, $A \wedge 1 = A$, $A \vee 0 = A$, $A \vee 1 = 1$.
- Complementation laws: $A \wedge \neg A = 0$, $A \vee \neg A = 1$.
- Distributive laws: $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$ and $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$.

It follows from these axioms that $\neg(\neg A) = A$, and that de Morgan's laws hold, $\neg A \vee \neg B = \neg(A \wedge B)$ and $\neg A \wedge \neg B = \neg(A \vee B)$.

The laws of logic for "and," "or," and "not" are exactly the laws that hold in every Boolean algebra, with $A, B, C, \ldots$ playing the role of statements or propositions or conditions. Another case in which these axioms are satisfied is that $A, B, C, \ldots$ are *sets*, more precisely subsets of some set $\Omega$, $A \wedge B$ means the intersection $A \cap B$, $A \vee B$ means the union $A \cup B$, $\neg A$ means the complement $A^c = \Omega \setminus A$, 0 means the empty set $\emptyset$, and 1 means the full set $\Omega$. That is, every family $\mathscr{A}$ of subsets of $\Omega$ that contains $\Omega$ and is closed under complement and intersection (in particular, every $\sigma$-algebra) is a Boolean algebra. (It turns out that also, conversely, every Boolean algebra can be realized as a family of subsets of some set $\Omega$.)

[45] Gödel provides an exampe of a statement that is true about the natural numbers, so it follows from the Peano axioms, but cannot be derived from the Peano axioms using the standard rules of logic, thus showing that these rules are incomplete.

Now let $A, B, C, \ldots$ be *subspaces* of a Hilbert space $\mathscr{H}$ (more precisely, closed subspaces, which makes no difference in finite dimension where every subspace is closed); let $A \wedge B := A \cap B$, $A \vee B := \overline{\text{span}}(A \cup B)$ (the smallest closed subspace containing both $A$ and $B$), and $\neg A := A^{\perp} = \{\psi \in \mathscr{H} : \langle \psi | \phi \rangle = 0 \,\forall \phi \in A\}$ be the orthogonal complement of $A$; let $0 = \{0\}$ be the 0-dimensional subspace and $1 = \mathscr{H}$ the full subspace. Then all axioms except distributivity are satisfied. So this structure is no longer a Boolean algebra; it is called an *orthomodular lattice* or simply *lattice*. Hence, a distributive lattice is a Boolean algebra, and the closed subspaces form a non-distributive lattice $\mathbb{L}(\mathscr{H})$.

That is nice mathematics, and we will see more of that in a moment. The analogy I mentioned holds between $\mathbb{L}(\mathscr{H})$ and Boolean algebras, often understood as representing the rules of logic. The analogy is that both are lattices. In order to emphasize the analogy, some authors call the elements of $\mathbb{L}(\mathscr{H})$ "propositions" and the operations $\wedge, \vee$, and $\neg$ "and," "or," and "not." They call $\mathbb{L}(\mathscr{H})$ the "quantum logic" and say things like, $A \in \mathbb{L}(\mathscr{H})$ is a yes-no question that you can ask about a quantum system, as you can carry out a quantum measurement of the projection to $A$ and get result 0 (no) or 1 (yes).

Here is why the analogy is rather limited. Let me give two examples.

- First, consider a spin-$\frac{1}{2}$ particle with spinor $\psi \in \mathbb{C}^2$, and consider the words "$\psi$ *lies in* $\mathbb{C}|\text{up}\rangle$." These words sound very much like a proposition, let me call it $\mathscr{P}$, and indeed they naturally correspond to a subspace of $\mathscr{H} = \mathbb{C}^2$, viz., $\mathbb{C}|\text{up}\rangle$. Now the negation of $\mathscr{P}$ is, of course, "$\psi$ *lies in* $\mathscr{H} \setminus \mathbb{C}|\text{up}\rangle$," whereas the orthogonal complement of $\mathbb{C}|\text{up}\rangle$ is $\mathbb{C}|\text{down}\rangle$. Let me say that again in different words: The negation of "spin is up" is not "spin is down," but "spin is in any direction but up."

