Introduction and overview

Quantum mechanics is our most successful physical theory. It underlies our very detailed understanding of atomic physics, chemistry, and nuclear physics, and the many technologies to which physical systems in these regimes give rise. Additionally, relativistic quantum mechanics is the basis for the standard model of elementary particles, which very successfully gives a partial unification of the forces operating at the atomic, nuclear, and subnuclear levels.

However, from its inception the probabilistic nature of quantum mechanics, and the fact that "quantum measurements" in the orthodox formulation appear to require the intervention of non-quantum mechanical "classical systems," have led to speculations by many physicists, mathematicians, and philosophers of science that quantum mechanics may be incomplete. Among the Founding Fathers of quantum theory, Einstein and Schrödinger were both of the opinion that quantum mechanics is in some way unsatisfactory, and this view has been amplified in more recent profound work of John Bell, among others. In an opposing camp, many others in the physics, mathematics, and philosophy communities have attempted to provide an interpretational foundation in which quantum mechanics remains a complete and self-contained system. Among the Founding Fathers, Bohr, Born, and Heisenberg maintained that quantum mechanics is a complete system, and a number of recent proposals have been made to improve upon or to provide alternatives to their "Copenhagen Interpretation." The debate continues, and has spawned an enormous literature. While it is beyond the scope of this book to give a detailed review of all the proposals that have been made, to set the stage we give a brief discussion of the measurement problem in Section 1, and we survey some of the current proposals to revise the interpretational foundation of quantum mechanics in Section 2.

The rest of this book, however, is based on the premise that quantum mechanics is in fact not a complete system, but rather represents a very accurate asymptotic approximation to a deeper level of dynamics. Motivations for pursuing this track are given in Section 3. The detailed proposal to be developed in this book is that *quantum mechanics is not a complete theory, but rather is an emergent phenomenon arising from the statistical mechanics of matrix models that have a global unitary invariance*. We use "emergent" here in the sense that it is used in condensed matter, molecular dynamics, and complex systems theory, where higher level phenomena (phonons, superconductivity, fluid mechanics, etc.) are seen to arise or "emerge" as the expressions, in appropriate dynamical contexts, of an underlying dynamics that at first glance shows little resemblance to these phenomena. Initial ideas in this direction were developed by the author and collaborators in a number of papers dealing with the properties of what we termed "generalized quantum dynamics" or, in the terminology that we shall use in this exposition,"trace dynamics." The purpose of this book is to give a comprehensive review of this earlier work, with a number of significant additions and modifications that bring the project closer to its goal. We shall also relate our proposal to a substantial body of literature on stochastic modifications of the Schrödinger equation, which we believe provides the low energy phenomenology, expressed in terms of experimentally accessible observables, for the pre-quantum dynamics that we develop here. A quick overview of what we intend to accomplish in the subsequent chapters is given in Section 4, and some brief remarks on the history of this project are given in Section 5.

Certain sections of this book are more technical in that they involve some knowledge of supersymmetry techniques and, although included for completeness, are not essential to follow the main line of development; these are marked with an asterisk (*) in the section head. The exposition of the text is based on dynamical variables that are matrices in complex Hilbert space, but many of the ideas carry over to a statistical dynamics of matrix models in real or quaternionic Hilbert space, as sketched in Appendix A. Discussions of other topics needed to keep our treatment self-contained are given in further appendices, and our notational conventions are reviewed in the introductory paragraphs preceding Appendix A.

1 The quantum measurement problem

Quantum mechanics works perfectly well in describing microscopic phenomena, and even in describing phenomena in which many particles act coherently in one or a small number of quantum states, as in Bose–Einstein condensates, superfluids, and superconducting Josephson junctions. Conceptual problems arise only when one tries to apply the rules of quantum mechanics simultaneously to a microscopic system and to the macroscopic apparatus that is measuring the state of the microscopic system; this is the origin of the notorious "quantum measurement problem." We shall give here a simplified, "bare bones" description of the measurement problem, taking as an example a variant of the familiar Stern–Gerlach experiment. (For a selection of papers on the measurement problem, see the reprint volume Wheeler and Zurek, 1983.)

