Quantum Mechanics and Reality
An Interpretation of Everett's theory

A Dissertation
Submitted to the Department of Philosophy
and the Committee on Graduate Studies
of Stanford University
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

Christoph Albert Lehner
August 1997

ABSTRACT

The central part of Everett's formulation of quantum mechanics is a quantum mechanical model of memory and of observation as the recording of information in a memory. To use this model as an answer to the measurement problem, Everett has to assume that a conscious observer can be in a superposition of such memory states and be unaware of it. This assumption has puzzled generations of readers.

The fundamental aim of this dissertation is to find a set of simpler assumptions that are sufficient to show that Everett's model is empirically adequate. I argue that Everett's model needs three assumptions to account for the process of observation: an assumption of decoherence of observers as quantum mechanical systems; an assumption of supervenience of mental states (qualities) over quantum mechanical properties; and an assumption about the interpretation of quantum mechanical states in general: quantum mechanical states describe ensembles of states of affairs coexisting in the same system. I argue that the only plausible understanding of such ensembles is as ensembles of possibilities, and that all standard no-collapse interpretations agree in this reading of quantum mechanical states. Their differences can be understood as different theories about what marks the real state within this ensemble, and Everett's theory as the claim that no additional 'mark of reality' is necessary.

Using the three assumptions, I argue that introspection cannot determine the objective quantum mechanical state of an observer. Rather, the introspective qualities of a quantum mechanical state can be represented by a (classical) statistical ensemble of *subjective states*. An analysis of these subjective states and their dynamics leads to the conclusion that they suffice to give empirically correct predictions. The argument for the empirical adequacy of the subjective state entails that knowledge of the objective quantum mechanical state is impossible in principle. Empirical reality for a conscious observer is not described by the objective state, but by an Everettian relative state conditional on the subjective state, and no theoretical 'mark of reality' is necessary for this concept of reality. I compare the resulting concept of reality to Kant's distinction between empirical and transcendental reality.

ACKNOWLEDGMENTS

On the long way from my first interest in Everett's theory to the completion of the present thesis, I have been fortunate to have had the advice of many excellent teachers. I would like to express my thanks to Nancy Cartwright and Peter Galison for helping me to turn my vague fascination into a set of concrete questions, and to the members of my final reading committee for their support (and patience) while I was trying to find answers to these questions: to Fred Dretske, who always had an open ear for my theoretical and practical worries; to Yair Guttmann for countless discussions and valuable comments that shaped this dissertation; and especially to Patrick Suppes: Not only was he willing to take on the responsibilities of an adviser for a half-finished and not very straightforward dissertation project, but he played a big role in bringing it to completion with his vast knowledge, his open-mindedness for unconventional approaches, and never-ending encouragement and support.

For valuable discussions I am thankful to David Albert, Lanier Anderson, Jordi Cat, Meir Hemmo, and Martin Jones. Dagfinn Føllesdal and Peter Godfrey-Smith read the first draft of this thesis and gave me helpful comments. I had a wonderful time at the Stanford Philosophy Department, and I want to thank everybody there, my fellow students, the faculty, and especially the always supportive staff: Jill Johnston, Rosemary Leute, Nancy Steege, and Zita Zukowsky. For my livelihood, I am greatly indebted to the Department, the Mellon Foundation, and my parents.

I am immensely grateful to my friends and family for everything they have done to keep me safe and sane during this project: Eric Ganther typed parts of the manuscript. Daniel Harris, an accomplished writer himself, took pity on my writer's block and let me dictate the first sections to him, besides typing and proofreading major parts of the thesis. My parents, Rudolf and Dorothea Lehner, have given me so much love and support all through my life that a few words seem hardly sufficient to express my gratitude. Finally, Gregory Nelson has been by my side through all the terror and exaltation, feeding me when I was writing frantically and comforting me when I was not; and, most importantly, reminding me of all the beauty of life and that we don't understand a bit of it.

At this place, I also want to remember Karl Lorenz Beck. Without him, I would never have taken up this adventure. I wish he could see what it all has come to—but that might not

be as important as I used to think. At least, that is what I sometimes think can be learned from the enigmas of quantum mechanics.

TABLE OF CONTENTS

Chapter 1

Metaphysics and the Interpretation of Quantum Mechanics

This thesis is about the interpretation of quantum mechanics. However, after seventy years of attempts at interpreting quantum mechanics, no agreement has emerged as to what it means to give such an interpretation. In this introduction, I want to sketch what I see as the problem about understanding quantum mechanics that requires an interpretation and what such an interpretation should do. There is agreement that the present mathematical formulation of quantum mechanics is well developed and—mathematically—well understood. There is also widespread agreement that the empirical application of quantum mechanics has been extremely successful. Although it has proven to be one of the most intractable problems in philosophy of science to explain how theories apply to experimental practice and its results, it is fair to say that whatever the criteria for empirical success of a theory may be, quantum mechanics has not failed them or we would have heard from experimental physicists about it.

What else could be required of a theory than a clear formulation and its successful application to experimental practice? One could express the desire for an interpretation thus: knowing a mathematical formalism and being able to apply it in certain situations does not yet amount to an understanding of what it is that this formalism describes. It seems like quantum mechanics puts us into the situation of the uninitiated user of an elaborate piece of software who has learned to push certain buttons to make the computer do certain things, but who has no idea of what is going on inside the mysterious machine. However, while it is possible to learn about the inner workings of a computer, it may be intellectual hubris to expect to learn about the inner workings of nature. This sentiment was famously expressed by Richard Feynman:

> I am going to tell you what nature behaves like ... Do not keep saying to yourself, if you can possibly avoid it, "But how can it be like that?" because you will get "down the drain," into a blind alley from which nobody has yet escaped. Nobody knows how it can be like that.[1]

[1](Feynman 1965, 129), quoted after (Hughes 1989, 1)

What Feynman expressed here is a conviction that has been almost universally shared by physicists of his generation (and, mostly, up to this day): physics can give us successful predictions for experiments, it enables us to build all kinds of devices that operate just as we want them to, but to expect physics to tell us what 'nature is like' is a misguided demand of an unenlightened past for a god-like insight into the essence of things, which leads into nothing but fruitless speculation. This blind alley is known under the name of metaphysics.

The repudiation of metaphysics is not the invention of physicists, though: they had predecessors in the logical positivists who wanted to clear science (and meaningful discourse in general) of metaphysical ballast. To this end, they created a normative picture of science that has become known as the 'received view' (Putnam 1962). Two fundamental postulates of this picture were that scientific theories are axiomatic logical theories that operated on symbols without regard for their meaning, and that meaning is given to the terms of the theory only by a set of correspondence rules that connect theoretical terms to observable events.[2]

This clean-cut picture of physics started to crack on both ends by the 1950s: In physics, it turned out that the restriction of quantum mechanics to experimental predictions as the only aim was impossible in cosmology, which deals with a system not confined to a laboratory. The classical version of quantum mechanics was unsuitable for this task because it makes explicit use of acts of measurement in the formulation of the dynamics. This problem led Hugh Everett to formulate a version of quantum mechanics that treats measurement as a purely quantum mechanical interaction. But, as John Wheeler put it, this picture carried too heavy a metaphysical ballast to become widely accepted (Wheeler 1977). In order to give an internal account of measurement, it was seen as postulating the constant splitting of our world into countless copies of itself. Therefore, it overstepped the boundary between physics and metaphysics that the received view had erected and fell under Feynman's ban: a theory like Everett's was mere metaphysical speculation, not science. Nevertheless, it turned out that Everett's theory could be used to solve a long-standing puzzle about quantum mechanics: its relation to classical mechanics. Starting with the work of Zeh,[3] it was shown that within Everett's theory the constant interaction of a quantum mechanical system with its environment resulted in the emergence of a classical behavior of that system. This phenomenon, called decoherence, has increasingly become accepted as a field of research by theoretical physicists.[4]

---

[2]See the detailed discussion in (Suppe 1977, 16-61).
[3](Zeh 1971)
[4]See the references in section 3.2, which discusses decoherence in more detail.

In philosophy of science, the shortcomings of the received view became increasingly clear: the characterization of scientific theories as axiomatic systems of first order logic is too impoverished to give a realistic account of scientific theories, and the correspondence rules postulated by the received view are not sufficient to fix the meaning of theoretical terms. (Suppe 1977) Beginning with the work of Patrick Suppes, a new view of scientific theories emerged called the *semantic* view:[5] science should not be understood as producing a formal language, but rather as offering models of phenomena. Theoretical statements are justified semantically (because they are true of the models in the theory) and not syntactically, i.e. by deduction from a set of axioms. The models of physics are mathematical structures. They are models in two senses: they are mathematical models in the sense mentioned, i.e. giving truth conditions for the theoretical statements; on the other hand, they are representative (or iconic) models of the objects of the theory, i.e. they resemble the objects in some important respect, and can be used to represent them within the theory.