- Second, consider the delayed-choice experiment in the form discussed at the end of Section 18.4: forget about the interference region and consider just the two options of either putting detectors in the two slits or putting detectors far away. The first option has the PVM $P_{\text{left}} + P_{\text{right}} = I$, the second to the PVM $U^\dagger P_{\text{far right}} U + U^\dagger P_{\text{far left}} U = I$, where $U$ is the unitary time evolution from the slits to the far regions where the detectors are placed. The two PVMs are identical, as $U^\dagger P_{\text{far right}} U = P_{\text{left}}$ (and likewise for the other projection); that is, we have two experiments associated with the same observable. If we think of subspaces as propositions, then it is natural to think of *the particle passes through the left slit* as a proposition and identify it with the subspace $A$ that is the range of $P_{\text{left}}$. But if we carry out the second option, detect the particle on the far right, and say that we have confirmed the proposition $A$ and thus that the particle passed through the left slit, then we have committed Wheeler's fallacy.

The philosophical idea that I mentioned is that logic as we know it is false, that it applies in classical physics but not in quantum physics, and that a different kind of logic with different rules applies in quantum physics—a *quantum logic*. Why did I call that a rather silly idea? Because logic is, by definition, what is true in every conceivable situation. So logic cannot depend on physical laws and cannot be revised by empirical science. As Tim Maudlin once nicely said:

> There is no point in arguing with somebody who does not believe in logic.

Bell wrote in *Against "measurement"* (1989, page 216 in the 2nd edition of *Speakable and unspeakable in quantum mechanics*):

> When one forgets the role of the apparatus, as the word "measurement" makes all too likely, one despairs of ordinary logic—hence "quantum logic." When one remembers the role of the apparatus, ordinary logic is just fine.

Nevertheless, there is more mathematics relevant to $\mathbb{L}(\mathscr{H})$, something analogous to probability theory. Recall that a probability distribution on a set $\Omega$ is a normalized measure, that is, a mapping $\mu$ from subsets of $\Omega$ to $[0,1]$ that is $\sigma$-additive and satisfies $\mu(1) = \mu(\Omega) = 1$. The domain of definition of $\mu$ is a $\sigma$-algebra, which is a Boolean algebra with slightly stronger requirements. By analogy, we define that a *normalized quantum measure* is a mapping $\hat{\mu} : \mathbb{L}(\mathscr{H}) \to [0,1]$ that satisfies $\hat{\mu}(1) = \hat{\mu}(\mathscr{H}) = 1$ and is $\sigma$-additive, i.e.,

$$\hat{\mu}\Big(\bigvee_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \hat{\mu}(A_n) \tag{22.1}$$

whenever $A_n \perp A_m$ for all $n \neq m$. (The relation $A \perp B$ can be expressed through lattice operations as $A \leq (\neg B)$, with $A \leq C$ defined to mean $A \vee C = C$ or, equivalently, $A \wedge C = A$. In $\mathbb{L}(\mathscr{H})$, $A \leq B \Leftrightarrow A \subseteq B$.)

**Theorem 22.1.** *(Gleason's theorem[46]) Suppose the dimension of $\mathscr{H}$ is at least 3 and at most countably infinite. Then the normalized quantum measures are exactly the mappings $\hat{\mu}$ of the form*

$$\hat{\mu}(A) = \mathrm{tr}(\rho P_A) \quad \forall A \in \mathbb{L}(\mathscr{H}) \,, \tag{22.2}$$

*where $P_A$ denotes the projection to $A$ and $\rho$ is a density matrix (i.e., a positive operator with trace 1).*

This amazing parallel between probability measures and density matrices has led some authors to call elements of $\mathbb{L}(\mathscr{H})$ "events" (as one would call subsets of $\Omega$). Again, this is a rather limited analogy, for the same reasons as above.