Consider a source emitting spin-1/2 particles with polarized spins, so that all particles have spin component up along the x axis; that is, the initial beam is in a state with $S_x = 1/2$. (We shall see in a moment how this is accomplished in practice.) The particles then go through an inhomogeneous magnetic field aligned along the z axis, which splits the beam into two spatially displaced components, corresponding to components of the beam with spin component $S_z = 1/2$ and $S_z = -1/2$, as shown in Fig. 1a. The quantum mechanical description of what has happened so far is simply the spin state decomposition (with appropriate phase conventions)

$$|S_x = 1/2\rangle = \frac{1}{\sqrt{2}}(|S_z = 1/2\rangle + |S_z = -1/2\rangle). \tag{1a}$$

At this point *no measurement has been made*; if we pass the split beams through a second inhomogeneous field with the direction of inhomogeneity reversed, as in Fig. 1b, and devote great care to the isolation of the beams from environmental influences, the two components of the beam merge back into one and what emerges from the combined apparatus is the original state $|S_x = 1/2\rangle$. (An analysis of issues involved in achieving spin coherence, and further references, are given in Sculley, Englert, and Schwinger, 1989.)

To make a measurement, one must intercept one or both beams with a macroscopic measuring apparatus that absorbs the beam and registers a count in some form. When the measuring apparatus A intercepts both beams, we get the conventional Stern–Gerlach setup pictured in Fig. 1c. This is described, in the von Neumann (1932) model of measurement, by the evolution of the initial state $|S_x = 1/2\rangle|A_{\text{initial}}\rangle$ into a state in which the measured system and the apparatus are entangled

$$\frac{1}{\sqrt{2}}(|S_z = 1/2\rangle|A_+\rangle + |S_z = -1/2\rangle|A_-\rangle), \tag{1b}$$

where $|A_+\rangle$ is an apparatus state with a count shown on the upper counter and none on the lower counter, while $|A_-\rangle$ is an apparatus state with a count shown on the lower counter and none on the upper counter.

Once an apparatus intervenes in this way, two salient features become apparent. The first is that it is impossible in practice to coherently recombine the total system consisting of beam and apparatus so as to regain the initial state $|S_x = 1/2\rangle$. This feature, that the two legs of the apparatus have decohered, can be understood within the framework of quantum mechanics: since the apparatus state is a complex, large system, reversing the joint evolution of beams and apparatus with sufficient accuracy to preserve interference requires an unachievable control over the apparatus state. This is all the more so because in general the apparatus is in interaction with an external environment, into which phase coherence information is rapidly dissipated, making a coherent recombination of the beams a practical impossibility. In density matrix language, the off-diagonal components of the density matrix, when traced over the internal states of the apparatus and the environment, rapidly vanish because of decoherence effects, leaving just diagonal components that represent the probabilities for seeing the apparatus register an up or a down S_z spin component. (For further discussions of decoherence theory, see Harris and Stodolsky, 1981; Joos and Zeh, 1985; Zurek, 1991; and Joos, 1999.)

The second salient feature is that while there are definite probabilities for the apparatus to register a spin up or a spin down component, the outcome of any given run of a particle through the apparatus cannot be predicted; part of the time it registers in the "up" counter, and part of the time it registers in the "down" counter. (In the above example, the probabilities for registering "up" and "down" are both $1/2$, but for general orientations of the apparatus axis the probabilities will be $\sin^2 \theta/2$ and $\cos^2 \theta/2$, with θ the angle by which the inhomogeneous magnetic field is rotated with respect to the x axis.) This unpredictability of individual outcomes is the origin of the quantum measurement problem. If we maintain that quantum mechanics should apply to both the particle passing through the apparatus *and* to the measuring apparatus itself, then the final state at time t is described by a unitary evolution $U = \exp(-iHt)$ applied to the initial state, and this describes a superposition as in Eq. (1b), not an either–or choice between outcomes that are described by orthogonal states in Hilbert space. Since environmental decoherence effects still involve a unitary evolution (in an enlarged Hilbert space describing the system, apparatus, and environment), they cannot account for this either–or choice observed in the experimental outcomes. (See Adler, 2003b for a more detailed discussion of this point, and for extensive literature references. For an opposing viewpoint, see the review of Zurek, 2003.)