The semantic view gives a more realistic account of scientific theories.[6] However, it still does not give us a place for the notion of an interpretation of a theory that we seek to explain. Certainly, the mathematical model is an interpretation of the theory in a sense. But this does not seem to be the sense that we are after. In quantum mechanics, we have a beautiful mathematical model, the Hilbert space, and there still seems to be something that we do not understand about it. Take, for example, the notion of a superposition: in the Hilbert space the state of a system is represented by a vector. Because the Hilbert space is a linear space, we can form linear combinations of such vectors called linear superpositions. But what does such a linear combination mean? It seems that we need to look at the representational function of models to answer this question. However, in the case of quantum mechanics, the look at the phenomena is not very helpful. It is certainly true that the mathematical structure of the Hilbert space reflects a phenomenal structure of experimental outcomes (here, the probabilities for different measurements), but the phenomenal structure is not any less arcane than the mathematical structure of the Hilbert space. That nature *behaves* like that does not answer our question of *how it can be* like that.

For an answer to this question, a brief excursion is necessary. As Nancy Cartwright has argued in detailed case studies (Cartwright 1983), the application of theoretical models to physical phenomena is far from straightforward. It is the art of the physicist to build concrete models from the elements of the theory so that the concrete model simulates a given

[5]See, e.g., (Suppes 1967).
[6]For a detailed exposition, see Suppe 1989.

experimental situation. Cartwright calls such models simulacra, and compares them to the notion of a thinker toy model used by Cushing (Cushing 1982). This model building is a highly non-trivial task: the tinker toy model is neither a realistic picture of the system under consideration, nor, properly speaking, a mathematical model of the underlying theory. It uses all kinds of ad hoc assumptions, even assumptions that are in conflict with the theory. Nevertheless, Cartwright argues, without such model building fundamental theories cannot be applied to any real situations. I will call such models *phenomenal* models.

How does Cartwright's argument relate to the semantic view? Here is my attempt at reconciliation: the underlying theory furnishes us with elements to build a phenomenal model for a concrete situation. However, these elements are not mathematical structures, i.e. uninterpreted formal concepts. They are mass points, force fields, physical properties, and their likes, that is, concepts that are at least partially grounded in our experience. When we use such concepts in physics, we make substantial assumptions about what the world is like, and over time, we might be forced to revise these assumptions. I will call the totality of these assumptions the *physical model* of the theory. Unlike the mathematical model, which is a formal structure, the physical model is a conceptual model, and its elements are interpreted concepts that are never fully abstract.

In eighteenth century rigid body mechanics, for example, the concept of a rigid body was treated as elementary: Rigidity was part of the definition of a mechanical object, that is, part of the physical model of mechanics. With the development of continuum mechanics this concept was analyzed and explained as the effect of the inner forces of solid bodies, and it turned out that perfect rigidity was not only not a fundamental concept, it was an unrealistic idealization of the properties of solid matter. The concept of rigidity, part of the physical model of rigid body mechanics, becomes a phenomenon to be explained within continuum mechanics. We may think that all the assumptions in our physical models are of such provisional nature. Nevertheless, it seems hard to imagine what a theory without such a model would be like: It is this basic intuitive interpretation that relates the theory to our experience (we have a pretheoretical notion of rigidity, and therefore we know how a theory that contains this concept will relate to our practice.)

But all of this means that the two functions of a model in the semantic view are performed by two different structures: there is the mathematical model, given by a mathematical space and diverse mathematical structures defined on it (transformation groups, algebras, measures, and so on); and there is the physical model, given by mass points, rigid rods, forces, light rays, or what have you. The mathematical model gives a mathematical *representation* of the physical model. The physical model, on the other hand, gives an *interpretation* to the mathematical model. Without such a model, all we have are "formulas with holes in them, bearing no relation to reality." (Cartwright 1983, 159).

It is the interpretation through a physical model that embeds the phenomenal model of an atom, or a laser, or any concrete physical structure into a physical theory as electromechanics or quantum mechanics: Only after we have built a phenomenal model from the elements of the physical model, we can translate its structure into mathematical terms of the theory. (And again, this translation is not trivial at all. We use all kinds of approximations and simplifying assumptions to arrive at a mathematical model that we actually can work with.) It is through this embedding into a physical model that the phenomenal model leads to empirical predictions: Phenomenal models don't come with their own correspondence rules. We need the theory itself to tell us what the observational effects of the phenomenal model will be.

Take the Bohr model of the atom as an example:[7] it employs elements of the physical model of Maxwell's theory of electromagnetism to represent an atom, such as charged particles, or the electromagnetic field. And although the orbits of electrons in Bohr's model are not possible motions for charged particles in an electrical field (because they are restricted by quantum conditions), the model nevertheless portrays electrons as charged particles, their energies as electromagnetic potential and kinetic energies, and their interactions as electromagnetic interactions. It is only through this embedding that the Bohr model leads to empirical predictions: the underlying electromagnetic theory is necessary to relate the electron jumps in the Bohr model to another theory, Planck's quantum theory of radiation. And without this relation, the Bohr model would not be able to say anything about the relation of atomic spectra to the parameters calculated with the model.

Physical theories are not mathematical theories: the physical models are an integral part of physics. This explains why we hardly ever find the separate correspondence rules postulated by the received view in the practice of physicists: the elements of physical theories are not uninterpreted formal structures that need to be linked to observations by explicit correspondence rules. Hence, the meaning of theoretical terms is not given by correspondence rules. It is given by their interpretation in terms of a physical model.

What, then, is the physical model for a fundamental physical theory such as mechanics? What are the basic elements we need to build the countless phenomenal models

---

[7](Heilbron 1977)

that physics employs? I will look at this question in Chapter 2 in more detail, but I want to sketch the general answer already here: these basic elements are entities like objects and their properties, or states and their change. These are the kinds of things that fundamental mechanical theories are interpreted in: we say that a point in phase space represents a state, or that a one-parameter transformation represents its change. The theory of such entities has had a name since Aristotle's times: metaphysics. I will therefore, boldly, call the models of our fundamental physical theories *metaphysical models.* They are metaphysical because they interpret the mathematical models of these theories in metaphysical terms. But they are models nonetheless: we do not need to believe that they tell us how the world truly is in order to use them, just that the world is *like* these models say in certain regards. In theoretical physics, they are generally termed 'assumptions of the model', i.e. they are assumptions that we implicitly make by choosing a certain mathematical model. More importantly, they do not reflect a non-empirical knowledge about the world, but they are tools we construct to interpret the mathematical models. They are not prior to these mathematical models, and in the history of science they often have developed considerably later than the theory they interpreted. The notion of a field only slowly emerged in the electromagnetic theory of the nineteenth century, and it found its final formulation only in the theory of relativity; or, most famously, Newton's mechanics led to a long process of reformulating the mechanical picture of nature that lasted into the nineteenth century. In Chapter 2, I will try to make the notion of a metaphysical model more concrete by discussing such models in various mechanical theories.

A metaphysical model in this sense, then, is not something we can impose a priori on a physical theory. It evolves slowly from the complex process of applying a theory to concrete situations. Such a model, on the other hand, is not a useless formal embellishment: it is a tool we need for the construction of phenomenal models. But it should do more than that: the model of a fundamental theory has to enable us to embed other theories. The best-known example for this is statistical mechanics, which allowed thermodynamics to be embedded into classical mechanics. This was not a reduction in the sense of the received view: statistical mechanics employs fundamentally new principles that are not found in classical mechanics. But it formulates these principles on the elements of classical mechanics (ensembles of atoms seen as mechanical bodies), and therefore allows us to see how thermodynamics and classical mechanics are related.

This is the deeper sense why I propose the name of a 'metaphysical model': we know very well that we cannot deduce thermodynamics from mechanics, and much less chemistry, biology, or psychology. Nevertheless, there is a widespread conviction that all these sciences do not operate completely independently from the mechanical model of nature. Few people find animism in biology or dualism in psychology a promising paradigm. But it is not very clear what exactly the assumption of physicalism in these sciences means. A way to understand physicalism is as the principle that it should be possible to embed these sciences into the metaphysical model of mechanics.

This, then, is the answer I propose to the question of what an interpretation of quantum mechanics should be: it should be the description of a metaphysical model for quantum mechanics. Such a metaphysical model might revise some of the assumptions about nature embedded in the metaphysical model of classical mechanics. Nonetheless, the relation to the classical model has to be clearly formulated. The easiest way this can be achieved is by embedding one model into the other. Of course, this enterprise might fail, and we might have to live with a metaphysical dualism of a quantum and a classical realm. This was Bohr's conviction, and it has been, under the name of the 'Copenhagen interpretation', more or less the standard position ever since. However, the ongoing debate over the interpretation of quantum mechanics shows that it is not a very satisfying conclusion. And different from the other great dualism debate in the philosophy of mind, we have not even been able to establish anything like a clear border between the two realms.