[46] A.M. Gleason: Measures on the closed subspaces of a Hilbert space. *Indiana University Mathematics Journal* **6**: 885–893 (1957)

# 23 No-Hidden-Variables Theorems

This name refers to a collection of theorems that aim at proving the impossibility of hidden variables. This aim may seem strange in view of the fact that Bohmian mechanics is a hidden-variable theory, is consistent and makes predictions in agreement with quantum mechanics. So how could hidden variables be impossible? A first observation concerns what is meant by "hidden variables." Most no-hidden-variable theorems (NHVTs) address the idea that every observable $A$ (a self-adjoint operator) has a true value $v_A$ in nature (the "hidden variable"), and that a quantum measurement of $A$ yields $v_A$ as its outcome. This idea should sound dubious to you because we have discussed already that observables are really equivalence classes of experiments, not all of which yield the same value. Moreover, we know that in Bohmian mechanics, a true value is associated with position but not with every observable, in particular not with spin observables. Hence, in this sense of "hidden variables," Bohmian mechanics is really a *no-hidden-variables theory.*

But this is not the central reason why the NHVTs do not exclude Bohmian mechanics. Suppose we choose, in Bohmian mechanics, one experiment from every equivalence class. (The experiment could be specified by specifying the wave function and configuration of the apparatus together with the joint Hamiltonian of object and apparatus as well as the calibration function.) For example, for every spin observable $\boldsymbol{n}\cdot\boldsymbol{\sigma}$ we could say we will measure it by a Stern-Gerlach experiment in the direction $\boldsymbol{n}$ and subsequent detection of the object particle. Then the outcome $Z_{\boldsymbol{n}}$ of the experiment is a function of the object wave function $\psi$ and the objection configuration $Q$, so we have associated with every observable $\boldsymbol{n}\cdot\boldsymbol{\sigma}$ a "true value" which comes out if we choose to carry out the experiment associated with $\boldsymbol{n}\cdot\boldsymbol{\sigma}$. And it is this situation that NHVTs claim to exclude! So we are back at an apparent conflict between Bohmian mechanics and NHVTs.

It may occur to you that even a much simpler example than Bohmian mechanics will prove the possibility of hidden-variable theories. Suppose we choose, as a trivial model, for every self-adjoint operator $A$ a random value $v_A$ independently of all other $v_{A'}$ with the Born distribution,

$$\mathbb{P}(v_A=\alpha)=\|P_\alpha\psi\|^2\,. \tag{23.1}$$

Then we have not provided a real theory of quantum mechanics as Bohmian mechanics provides, but we have provided a clearly consistent possibility for which values the variables $v_A$ could have that agrees with the probabilities seen in experiment. Therefore, all NHVTs must make some further assumptions about the hidden variables $v_A$ that are violated in the trivial model as well as in Bohmian mechanics. We now take a look at several NHVTs and their assumptions.

## 23.1 Bell's NHVT

Bell's theorem implies a NHVT, or rather, the second half of Bell's 1964 proof *is* a NHVT. Let me explain. In the trivial model, we have not specified how the $v_A$ change with time. They may change according to some law under the unitary time evolution;

more importantly for us now, they may change whenever $\psi$ collapses. That is, when a quantum measurement of $A$ is carried out, we should expect the $v_{A'}$ ($A' \neq A$) to change. However, there is an exception if we believe in locality. Then we should expect that Alice's measurement of $\boldsymbol{\alpha} \cdot \boldsymbol{\sigma}_a$ (on her particle $a$) will not alter the value of any spin observable $\boldsymbol{\beta} \cdot \boldsymbol{\sigma}_b$ acting on Bob's particle. But Bell's analysis shows that this is impossible. To sum up:

**Theorem 23.1.** *(Bell's NHVT, 1964) Consider a joint distribution of random variables $v_A$, where $A$ runs through the collection of observables*

$$\mathscr{A} \cup \mathscr{B} = \left\{ \boldsymbol{\alpha} \cdot \boldsymbol{\sigma}_a : \boldsymbol{\alpha} \in \mathbb{S}(\mathbb{R}^3) \right\} \cup \left\{ \boldsymbol{\beta} \cdot \boldsymbol{\sigma}_b : \boldsymbol{\beta} \in \mathbb{S}(\mathbb{R}^3) \right\}. \tag{23.2}$$

*Suppose that a quantum measurement of $A \in \mathscr{A}$ yields $v_A$ and does not alter the value of $v_B$ for any $B \in \mathscr{B}$, and that a subsequent quantum measurement of $B \in \mathscr{B}$ yields $v_B$. Then the joint distribution of the outcomes satisfies Bell's inequality* (16.32). *In particular, it disagrees with the distribution predicted by quantum mechanics.*

In short, local hidden variables are impossible.