It is not necessary for the apparatus to intercept both beams for a measurement problem to be apparent. Consider the apparatus illustrated in Fig. 1d, which intercepts only the "down" leg of the experiment. If the particles are gated into the apparatus at definite time intervals, then a count on the "down" meter indicates that a particle has been detected there, and subsequent downstream measurements in the "up" leg will detect no particle there. If there is no count on the "down" meter (i.e., a "down" meter anti-coincidence), then one can say with certainty that the particle has passed through the "up" leg of the apparatus and is in a polarized state $|S_z = 1/2\rangle$; this is how one produces a polarized beam. Decoherence accounts for

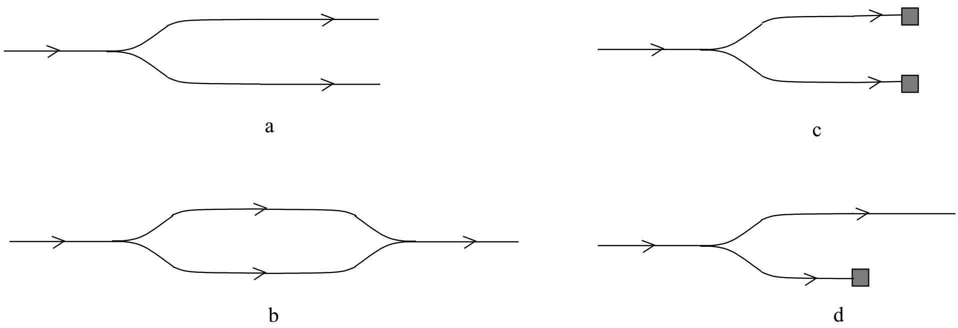

Figure 1 Beam paths through variants of the Stern-Gerlach experiment. Where the beams separate or recombine, there are magnetic fields that are not shown. a. Spin up and down components are separated and continue to propagate. b. Spin up and down components are separated, propagate, and then are coherently recombined. c. Spin up and down components are separated and each impinges on a detector. d. Spin up and down components are separated, the down component impinges on a detector, while the up component continues to propagate, producing a spin up polarized beam.

the fact that we cannot in practice reconstitute the original state $|S_x = 1/2\rangle$, but it cannot account for the stochastic pattern in which polarized particles emerge from the “up” leg of our apparatus.

There are two conventional ways to try to avoid the measurement dilemma just stated. The first is to assert that quantum mechanics has only a statistical interpretation, and should only be applied to describe the statistical properties of multiple repetitions of an experiment, but not to any individual run. However, with the advent of our ability to trap individual particles for long periods, and to manipulate their quantum states (e.g., the particle emerging from the “up” beam in Fig. 1d could be run into a trap, and manipulated there), this interpretation of quantum mechanics becomes dubious. The second is to adopt the Copenhagen interpretation, and to state by fiat that the unitary state vector evolution of quantum mechanics does not apply to measurement situations. One then adds to the unitary evolution postulate a second postulate, that of state vector reduction, which states that after a measurement one sees a unit normalized state corresponding to the measurement outcome $|f\rangle$, with a probability given by the Born rule $P_f = |\langle f|\Psi\rangle|^2$ as applied to the initial state $|\Psi\rangle$ being measured.

While perfectly consistent for all experiments that have been performed to date, the Copenhagen interpretation is at odds with the our belief that quantum mechanics should have universal applicability, and should describe the behavior of large systems (such as a measuring apparatus) as well as microscopic ones. It also has the bizarre feature of erecting a probabilistic theory, without an underlying sample space of individual events, the coarse-grained behavior of which is described by the probabilities. In all other applications of probability theory, probabilities emerge from the fact that one cannot observe, or chooses not to observe, individual details which deterministically specify the outcomes. Quantum mechanics is unique in that probabilities (or in some formulations, expectation values) are introduced as a postulate, without emerging by some well-defined rule from an underlying sample space of predictable individual events.

There are two logical possibilities for dealing with the problems just sketched. The first is to maintain that quantum mechanics is exactly correct, but in need of an improved conceptual foundation. One way to do this is to generalize the Copenhagen interpretation, so as to eliminate the apparently arbitrary distinction between "system" and "apparatus," and to give a set of extended interpretive rules with general applicability. This is the goal of the "consistent histories" approach to quantum foundations. Another way to do this is to extend the kinematic rules of quantum mechanics so as to give a concrete specification of a hidden sample space, that is constructed so as to be in principle unobservable, which leads to Born rule probabilities because full details of the sample space cannot be seen. This is what is done in certain versions of the "many worlds" approach, and in the Bohmian and Ax–Kochen approaches to quantum theory.

The second logical alternative is to consider the possibility that quantum mechanics is only a very accurate approximation to a deeper level of dynamics, which in turn gives a unified understanding of both unitary Schrödinger evolution and measurement dynamics. In this case the sample space that is created is not constructed so as to be unobservable, and detectable deviations from quantum mechanics become possible, leading to experimental constraints on the model parameters. As in any approach that proceeds by creating a sample space, there are so-called "hidden variables," and so important constraints imposed by no-go theorems coming from the work of Kochen and Specker (1967), Bell (1964, 1987), and others, have to be observed.