According to my claim that metaphysical models are devices to construct phenomenal models, this lack of a coherent metaphysical model should have an impact on our ability to apply quantum mechanics to concrete phenomena. And I think that exactly this is the case in quantum mechanics: Bohr's interpretation basically offers two incompatible metaphysical models for quantum mechanics and the general rule that it can only be decided *ex post* (i.e. after an actual experiment is performed) which of these to apply. This is the thesis of wave-particle dualism. It is plausible that such a metaphysical model makes the building of phenomenal models a rather daunting task. Physicists have developed several strategies to deal with this problem. The most common is the building of semi-classical models, that is: one first constructs a classical model for a phenomenon and then, after a mathematical representation of this model is found, and generally after it has been solved, one constructs an approximate translation of this representation into a Hilbert space representation. One main problem about such methods is that they work generally only if we restrict our attention to a single particle. (A typical example is the Hartree-Fock method.) Because of this, the quantum mechanical treatment of complex systems is still in its infancy. This also has far-reaching consequences for the applicability of quantum mechanics to theories that it should clearly be applied to, e.g. general relativity and cosmology, chemistry,[8] or, for a more specific and very telling example, chaos theory.

For the optimists and heretics who want to pursue a unification program, there are two strategies: One is the attempt to embed quantum mechanics into the classical model. This was Einstein's aim and it is the goal of the different types of hidden variable theories and the modal interpretations. All of these interpretations try to construct an essentially classical metaphysical model underlying quantum mechanics. The main arguments against these attempts are various 'no hidden variable proofs' that try to show that no such model is possible. I will discuss several such arguments in the course of this thesis. While none of these arguments can exclude the possibility of any kind of hidden variable theory, they have established that any such theory has to admit some nonclassical elements. The main question for the classicist camp today is which toad is easiest to swallow.

The second strategy is to find a metaphysical model for quantum mechanics that is able to embed classical mechanics. Its early proponent was von Neumann.[9] Later, this enterprise was taken up by Everett and his defenders. However, because this group has almost exclusively consisted of physicists, the focus of their work has been the mathematical formulation of a "pure" quantum mechanics or quantum field theory, without trying to formulate a consistent metaphysical model. This thesis attempts a metaphysical interpretation of Everett's formulation. I believe it is a worthwhile task to find out what the world is like according to quantum mechanics, even if one, in the end, does not want to believe that ours is such a world.

[8]Theoretical chemists have found their own highly ingenious ways to deal with the application of quantum mechanics to molecular structures. They are mostly qualitative, but unlike many models in physics they don't rely on classical approximations.

[9]I will argue for this unorthodox claim in section 3.4

Chapter 2

The Interpretation of Mechanical Theories

Since the seminal work by Heisenberg, Jordan, Dirac, and Schrödinger, the Hilbert space has generally been accepted as the mathematical structure for the formulation of quantum mechanics.[10] Von Neumann's *Mathematische Grundlagen der Quantenmechanik was* not only the first work to give a rigorous exposition of the Hilbert space formulation of quantum mechanics, it also gave the first systematic attempt at the interpretation of this formalism. I will outline the interpretation of the Hilbert space formalism given by von Neumann before I turn to a discussion of Everett's model. This will lead to a specific setting for the measurement problem as a problem of the interpretation of quantum mechanics and give a framework for the solution of this problem proposed in this thesis. Von Neumann develops the interpretation of quantum mechanics in analogy to the interpretation of classical point mechanics and classical statistical mechanics.[11] Therefore, I will give an account of the fundamental mechanical concepts and their application to quantum mechanics in this chapter.

2.1 Classical Mechanics

Every mechanical theory assumes that it is possible to pick out a mechanical system that is distinguished from its environment. If the system is sufficiently isolated from its environment, that is, if the action of the environment on the system is negligible, then it is assumed that we are able to make predictions about the behavior of the system. Mechanics is the theory of the behavior of isolated systems. For example, we might consider a mechanical clock. The assumption that we have a mechanical system means that we are able to tell what is part of the clock and what is not and we are to shield its functioning from possible disturbances by the surroundings. The system of the clock would consist of the wheels, the springs, the dial, the hands but not the air between the wheels, the dust, etc. We assume that we can disregard the effect of the air and the dust on the functioning of the clock's parts. Another example of a system is a volume of gas in a given container. The container is not

---

[10] For an excellent, much more extensive introduction to the formalism, see (Hughes 1989).
[11] Von Neumann, ibid., p. 1: "Vor allem sollen die schwierigen und vielfach noch immer nicht restlos geklärten Interpretationsfragen näher untersucht werden. Besonders das Verhältnis der Quantenmechanik zur Statistik und zur klassischen statistischen Mechanik ist hierbei von Bedeutung."

part of the system but it isolates the system from its environment, i.e., the air surrounding the container.

Mechanical systems are represented by different mathematical structures in classical and quantum mechanics. In classical mechanics, a system is represented by its phase space, in quantum mechanics, by its Hilbert space. We will discuss these mathematical representations in more detail later. For now it is only important that they both are representations of the same metaphysical concept of *substance*. Compare, e.g., the concept of a system with Aristotle's notion of substance.[12] The main characteristics Aristotle gives for the concept of substance are:

- While it is the subject of attributes and predicates, it is itself not an attribute or predicate of something else.
- It is individual.
- While remaining one and the same, it is capable of admitting contrary qualities[13].

Mechanical systems are the fundamental and individual subjects of theoretical description in mechanics.

The counterpart to the concept of substance is the concept of attribute. In mechanics, this concept is represented by the concept of a variable. In general, we call mechanical variables all changeable properties of systems as opposed to the permanent and essential properties that define the system, for example, the mass of a mass point, the radius of a wheel in the clock or the number of molecules in the container of gas. Von Neumann distinguishes two kinds of attributes in mechanics: quantities (Größen) and properties. Their difference is that quantities have numerical values (the real numbers or a subset of them) whereas properties only have two values (yes or no). The variables in the example above are quantities. Examples for properties in von Neumann's sense would be the property that the temperature of the gas is under its condensation temperature, or that a wheel of the clock moves counterclockwise. Properties can be regarded as quantities that only assume two values (0 or 1), while quantities can be considered as sets of properties (e.g., 'the value of the quantity is smaller than a for any real number a).

The central idea of mechanical theories is that all variables are determined by a set of fundamental variables of the system, the so-called dynamical variables. I will call this assumption the mechanistic postulate. Often, the fundamental variables are the positions and

---

[12] Categoriae 2^a-4^b or Metaphysica Δ 1017b

[13] The last point concerns the relation of substances to time (which we will discuss in section 2.2).

momenta of its parts, but they can also be angles or angular momenta, or, in a field theory, the field strength at a specific point in space and time. Any other facts about the system are considered irrelevant. In our example of the clock, we would not give a geometrical description of a wheel or a description of what the wheel is made of but only its angular coordinate, i.e., the angle in a given frame of reference, and its angular velocity, i.e., the speed at which it is turning.

In the example of the gas, the dynamical variables are the positions and velocities of all the molecules in the gas. In this case, we are not in practice able to measure a single dynamical variable that is the position or velocity of one gas molecule because we cannot pick out a single gas molecule with any physical measurement apparatus that we have today. Here, the dynamical variables are practically unobservable. Still, it is assumed that they obey the same physical laws as observable positions and momenta (like those of macroscopic objects) and that they could be observed in principle.

The task of mechanics is to formulate the dependence of variables from the dynamical variables mathematically. It was, for example, an important achievement of Maxwell's kinetic gas theory to express temperature and pressure of an ideal gas in terms of the dynamical variables of its molecules. To be able to achieve this we have to define what kinds of mathematical objects suitably represent mechanical variables, i.e. find the algebraic rules that govern their relations. The algebra of the dynamical variables defines the fundamental structure of the mechanical theory. All other variables can be represented by algebraic expressions of the dynamical variables. In classical mechanics, the algebra of the dynamical variables is simply the familiar algebra of the real numbers (their addition, multiplication, and their powers). This means that we can do algebraic operations with the dynamical variables as if they were numbers, and that all other variables are simply real functions of the dynamical variables. Also the mechanical properties form an algebra: they are connected by the usual logical connectives and therefore form a Boolean algebra.

Because of this algebraic representation, a specification of the dynamical variables also specifies the other mechanical variables. Such a specification is called a mechanical state of the system. In the simplest case, this specification consists of giving numerical values to all dynamical variables. This is the case in classical point mechanics. In the example of the clock, the mechanical state would be given by giving values for the positions and momenta of all its parts.