## 23.2 Kochen and Specker's NHVT

This theorem addresses the idea that while in general a quantum measurement of $A$ may change the values of $v_{A'}$ for $A' \neq A$, this should not happen if $A$ and $A'$ can be "simultaneously measured."

**Theorem 23.2.** *(Kochen and Specker's NHVT[47], 1967) Suppose $3 \leq \dim \mathscr{H} < \infty$ and $\psi \in \mathbb{S}(\mathscr{H})$, let $\mathscr{A}$ be the set of all self-adjoint operators on $\mathscr{H}$, and consider a joint distribution of random variables $v_A$ for all $A \in \mathscr{A}$. Suppose that whenever $A, B \in \mathscr{A}$ commute, then a quantum measurement of $A$ yields $v_A$ and does not alter the value of $v_B$, and that a subsequent quantum measurement of $B$ yields $v_B$. Then the joint distribution of the outcomes disagrees, at least for some $A$ and $B$, with the distribution predicted by quantum mechanics using $\psi$.*

This result is actually weaker than Bell's (although it was proved later), as it makes a stronger assumption: indeed, any $\boldsymbol{\alpha} \cdot \boldsymbol{\sigma}_a$ commutes with any $\boldsymbol{\beta} \cdot \boldsymbol{\sigma}_b$, so Kochen and Specker's assumption implies that of Bell's NHVT. In particular, the assumption is violated in Bohmian mechanics.

## 23.3 Von Neumann's NHVT

John von Neumann presented a NHVT in his 1932 book.[48] It is clear that for a hidden-variable model to agree with the predictions of quantum mechanics, every $v_A$ can only

[47]S. Kochen and E.P. Specker: The Problem of Hidden Variables in Quantum Mechanics. *Journal of Mathematics and Mechanics* **17**: 59–87 (1967)

[48]J. von Neumann: *Mathematische Grundlagen der Quantenmechanik.* Berlin: Springer-Verlag (1932). English translation by R. T. Beyer published as J. von Neumann: *Mathematical Foundation of Quantum Mechanics.* Princeton: University Press (1955)

have values that are eigenvalues of $A$, and its marginal distribution must be the Born distribution. Von Neumann assumed in addition that whenever an observable $C$ is a linear combination of observables $A$ and $B$,

$$C = \alpha A + \beta B\,, \quad \alpha, \beta \in \mathbb{R}\,, \tag{23.3}$$

then $v_C$ is the same linear combination of $v_A$ and $v_B$,

$$v_C = \alpha v_A + \beta v_B\,. \tag{23.4}$$

**Theorem 23.3.** *(von Neumann's NHVT, 1932) Suppose* $2 \leq \dim \mathscr{H} < \infty$ *and* $\psi \in \mathbb{S}(\mathscr{H})$*, let* $\mathscr{A}$ *be the set of all self-adjoint operators on* $\mathscr{H}$*, and consider a joint distribution of random variables* $v_A$ *for all* $A \in \mathscr{A}$*. Suppose that* (23.3) *implies* (23.4)*. Then for some* $A$ *the marginal distribution of* $v_A$ *disagrees with the Born distribution associated with* $A$ *and* $\psi$*.*

As emphasized by Bell,[49] there is no reason to expect (23.4) to hold. For example, let $\mathscr{H} = \mathbb{C}^2$, $A = \sigma_1$, $B = \sigma_3$, and $C$ the spin observable in the direction at 45° between the $x$- and the $z$-direction; then $C = \frac{1}{\sqrt{2}}A + \frac{1}{\sqrt{2}}B$. However, the obvious experiment for $C$ is the Stern-Gerlach experiment in direction $\boldsymbol{n} = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$, whereas those for $A$ and $B$ would have the magnetic field point in the $x$- and the $z$-direction. Of course, the experiment for $C$ is not based on measuring $A$ and $B$ and then combining their results, but is a completely different experiment. Thus, there is no reason to expect that its outcome was a linear combination of what we would have obtained, had we applied a magnetic field in the $x$- or the $z$-direction. So von Neumann's assumption is not a reasonable one.

[49] J.S. Bell: On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics* **38**: 447–452 (1966)

# Contents