In Section 2 immediately following, we shall briefly describe the approaches that proceed from the assumption that quantum theory is exact but requires a new conceptual foundation. In Section 3 we shall give motivations for considering the possibility that quantum mechanics is in fact not an exact, final theory, which leads into the main themes of this book.

2 Reinterpretations of quantum mechanical foundations

A number of approaches to the reinterpretation of quantum foundations, assuming that quantum theory is exact, have been explored in recent years. Our aim in this section is to give a brief overview with entry points to the relevant literature, without attempting either a detailed exposition or a critique.

2.1 Histories

The histories approach is a generalization of the Copenhagen interpretation, that replaces the imprecise notions of an "apparatus" and a "measurement" with more precise concepts based on histories. The basic objects in this approach are time-dependent projectors $E_k(t_k)$ associated with events (defined as properties at given times) occurring in a history, and the probability of a history is then postulated to be given by

$$p = \mathrm{Tr}[E_n(t_n)\ldots E_1(t_1)\rho E_1(t_1)\ldots E_n(t_n)], \tag{2a}$$

with ρ the initial density matrix. This definition, supplemented by the notion of a family of decohering histories, which describes mutually exclusive evolutions with probabilities that sum to unity, can be argued to lead to all of the usual properties of quantum mechanical probabilities. In this interpretation, state vector reduction appears only as a Bayesian statistical rule for relating the density matrix after a measurement to that before the measurement. Detailed accounts of the histories approach can be found in the book of Griffiths (2002), the review and books of Omnès (1992, 1994, 1999), and the lectures of Hartle (1992). The histories approach involves no enlargement of the basic mathematical apparatus of quantum mechanics, and may still be relevant as a detailed description of quantum behavior even if quantum mechanics turns out to be an approximation to a deeper level of dynamics.

The three approaches that we discuss next all enlarge the mathematical structure of quantum mechanics, so as to create a sample space which forms the basis for the probabilistic interpretation. However, in all three cases the attributes that distinguish "individuals" in the sample space are not observable, so that there are no predictions that differ from those of standard quantum mechanics. Because these theories reproduce the results of quantum mechanics, it is evident that the assumptions of the Kochen and Specker (1967) and Bell (1964) no-go results are evaded. In the Bell case, for example, this results from nonlocality in the construction of the hidden sample space.

2.2 Bohmian mechanics

In Bohmian mechanics (Bohm, 1952), in addition to the Schrödinger equation for the N-body wave-function $\psi(q_1, \ldots, q_N, t)$ that obeys

$$i\hbar\frac{\partial\psi}{\partial t} = \left(-\sum_{k=1}^{N}\frac{\hbar^2}{2m_k}\nabla_{q_k}^2 + V\right)\psi, \tag{2b}$$

one enlarges the mathematical framework by introducing hidden "particles" moving in configuration space with coordinates Q_k and velocities

$$v_k = \frac{dQ_k}{dt} = \frac{\hbar}{m_k} \mathrm{Im} \nabla_{Q_k} \log \psi(Q_1, \ldots, Q_N, t). \tag{2c}$$

The state of the individual system is then specified by giving both the wave function and the coordinates Q_k of the hidden particles. If the probability in configuration space is assumed to obey the Born rule $p = |\psi|^2$ at some initial time, the Bohmian equations then imply that this continues to be true at all subsequent times. Arguments have been given that the Bohmian initial time probability postulate follows from considerations of "typicality" of initial configurations. For detailed expositions, see Bub (1997), Dürr, Goldstein, and Zanghi (1992), and Dürr, Goldstein, Tumulka, and Zanghi (2003).

2.3 The Ax–Kochen proposal

Ax and Kochen (1999) extend the mathematical framework of quantum theory to encompass the "individual," by identifying the ray with the quantum ensemble, and the ray representative, i.e., the $U(1)$ phase associated with a particular state vector, with the individual. They then give a mathematical construction to specify a unique physical state from knowledge of the toroid of phases. They argue that if the a priori distribution of phases is assumed to be uniform, then their construction implies that the probabilities of outcomes obey the usual Born rule.

2.4 Everett's "many worlds" interpretation

In the "many worlds" interpretation introduced by Everett (1957), there is no state vector reduction, but only Schrödinger evolution of the entire universe. In this interpretation, to describe N successive quantum measurements requires consideration of an N-fold tensor product wave function. The mathematical framework can be enlarged to create a sample space by considering the space of all possible such tensor products, and defining a suitable measure on this space. This procedure, given in the De Witt and Graham (1973) versions of many worlds, is the basis for arguments obtaining the Born rule as the probability for the occurrence of a particular outcome, that is, as the probability of finding oneself on a particular branch of the universal wave function.