The mechanical state is the mechanical correspondence to the concept of a proposition, i.e., a substance's having an attribute. The mechanistic postulate implies a kind of propositional atomism, i.e., that all propositions depend on a set of fundamental propositions, the mechanical states. The totality of all possible mechanical states forms the state space of the system. The mechanistic postulate implies that all properties can be represented by subsets of the state space, and all variables by functions on the state space. The logical operations on properties are mirrored by set theoretical operations on the subsets (e.g. the property $A \vee B$ is represented by the subset $M(A) \cup M(B)$). As is well known, also set theoretic operations form a Boolean algebra and this algebra is isomorphic to the Boolean algebra of the logical operations on properties, hence this representation is consistent.

While the mechanical state describes the reality of the system (in the sense of the actual matters of fact about it), its state space can be understood as giving an account of its possibilities (in an absolute sense, i.e., everything the system could possibly be, not just its possible future states). The state space therefore is the fundamental account of the nature of the system, i.e., all its intrinsic essential attributes. The dynamical variables are coordinates in the state space, that is, a specific way of describing elements of the state space. But this means that dynamical variables are not unique: because any state space allows for many different coordinate systems, there are many different sets of dynamical variables that can describe the same system.

2.2 Time and Space in Mechanics

So far, we have not explicitly addressed the role of time in mechanics. It turns out that time plays a fundamentally different role than the mechanical variables. Mathematically, this difference is represented by the fact that the mechanical state is a function of time, i.e., time is a free variable, while the mechanical variables are bound variables, their values at every point of time given by the mechanical state. There is a certain ambiguity of formulation here: we can either take the mechanical state as the function itself or as the value of the function at a given point of time. The former is more commonly called the trajectory of the system, and I will restrict the notion of a mechanical state to the momentary value.

The goal of mechanics is to establish relations between mechanical states at different times that allow the derivation of a mechanical state at one time from the mechanical state at another time. These relations are the equations of motion for the system. They are either given in integral form (like the principle of least action) describing possible trajectories, or in differential form, describing how the mechanical state changes with time. The mathematical equivalence of these two formulations shows that it does not matter whether we take trajectories or mechanical states as the fundamental propositions of mechanics. Corresponding to the choice of mechanical states as fundamental propositions, I will consider the equations of motion in differential form as the basic dynamical law of mechanics.

Notice that the equations of motion do not merely describe the actual motion of a system, but all of its possible motions. The actual motion of the system is determined by boundary conditions, i.e., the designation of some mechanical state as the actual one at a certain time. The equations of motion in the Hamiltonian formulation of mechanics can be expressed as first-order differential equations. This possibility has important consequences: under reasonable restrictions (no singularities), the equations have exactly one solution if one is given the mechanical state at one time. This uniqueness of the solution is taken as an expression of the determinism of classical mechanics: a complete specification of the dynamical variables at one time fixes all of the system's past and future variables, too.

It has been a long-standing problem in metaphysics how to explicate the role of time in mechanics. One possible way is to assume that objects at different times are not identical, so that time is really part of the intrinsic or essential specifications of the object. What we commonly think of as the same object at two different times are really two objects. Their 'identity' is really a relation (Lewis, 1986, 202-4). This position has been attractive to logicians and philosophers, because it removes the apparent contradiction between change and the logical law of contradiction (a thing cannot be both A and non-A). Nevertheless, it does not agree well with mechanics, because this relation of transtemporal identity cannot be based on mechanical variables. As we have seen in the definition of a mechanical system, it is part of the notion of the individuality of a mechanical system.

A second way of thinking of the temporal structure of mechanics is by taking systems as simple objects and differentiating their properties by their time of occurrence, i.e. the proposition that a system has property A at time t has to be properly understood as the proposition that the system has the property 'A-at-time-t'. This representation corresponds to the assumption mentioned above those trajectories, not mechanical states are the fundamental propositions in mechanics. Again, this leads to a problem of transtemporal identity, this time for the properties.

The most intuitive way to represent the role of time in mechanics is by assuming that physical systems have temporal parts very much like a spatially extended object has spatial parts. The mechanical state at a given time is a description of the part of the system at that time. This picture is called the geometrization of time. It does not attempt to reduce time to a specification of systems or of properties, and it takes the transtemporal identity of systems as a given. It therefore seems the most suitable to represent the special role of time in mechanics. The geometrization of time does not only allay the old worry of logicians how temporal change can be consistent with the law of contradiction, but it also sits well with the symmetry of space and time demanded by the theory of relativity. In this picture, the equations of motion define the relations between variables at different temporal parts of the system. The fact that these equations of motions can be expressed as differential equations means that these relations are defined locally, i.e., only the variables at contiguous temporal parts are related.

There is a fundamental intuition about time that is not represented adequately in this interpretation, though: namely, the tense structure of time, i.e., the qualitative difference between past, present, and future. This is not the place to go into the many metaphysical theories of time that have been offered in the long history of philosophy, but I think it is fair to say that not a single cogent metaphysical representation of our intuitions about the tense structure of time has emerged so far.[14] Nevertheless, there are two fundamental features about our experience of time that any metaphysical picture has to address:

- The future seems 'less real' than past and present. The idea that we can influence future events by our actions seems to require that, as of now, there is no matter of fact about future events. Past matters of fact, on the other hand, seem to be well-defined like present matters of fact.
- Even more elusive is the distinction between past and present: both are factual (in the sense sketched above); nevertheless, there seems to be a sense in which only the present is actual while the reality of the past is merely inferred.

Both these intuitions are not represented in the mechanical model. The actuality of the present could be represented by an extraneous statement "the time t_0 is now" which depends on the time of utterance. This statement would designate the 'temporally actual' state much as the boundary conditions designate the real state. However, while it seems unproblematic to think of the boundary conditions as representing an objective matter of fact, because their truth-value is universal, the definition of the present changes with time, and therefore seems to require another notion of temporal change outside of mechanics.[15]

---

[14] (Van Benthem 1983)

[15] Notice, though, that any such extraneous definition of the present does not agree well with special relativity: it introduces an absolute concept of simultaneity which makes the Lorentz symmetry of relativistic mechanics appear like a strange coincidence.

The distinction between past and future is even harder to represent in mechanics. Not only are future mechanical states represented in the same way as past states, but the determinism of mechanics does not seem to leave room for contingency of future states. Moreover, for most mechanical systems the equation of motion is symmetrical under time reversal.[16] Therefore, a dynamical definition of the distinction (through processes that have a definite direction in time) is also excluded. On the other hand, the distinction of past and future is closely connected to the second law of thermodynamics, which does explicitly introduce this distinction. This is a fundamental problem for the unification of thermodynamics and classical mechanics.

Classical point mechanics represents spatial properties of systems as variables, i.e., as properties of the mass points that make up the system. This is not a satisfying representation for spatially extended objects. Already Newton introduced the use of differential calculus to account for such cases: we can think of an extended object as made up by infinitesimal parts that only interact with their neighbors. Otherwise, this model (continuum mechanics) uses the same categories as classical mechanics. We can distinguish between local and global variables: local variables are defined on a single infinitesimal element, their values represented as functions over space; the global variables are defined as space integrals of the local variables.

Again, all local variables can be defined as functions of a set of local dynamical variables, called *fields.* The mechanical state is a specification of the fields, i.e., given by a set of functions over space. The assumption that a spatially extended object can be thought of as a composite of infinitesimal parts is reflected in the concept of the *separability* of the state: the state of a system always is the logical conjunction of the states of a complete set of spatial parts of the system. For the dynamics, the assumption that infinitesimal elements only interact with their neighbors leads to the assumption of *locality*: the evolution of the system can be expressed by a set of differential equations in space and time, i.e., only the values of the dynamical variables and their spatial differentials determine the dynamics of the system. Notice that in this model, space mathematically plays a similar role as time: it is a free variable, and properties are defined as functions of space. Unlike in the case of time, though, the interpretation of a system having spatial parts, and of all these parts coexisting, poses no difficulty.

---

[16]The criterion for reversibility is that the Hamiltonian is time-independent and an even function of the momentum, which is always fulfilled for isolated and dispersion-free systems.

Notice further that continuum mechanics takes its notion of a substance from classical mechanics, i.e., the infinitesimal elements are thought of as individuals, and as having transtemporal identity. When in the nineteenth century continuum mechanics was applied to electromagnetic phenomena and to light, it was therefore natural to assume such a substance as the bearer of these phenomena: the ether. Einstein's revolutionary act in the special theory of relativity was to emancipate the notion of a field from the idea of a substance whose property the field is. This move is reflected in the abandonment of the idea of a transtemporal identity of spatial elements, i.e., the ideal of a velocity of the ether. Nevertheless, even general relativity maintains the concept of space-time elements as individuals and as the bearers of properties (the fields).

In summary, the interpretation of classical mechanics is for the most part a rather unambitious endeavor: the fundamental concept of mechanics fit easily into a logical structure that agrees with basic intuitions of objects and their properties. The most problematic issue is the mechanical concept of time. We will return to this topic in the context of the interpretation of quantum mechanics, where I will argue that quantum mechanics, suitably interpreted, can offer a more satisfying representation of temporal change.

2.3 Statistical Mechanics

In classical mechanics, the state of a system is given by a complete specification of all its dynamical variables. However, such complete knowledge of the dynamical variables is generally neither possible nor necessary. In statistical mechanics, a more general notion of a physical state is used: the statistical state does not give a unique value for the dynamical variables, but a distribution over values, which is understood as giving the probabilities for different values. Mathematically, the concept of probability is well understood: it is a measure over the state space. The definition of a measure on a subset of the state space requires the definition of a suitable set of subsets, i.e. a restriction of the Boolean algebra of subsets, which is called a σ-algebra of measurable subsets. A measure maps these subsets on positive numbers (in our case, the probabilities).[17] Such statistical states have proven extremely useful for the treatment of systems with many degrees of freedom like our

[17]The classical text on the foundation of probability theory is Kolmogorov (1933)

container of gas, where it is practically impossible to determine the exact positions and momenta of every single molecule.

Nonetheless, they pose a problem for the interpretation of statistical mechanics: a state in classical mechanics can be interpreted as a proposition (a substance having an attribute, or a set of attributes) and ascribing a state to a system simply means asserting a fact. But statistical states have a fundamentally different structure: They do not give a value to a variable, but they define a function over a range of values for the variable. Mathematically speaking, the mechanical variables are treated as free variables in the statistical state, not as bound variables (as in the case of the classical mechanical state). This is similar to the role of time and space discussed in the last section, which lead to the problematic notion of the coexistence of all times. The problem seems even worse now: the coexistence of different values of a variable is a logical impossibility if variables are to be understood as attributes of a substance. This problem is not confined to mechanics: it applies to the use of probabilities quite generally. It has given rise to a lot of attempts to clarify the metaphysical status of probabilistic statements. I will only sketch the interpretations that are most relevant for mechanics.[18]

In the most straightforward interpretation, which goes back to Laplace's *Philosophical Essay on Probabilities* (Laplace, 1995), probabilities are an expression of our incomplete knowledge of the real mechanical state of the system. They are determined by the so-called "principle of insufficient reason": if we have no reason to predict the occurrence of an event rather than an alternative, we should give equal probability to both. But this principle fails to define a unique probability distribution for continuous variables. Here, the notion of equidistribution depends on the chosen parametization of the variable.[19] This problem has been perceived as a result of the fact that the principle of insufficient reason is merely an epistemic principle: it only gives a criterion for attributions of probability to our beliefs. Therefore, it is not surprising that it does not define a unique notion of probability.

There are two responses to this problem:

One can accept the fact that probabilities are not objective and attempt to build a theory of probability that does not presuppose that there is a unique and objective notion of probability. This is the premise for subjective theories of probability (de Finetti, 1937; Ramsey, 1950). Such an interpretation of probability is problematic for statistical mechanics,

[18]For an overview of interpretations of probability see Gillies (1973), Introduction
[19]A well-known case for this is Bertrand's paradox: there is no natural notion of equiprobability for the mixing ration of two liquids. See, e.g. (Mises 1957, 77-79)

where a well-defined notion of probability is of central importance. For example, the entropy of a system is defined by the distribution of its statistical state. Nevertheless, it is a well-defined physical quantity that plays a fundamental role in the thermodynamic equation of state.

Alternatively, one has to try to find another way of defining probabilities that can be understood as representing an objective matter of fact. The most widely accepted approach takes probabilities as properties not of a single system, but of an ensemble of systems. The law of large numbers of the mathematical theory of probability asserts that for a sufficiently large ensemble of probabilistic events, the frequency of an outcome is close to the probability of that outcome with a high probability. In the limit of infinite ensembles, the frequency converges towards the probability. Frequentists (Mises, 1957) have attempted to use this law for the definition of probability as the limit of the frequency of an outcome. This approach is quite plausible for many situations in statistical mechanics, where the system is actually composed of a large number of similar systems (e.g., a volume of gas is composed of a gigantic number of molecules of the same kind). A statistical state can then simply be understood as the idealized list of the actual ratios of molecules that have a given value of a variable. In this case, the statistical state is an abbreviated way of writing a large number of mechanical states for single molecules. But, strictly speaking, the use of the law of large numbers already presupposes a measure on the state space, because it only makes probabilistic statement for frequencies. Therefore it already assumes a concept of probability.

Furthermore, statistical mechanics is not limited to such simple cases, and Gibbs has demonstrated that the power of statistical physics can be greatly improved if we take the whole of the system as the basis of our description (not the single molecule) and consider a statistical state in the state space of the whole system (Gibbs 1902). Here, the simple interpretation of the statistical state in terms of ratios of molecules is impossible. Gibbs himself invoked the notion of an imaginary ensemble of systems, but this move again opens up the question: how can the so defined probabilities be objective features of the physical system? One possible answer is given by ergodic theory that takes probabilities not as defined by frequencies in simultaneous ensembles but by frequencies in the temporal succession of states of a system. While this approach makes the probability distribution an objective feature of the single mechanical system, it still is not a mechanical property as defined in the last section, because it is not a function of any one mechanical state but a time integral over many such states.

Notice that all these interpretations of statistical states are interpretations in term of classical mechanics, i.e., they formulate the statistical state under the assumption that an objective description of a system can be given by its mechanical state. Especially in the context of the interpretation of quantum mechanics states, this premise has been called the 'ignorance interpretation' of statistical states because it assumes that the statistical description of a system implies our ignorance about the objective mechanical state of the system. Regarding the distinction between subjective and objective interpretations, this coinage can be misleading because only subjective interpretations understand probabilities as epistemic concepts, and the notion of 'ignorance interpretation' seems to imply an epistemic explanation of probability. What is really meant is simply an interpretation in terms of classical mechanics.

Nevertheless, none of the interpretations above offers a way to understand the statistical state as a property of the individual system in the mechanical sense, whether they interpret statistical states subjectively or as properties of some ensemble. But this leads to a kind of epiphenomenalism for statistical states: because the fundamental assumption of mechanical dynamics is that the future development is completely determined by the mechanical state, it seems impossible that a statistical state can play a causal role, and neither can a intrinsically statistical quantity like entropy. The problems about the interpretation of time and of probability are connected in a rather mysterious way in the second law of thermodynamics, which says that in any thermodynamical process, the entropy never decreases with time. But this means: entropy, which does not seem to be a mechanical property of systems, distinguishes between future and past, which classical mechanics cannot. What could be the missing part?

2.4 The Hilbert Space

We will now turn to the interpretation of the Hilbert space formalism. From early on, quantum mechanics was perceived as necessitating a revision of the fundamental metaphysical ideas in mechanics. But at the same time, there are striking similarities in the structure of quantum mechanics and classical mechanics. It is plausible to assume that analogous structures in both theories should be interpreted in the same way. This is not a far-fetched methodological claim: it was exactly around these metaphysical analogies that quantum mechanics developed.[20] The question is therefore how far these analogies can be taken and where the difference of the formalism forces us to abandon metaphysical tenets of classical mechanics. In this section I will follow von Neumann's analysis of the analogies and disanalogies between quantum and classical mechanics.

Quantum mechanics uses exactly the same concept of a mechanical system as classical mechanics, i.e. the notion of individuated substances that persist through time. This may seem surprising at first. Quantum mechanics, after all, is perceived to be a theory about atoms and elementary particles, the microscopic constituents of matter. One would expect, therefore, that there is a statement in the theory to this effect: that it is a theory about atoms and not about cuckoo-clocks. But, as a matter of fact, there is no such statement: every system that can be described by classical mechanics, can be described by quantum mechanics as well.[21]

This fact will be of central importance to the later argument because once we have convinced ourselves that quantum mechanics is the correct theory to apply to atomic phenomena, there is nothing in the theory itself that keeps us from applying it to all other mechanical systems, too. (Of course, there might be extraneous reasons to limit the scope of the theory.)

The representation of variables in quantum mechanics is fundamentally different from that in classical mechanics. This reflects the fundamentally new phenomenon that forced the abandonment of classical mechanics in favor of quantum mechanics: the quantization of variables. It was the seminal discovery of quantum mechanics that this fact can be expressed by representing variables by linear operators (Heisenberg 1925). The values of the variables are then represented by the eigenvalues of these operators. As a special case, properties (in von Neumann's sense) are represented by projection operators, i.e. operators which only have the eigenvalues 0 and 1.[22]

---

[20]For example, the formal analogy between Hilbert space operators and classical variables given by their algebraic properties is the main heuristic tool in 'quantizing' equations of motion.