Since the reinterpretations of quantum theory sketched here all aim, by construction, to reproduce the entire body of predictions of nonrelativistic quantum theory, they cannot be experimentally falsified (unless deviations from quantum theory are eventually established). Thus, apart from issues of the extent to which they can be generalized to encompass relativistic quantum field theory, the choice between them is somewhat a matter of taste. Rather than join in the already extensive literature debating their strengths and weaknesses, we shall proceed now to consider an alternative possibility, that quantum mechanics is in fact not an exact, complete structure.

3 Motivations for believing that quantum mechanics is incomplete

As surveyed in the preceding section, one approach to the quantum measurement problem and associated "paradoxes" of quantum theory is to continue to assume that quantum mechanics is exactly correct, and to attempt to supply it with a new foundational interpretation. However, there is another logical possibility, which is to suppose that quantum mechanics is not exactly correct, but represents an extremely accurate approximation to a qualitatively different level of dynamics. Since quantum theory is an extraordinarily successful physical theory, one can ask why try to replace it with something else? We respond to this question by listing a number of motivations for considering the possibility that quantum mechanics, and quantum field theory, may require modification at a deeper level.

3.1 Historical precedent

The historical development of physics contains many examples of theories that seemed to be exact in the context for which they were developed, only to require modification when applied to a larger arena of phenomena. Newtonian mechanics and Galilean relativity appeared to be exact in the context of planetary orbits, until the need for their special and general relativistic extensions became apparent in the early twentieth century. Classical predictability appeared to be exact in the context of classical mechanics, thermodynamics, and statistical mechanics, until confronted with the problems of the blackbody radiation spectrum and the discreteness of spectral lines at the end of the nineteenth century. The Landau mean field theory of critical phenomena was considered to be exact, until confronted with experimental data showing anomalous critical scaling, requiring the modern Kadanoff–Fisher–Wilson theory of critical phenomena for its explanation. Given these historical precedents, there seems to be no compelling reason to assume that quantum mechanics is immune to the general rule, that theories are only valid within a given regime, and may require modification when extended beyond that regime.

3.2 The quantum measurement problem

As we have discussed in Section 1, the unitary evolution of standard quantum mechanics does not describe what happens when measurements are made, but conventionally has to be supplemented by an additional postulate of nonunitary state vector reduction when a "measurement" is performed by a "classical" apparatus. As many authors have stressed, an economical resolution of the measurement "paradoxes" would be achieved if one could find a more fundamental underlying dynamics, from which the unitary evolution and the state vector reduction aspects of conventional quantum mechanics would emerge in a natural way in the appropriate physical contexts. Such a resolution should show in a natural way why quantum mechanics is probabilistic, by endowing it with an underlying sample space, and should show how probabilities become actualities for individual outcomes.

3.3 What is the origin of "canonical quantization"?

The standard approach to constructing a quantum field theory consists in first writing down the corresponding classical theory, and then "quantizing" it by reinterpreting the classical quantities as operators, and replacing the classical Poisson brackets by $-i/\hbar$ times the corresponding commutators or anticommutators. However, since quantum theory is more fundamental than classical theory, it seems odd that one has to construct it by starting from the classical limit; the canonical quantization approach has very much the flavor of an algorithm for inverting the classical limit of quantum mechanics. Moreover, it is known through the theorem of Groenewold and van Hove (for a recent review, see Giulini, 2003) that the Dirac recipe of replacing Poisson brackets by commutators cannot consistently be applied to general polynomials in the canonical variables, but only to the restricted class of second-order polynomials. Additionally, what is the origin of Planck's constant $\hbar$? One might hope that in a new theory underlying quantum mechanics, one would work with operators from the outset and proceed directly to operator equations of motion without first starting from the classical limit, and that one would also achieve an understanding of why there is a fundamental quantum of action.

3.4 Infinities and nonlocality

An outstanding problem in quantum mechanics (or more specifically, in quantum field theory) is the presence of infinities arising from the local structure of the canonical commutation/anticommutation relations, and an outstanding puzzle in quantum mechanics is the nonlocality seen, for example, in Einstein, Podolsky, and Rosen (1935) type experiments. Both of these considerations motivate many studies that have been made of quantum foundations, and in our view suggest that