[21]It is this notion of a system that limits the applicability of quantum mechanics to particle physics. The indistinguishability of elementary particles conflicts with the notion of individuality. While this problem can be dealt with by introducing extraneous symmetry conditions imposed on the status of these particles, the possibility of creation and destruction of particles cannot be represented within the formalism of point mechanics at all. This is taken as an indication that not point mechanics is the fundamental theory of elementary particles but field mechanics.

[22]Again it can be shown that both notions are mathematically equivalent, but here the equivalence is far from trivial and only valid under certain restrictions. The proof of the mathematical equivalence of Hermitian operators with sets of projection operators is discussed by von Neumann (1932), 59-88, and forms the central part of his discussion of the eigenvalue problem, that is the question under which circumstances variables represented by operators can have values.

Linear operators are defined as mappings of a linear space onto itself. Therefore, the fundamental mathematical structure in quantum mechanics is a complex linear space, the Hilbert space. The linearity of the Hilbert space means that it allows for the addition of two elements and the multiplication of an element with a complex number. Furthermore, a scalar product is defined on the Hilbert space which maps pairs of elements on complex numbers. This scalar product defines the notion of orthogonality of two elements.
Quantum mechanical states are defined as one-dimensional subspaces of the Hilbert space: if two elements are only distinguished by a complex factor, they represent the same state. Linear combinations of two states are called superpositions of these states. This leads, intuitively speaking, to a large number of new elements in the space. A central problem in the interpretation of quantum mechanics is to understand this feature of the Hilbert space in metaphysical terms.

This structure implies that variables don't have values for all states.[23] Each eigenvalue of an operator defines a subspace of the Hilbert space, and subspaces belonging to different eigenvalues are orthogonal. This fact has an important corollary: only states that are orthogonal can be distinguished by the values of a variable or, equivalently, there is no property that can distinguish between two nonorthogonal states (in the sense that one state has the property while the other one does not).

The connection between variables and states is not as direct as in classical mechanics: we cannot define all states by their values for a set of dynamical variables because there is no variable that gives values to all states. Instead, the role of dynamical variables is taken by non-degenerate operators, that is operators who have for every eigenvalue a one-dimensional subspace of eigenstates. Such an operator defines a basis on the Hilbert space, and every element of the Hilbert space can be represented by its complex coefficients in this basis. Position is the standard example of such a non-degenerate operator and the representation of a Hilbert space element in position basis is called a wave function.[24]

The operators form an algebra with a noncommutative multiplication given by the concatenation of two operators. Two operators A, B commute if $AB = BA$. If and only if two operators commute, there is at least one basis whose elements are eigenstates of both

---

[23]Properly speaking, operators associated with unbounded variables have no eigenstates at all. Dirac introduced improper states (Delta functions) as eigenstates of such operators, while von Neumann insisted that this fact represented the only finite sharpness of values of unbounded variables.
[24]The basis vectors are only defined up to a complex phase by the non-degenerate operator. Therfore, properly speaking, we need a second noncommuting operator (the standard example is momentum) to define the relative phases of the basis. Theis leads to a situation analogous to classical mechanics, where the dynamical variables also come in pairs.

operators. The algebra again makes it possible to express variables as functions of the dynamical variables if we assume the mechanistic postulate, which now has the form (because of the linearity of the Hilbert space): every variable can be expressed by a linear operator on the Hilbert space, and every property by a subspace of the Hilbert space. But in the case of quantum mechanics, the mechanistic postulate has an important consequence: subspaces of the Hilbert space don't form a Boolean algebra, but a so-called partial Boolean algebra. The main difference is that the joint of two subspaces A∪B is not a subspace itself. Therefore, one defines instead the smallest subspace that contains A∪B as the *span* of A and B, written A⊕B. But this operation does not fulfill the distributive law of Boolean algebra, i.e. (A⊕B) ∩ C ≠ (A∩C) ⊕ (B∩C) for some A, B, C. The mechanistic postulate now implies that also properties in quantum mechanics don't form a Boolean algebra, i.e. their logic is not the familiar logic, but a non-distributive logic called quantum logic.[25] Many of the nonclassical phenomena in quantum mechanics can be represented by the difference between quantum and standard logic.

The representation of the dynamics of a system is given by a set of differential equations exactly like in classical mechanics. As in classical mechanics, these equations of motion can be represented by the system's Hamiltonian, i.e. the energy expressed as a function o the dynamical variables. In the Hilbert space, this leads to an especially simple characterization of the dynamics of a system: it can be described as a unitary evolution $\psi(t) = U(t)\psi(0)$ of the state vector ψ, where *U* is a unitary operator. This representation makes it obvious that the evolution is deterministic, like in classical mechanics.

So far, the interpretation of the Hilbert space formalism was analogous to the interpretation of the (Hamiltonian) formalism of classical mechanics, although the formalisms are quite different. The fundamental discrepancy is given by the fact that operators do not attribute values to all states. This is an obvious lacuna in the interpretation, because if a variable is measurable, then we should be able to measure it on any state of the system. Being measurable, after all, means that we can conceive of a procedure to measure the value of the observable. This procedure should (at least in certain limits) not depend on the state the system is in.

This question is answered by Born's rule (Born, 1926): If a property *A* is measured on any state ψ of a system, the probability for finding the property is given by $\langle\psi, P(A)\psi\rangle$,

---

[25]See Hughes (1989) 178-217

where $P(A)$ is the projection operator representing the property. Because of the spectral decomposition theorem this rule also gives the probability for the value α of any variable (quantity) A as $\langle\psi, P_A(\alpha)\psi\rangle$.

This rule gives empirical predictions for the measurement of quantum mechanical states, but it is not an interpretation in our sense, i.e., it does not make an assertion about the nature of the relation between variables and their values, but it gives empirical predictions for their measurement. Therefore, it leads to the question whether it can be taken as the basis of an interpretation of quantum mechanical states as statistical states in the sense of section 2.3.[26]

Before we turn to von Neumann's treatment of this question, I will give a brief review of the representation of statistical states in quantum mechanics which again is quite similar to what was said in section 2.3. First of all, it has to be noticed that an appropriate notion of probability in quantum mechanics has to accommodate the linear structure of the Hilbert space: Probability cannot be defined as a measure on subsets of the Hilbert space, but it must be defined on the linear subspaces. This leads to a generalized notion of probability, because the subspaces don't form a σ-algebra (compare footnote 8 in 2.3): Instead of the operation of joining two subsets we have the operation of the direct sum (span) of two subspaces, because the joint of two subspaces is not a subspace itself.

Gleason's theorem[27] asserts that any so defined probability measure on a Hilbert space (of dimension greater than two) can be represented by a positive definite Hermitian operator of trace 1. Such an operator is called a density operator. Especially, this means that a probabilistic distribution p_n over quantum mechanical states ψ_n is represented by a weighted sum $\rho = \sum_n p_n P(\psi_n)$ of the projection operators $P(\psi_n)$ defined by ψ_n (von Neumann, 157-8). On the other hand, ever density operator can be written in this form (von Neumann, 174). Therefore, density operators can always be understood as representing probabilistic mixtures of pure quantum mechanical states or, shorter, mixed states. This representation is not unique though, i.e., there is more than one way to represent a quantum statistical state (density operator) as a probability distribution over pure states. This feature distinguishes quantum statistical states from classical statistical states, where the representation is always unique.

---

[26]Von Neumann and Born see this rule as the central statement of a 'statistical interpretation' of quantum mechanics, but von Neumann makes very clear that this is not to be understood in our sense, i.e. as an interpretation of quantum mechanical states as (classical) statistical states. (Von Neumann (1932) 108-9). The latter is the meaning of 'statistical interpretation' that is commonly used nowadays.

[27]Gleason (1957)

We will return to the interpretation of statistical states in the context of the discussion of composite systems in the next chapter.

Back to the question of whether pure quantum mechanical states can be understood as statistical states. Von Neumann distinguishes these possibilities along the lines of the two possible ignorance interpretations of statistical states we discussed in the last section.

First, one could assume an ensemble interpretation of quantum mechanical states, i.e., one could explain the probabilistic dispersion of measurement outcomes postulated by Born's rule by assuming that the quantum mechanical state is not an irreducible pure state of a single system (or at a single time), but represents an ensemble of systems on which the measured observable has different values with frequencies given by the Born rule. Von Neumann shows that this interpretation is not possible if we assume that the Born rule is valid for any Hermitian operator defined in the Hilbert space. (v. Neumann, ibid., 167-71) Von Neumann's argument rests on the assumption that the value assignments obey the algebraic rules for all variables, that is even if two operators *A, B* don't commute, the expectation values given by the Born rule satisfy the equation $\mathrm{E}(A+B)=\mathrm{E}(A)+\mathrm{E}(B)$. This assumption has been criticized, because in general, we don't have an independent method of measurement for each possible quantum mechanical operator.

Kochen and Specker showed with an example that for any Hilbert space of dimension greater than two there is no value assignment for three noncommuting observables on a finite set of states (exactly, 117, in their example) that correctly reproduces the prediction of Born's rule,[28] weakening the premise of Gleason's theorem which only considers value assignments for all variables on the system and weakening von Neumann's assumption above: they only assume that the expectation values of *commuting* operators are additive. Of course, even in the face of Kocher and Specker's argument one could maintain the assumption that all variables have values if one gave up the assumption that these values were faithfully represented by the measurement outcomes. But this move would defeat the primary purpose of a statistical interpretation of quantum mechanics: to understand the quantum mechanical state as a statistical state.[29]

The second possible interpretation of the probabilistic dispersion is on the lines of a subjective interpretation of probability, namely, that the distribution of values in the Born rule is an effect of an inability to exactly measure the true value: the Born rule describes the

---

[28]Kochen and Specker (1967), see Hughes (1989) 162-175 for a discussion.
[29]We will discuss such "hidden variable" interpretations in section 4.3.

statistical distribution of a measurement error. This liberates us from the requirement that the probabilities in the Born rule be representable as frequencies in a statistical ensemble. But as von Neumann observes (von Neumann, ibid., 110-112), the measurement error would also affect the measurement of two compatible variables (i.e., variables whose operators commute). Therefore, it should be impossible to obtain exact correlations between the measurements of two compatible observables. As a special case, if one measures the same variable twice, the results of both measurements should not be correlated, but each should be distributed independently according to Born's rule. But this conclusion is in contradiction to our empirical evidence: it is possible to find correlations in variables that are much more exact than the spread of the distribution of these variables given by the Born rule. Von Neumann discusses the scattering experiments of Compton and Simon as an example. Here, the momenta of the scattered electron and photo are exactly correlated, but they are spread over a wide range of possible values. Therefore, also a subjective interpretation is ruled out.

Taken together, these two arguments mean that an ignorance interpretation of the probabilistic distribution of values for a variable given by the Born rule is impossible. A measurement cannot simply be the recording of an existing value of the observable, because there is no consistent way in quantum mechanics to attribute values to noncommuting variables, and explaining the inconsistency with an inexactitude of the measurement is in conflict with experimental evidence.

There is an alternative interpretation of the dispersion of values in a quantum mechanical state. It takes the state as a field, which for a single particle is quite plausible: the state can be described by a complex function in space, the wave function, and its unitary evolution can be expressed as a differential field equation, the Schrödinger equation. Variables that are not functions of position, such as momentum, can be understood as field operators, i.e. global variables of the field. This interpretation, known as wave mechanics, was advocated by Schrödinger and de Broglie, and was an extremely important heuristic tool in early quantum mechanics. Its main problem is to account for the probabilistic nature of the Born rule: While variables on fields show dispersion, the different values exist simultaneously (like different momenta in a wave), not stochastically. A further problem for the plausibility of a field interpretation is that many-particle systems cannot be described by fields in three-dimensional space.

Instead, von Neumann proposes a dynamical interpretation of the probabilities, which is called a propensity interpretation: if a state is not an eigenstate of an operator, the corresponding variable has no value for that state. But in the case of a measurement of the variable, the state 'produces' a value indeterministically. That is: it has a *propensity* for different values without any value being realized before a measurement is performed. The probabilities of the Born rule describe the strength of this propensity for different values.

It is instructive to look at the Compton-Simon experiment in yet another way: It can be interpreted as two successive measurements of the same variable. The existence of exact correlations means then that the state of the system after the first measurement must have changed: the second measurement does not show a dispersion in the measured variable anymore. Therefore, it can be shown the state after the first measurement must be an eigenstate of the measured observable (von Neumann, ibid., 112-114).

Von Neumann takes this as a general principle: the measurement of a variable changes the state into an eigenstate of the measured variable with a probability given by the Born rule.[30] This principle is called the reduction postulate and the eigenstate after the measurement the reduced state. Because the different possible reduced states are distributed probabilistically according to Born's rule, the resulting state can be represented by a density operator which is nontrivial (i.e. not a pure state) unless the initial state of the system was an eigenstate of the measured observable. This means that the process of measurement cannot be described by a unitary evolution, which always transforms pure states into pure states. This is quite obvious when we remember that unitary evolution is a deterministic process: Von Neumann's argument against an ignorance interpretation of quantum mechanical states implies that the process of reduction must be irreducibly indeterministic. But as we remarked earlier, quantum mechanics assumes that every temporal evolution of a system can be described by a deterministic evolution of its state. Therefore, von Neumann has to assume that two fundamentally different kinds of dynamical evolution exist in quantum mechanics: a 'normal', deterministic evolution and a special kind of indeterministic evolution that only takes place if a measurement is performed.

The impossibility of an ignorance interpretation of pure quantum mechanical states leads to a rather puzzling metaphysical picture: at any given time, there is no matter of fact about the value of most variables. It is only through acts of measurement that variables acquire a value. But this leaves us with several open questions: What kind of processes are

---

[30]This assumption of the repeatability of quantum mechanical measurement has been criticized, because many actual measurements are not repeatable: they change the measured state even if it is an eigenstate of the measured variable. But as we will see in the next chapter, non-repeatable measurements can be easily modeled in quantum mechanics once we have a model of repeatable measurements. Von Neumann's interest in this section is not to give the most general model of a physical measurement, but to discuss the problem of values in quantum mechanics. For that, the decisive fact is the *possibility* of repeatable measurements.

measurements and how can they have the power to actualize values of variables? How is it to be understood that measurement does not seem to obey the law of unitary evolution? And finally: macroscopic objects seem to have well-defined values for all their variables at any time. How is that possible if we assume that quantum mechanics is the more fundamental theory and underlies also macroscopic phenomena?

This problem is illustrated by Schrödinger's cat paradox: Schrödinger shows us a way to prepare a superposition of two states of a cat, one in which the cat is alive and one in which it is dead. Do we have to conclude that there only is a matter of fact about whether the cat is alive or dead once someone performs a measurement on the cat (or at least looks at it)?
The complex of these questions is known as the measurement problem. What feature could give measurement such a special status as a physical process? Of course, a measurement is not simply the evolution of an isolated system, but it involves the interaction with a measurement apparatus. Therefore, we have to discuss the quantum mechanical representation of interactions on a system before we can address these questions.

Chapter 3

The Measurement Problem

3.1 Composite Systems

It is a surprising feature of quantum mechanics that it makes nontrivial predictions about the behavior of composite systems.[31] Whereas in classical mechanics any composite system is completely characterized by describing the states of its parts, this is not true in quantum mechanics. It is easy to see why this is so. If a composite system S consists of two subsystems S_1 and S_2, then of course every pair of states ϕ of S_1, ψ of S_2 is a possible state for the composite system. This state is called the direct product of the subsystem states and is written $\phi \otimes \psi$. But then it follows from the linearity of the Hilbert space that also every linear combination of such direct products must be a possible state of the composite system, that is: if ϕ_i and ψ_i are possible states of the subsystems, and c_i are complex numbers, then

$$\omega = \Sigma_i c_i \phi_i \otimes \psi_i \qquad (3.1)$$

represents a possible state of the composite system. But ω in general cannot be written as a single direct product of any two states of the subsystems $\phi' \otimes \psi'$. This means: in general the composite system can be in a state that cannot be described by giving a pure state for each subsystem. Such states are called entangled states. There are no such states in classical mechanics, where any pure state of a composite system is fully described by the state of its components. This very important feature is often called the *holism* of quantum mechanics, as opposed to the atomism of classical mechanics.

There is an analogue in statistical mechanics to entangled states, though: Generally, a joint distribution of several variables cannot be written as a product of distributions over the single variables. This can be understood by observing that the joint distribution does not only tell us about how the values of every single variable are distributed, but also how the values for different variables are correlated. This means: assume we know the joint distribution of two variables, then we can calculate the *conditional* distribution of one variable given a distribution over the other. In quantum mechanics, we can make an analogous statement, and it turns out that if the given state ω of the composite system is pure, and if system S_1 is in a

[31] The formalism of composite systems is described in von Neumann (1932), ch. VI.2

state η, then the conditional state of system S_2 is also pure. It is called the *relative state* to η and is defined by:

$$\psi_{\text{rel}}(\eta) = N\sum_i \langle \eta \otimes \psi_i, \omega \rangle \psi_i \tag{3.2}$$

where N is a normalization constant, and ψ_i is any basis of S_2 (Everett 1973, 38).

All possible states for a composite system (given in the construction in (3.1)) form a Hilbert space, called the tensor product space of the Hilbert spaces of the subsystems H_1, H_2, written $H_1 \otimes H_2$,. It is easy to construct a basis for the tensor product space: if ϕ_i and ψ_j are bases of the subspaces, then the vectors $\phi_i \otimes \psi_j$ form a basis for the product space. The scalar product of two vectors $\phi_i \otimes \psi_j$, $\phi_k \otimes \psi_l$ is defined as $\langle \phi_i, \phi_k \rangle \langle \psi_j, \psi_l \rangle$. For all other states, the scalar product follows from linearity. The tensor product of two operators A on H_1 and B on H_2 is defined by the equation

$$(A \otimes B)(\phi_i \otimes \psi_j) = A\phi_i \otimes B\psi_j \tag{3.3}$$

and linear combinations. Again, there are operators on H that cannot be written as a tensor product of two operators on the subspaces, but any operator can be written as a sum of such tensor products. The operators A on H_1 and $A \otimes \mathbf{1}$ on H are equivalent, i.e. they have the same expectation values for any states ϕ on H_1 and ψ on H_2:

$$\langle \phi \otimes \psi, (A \otimes \mathbf{1})(\phi \otimes \psi) \rangle = \langle \phi, A\phi \rangle \langle \psi, \psi \rangle = \langle \phi, A\phi \rangle \tag{3.4}$$

Therefore we regard A and $A \otimes \mathbf{1}$ as representing the same physical quantity.

The fact that a general state of the composite system cannot be simply decomposed into two pure states of the subsystems leaves us with the question how to describe subsystems by a quantum mechanical state, if we are given the state of the composite system. After all, we can always make separate measurements on each of the subsystems and hence be able to deduce a state description of the subsystems from these measurements. Note especially that a composite system can be formed by any two systems even when these are perfectly isolated and noninteracting. In this case, it follows from our basic assumptions about mechanical theories that each of the subsystems must have a well-defined state. It can be shown (ibid., 41) that for any state ω of the composite system there exists a mixed state ρ_ω for subsystem S_1, so that the expectation value $\mathrm{Tr}(A\rho_\omega)$ for any variable A on S_1 is the same as the expectation value $\langle \omega, (A \otimes \mathbf{1})\omega \rangle$ for the equivalent variable $A \otimes \mathbf{1}$ on the composite system.

But this also means that any probabilistic prediction for the system S_1 that the two states yield will be the same. $\rho_\omega^{(S_1)}$ is always uniquely defined:

$$\rho_\omega^{(S_1)} = \sum_{ijk} \langle \phi_i \psi_k, \omega \rangle^* \langle \phi_j \psi_k, \omega \rangle \phi_i^\dagger \otimes \phi_j \tag{3.5}$$

where ϕ_i and ψ_k are bases on S_1 and S_2, respectively, and $\phi_i^\dagger \otimes \phi_j$ is the tensor product of the vectors ϕ_i and ϕ_j, i.e. the operator that maps ϕ_i onto ϕ_j and all vectors orthogonal to ϕ_i onto 0. Von Neumann calls the state $\rho_\omega^{(S_1)}$ the projection of ω onto the factor space H_1, but note that this is very different from the projection onto a subspace (which always produces a pure state.)

The fact that $\rho_\omega^{(S_1)}$ is uniquely defined implies that it is independent of which bases ϕ_i and ψ_k we choose to evaluate (3.5). But note that in the representation of $\rho_\omega^{(S_1)}$ given in (3.5) there are elements $\phi_i^\dagger \otimes \phi_j$ which cannot simply be read as projection operators. Therefore, (3.5) is not in a form that allows us to read $\rho_\omega^{(S_1)}$ as a probability distribution over pure states. As we have seen in the first chapter, there are always such representations for any mixed state, but they are not uniquely defined. Von Neumann has shown that for every $\omega \in H_1 \otimes H_2$ there is a representation of the form

$$\omega = \sum_i c_i \phi_i \otimes \psi_i \tag{3.6}$$

where the ϕ_i and ψ_i are elements of two bases on the two subspaces. If one writes the projected states in these bases, they are

$$\rho_\omega^{(S_1)} = \sum_i |c_i|^2 \phi_i^\dagger \otimes \phi_i \tag{3.7}$$

and the analogue for S_2. These mixtures are diagonal and can therefore be interpreted as a probability distribution over the elements of the bases. Furthermore, we can simply read off the relative states for any element of the bases: ϕ_i is the relative state to ψ_i and vice versa. This representation of an entangled state is called its *biorthogonal decomposition*. Also the biorthogonal decomposition is not always uniquely defined.

In general, any interaction of two systems will create an entangled state of the composite system. The interaction of two systems is described by a unitary transformation on the composite system. This transformation will in general not transform disentangled states into other disentangled states. Therefore, after any interaction the only state we can ascribe to a system is the projected state, which in general will be a mixture.

It is extremely important to realize that mixed states resulting from the projection of an entangled state cannot be given an ignorance interpretation like the mixtures of quantum mechanical states that we discussed in the last chapter. It is simple to see why this is so: if we give the projection mixture (3.7) an ignorance interpretation, what we say is that subsystem S_1 is really in some pure state ϕ_i, but we don't know in which one; it follows that S_2 is in the pure state $\psi_{\mathrm{rel}}(\phi_i)$ But this would mean that the composite system is in the disentangled state $\phi_i \otimes \psi_{\mathrm{rel}}(\phi_i)$, which is different from the pure state ω (3.6) we know it to be in. This leads to a contradiction in the value assignment to operators. Assume that the ϕ_i are eigenstates of a variable A. Hence, on an ignorance interpretation of (3.7), S_1 would be in an eigenstate of A. As we have seen, it follows from the properties of the tensor product that the state of S_1 and S_2, $\phi_i \otimes \psi_{\mathrm{rel}}(\phi_i)$ is an eigenstate of $A \otimes \mathbf{1}$. But ω is, by definition, not an eigenstate of $A \otimes \mathbf{1}$. Hence, we cannot maintain that the two descriptions are empirically equivalent, so the ignorance interpretation is not tenable.

In general, any quantum mechanical system we encounter may have somewhere in its past interacted with some other system and may therefore be in an entangled state with that other system. But then it follows from our reasoning above that the system is in a mixed state, and this mixed state cannot be given an ignorance interpretation. I will call a mixture of this kind an *objective* mixture. It might at first sight seem more plausible in a situation like this to maintain that there really is no state at all that describes the system, and that only the composite system has a well-defined quantum mechanical state. But besides the argument used earlier, that we can perform any measurement on each subsystem and hence must be able to summarize our measurement results in a state description, this line of reasoning would also lead us to the conclusion that no system but the universe as a whole will have a quantum mechanical state properly, because we can say with confidence for any system that at some point in the past it has interacted with another system. Hence, if we did not allow objectively mixed states for quantum mechanical systems, quantum mechanics would become inapplicable to any system in nature.

In the formalism introduced in section 2.4, we treated mixed states as the analogue of classical statistical states. But we did not discuss the interpretation of mixed states in quantum mechanics there. We implicitly assumed that all mixed states can be understood as ignorance mixtures in analogy to the approach generally taken in statistical mechanics. The argument above shows that this approach is not possible for quantum mechanics. We must accept mixed states as equally objective as pure states.

The argument against an ignorance interpretation of projection mixtures was based on the fact that this would lead to a different and experimentally distinguishable state of a composite system. But on the other hand, it is obvious that for any measurement confined to a single subsystem there is no difference in the empirical predictions of Born's rule whether we give the mixed state an ignorance interpretation or whether we treat it as an objective mixture. This follows simply from the fact that the mixed state itself gives us any empirical prediction that we can have. Its interpretation does not matter. Let us call this feature the *local indistinguishability* of objective and ignorance mixtures. While local indistinguishability cannot be an answer to the measurement problem, it is an important indicator of what the measurement problem is about: it is not a problem about certain empirical predictions of quantum mechanics that could be solved by changing the theory, but it is a problem about the interpretation of the theory, i.e. a problem about how to describe the reality that is reflected in these empirical predictions.

3.2 Decoherence

If we confine our attention to a single system and only consider measurements on this system we can therefore say: interactions will transform pure states of the system into mixtures. This process is called *decoherence*. It has only recently gained the interest of theoretical physicists in order to explain the emergence of classical behavior in quantum mechanical systems.[32] The theory of decoherence is important for an interpretation that wants to maintain that quantum mechanics is the fundamental theory of nature and classical mechanics only an approximation for macroscopic systems. Decoherence is supposed to answer the fundamental problem for such an interpretation: why is it that macroscopic systems always seem to behave classically if they are to be understood as large quantum mechanical systems?

The general line of the answer is that this happens because of the interaction of systems with their environment, especially the thermal electromagnetic field (which of course is always present unless at temperatures close to absolute zero). This interaction spontaneously turns superpositions of eigenstates of certain variables (namely, macroscopic

[32] For an introductory account, see Zurek (1991)

classical variables) into mixtures of these variables. Of course, on von Neumann's interpretation of possessed values discussed in the last chapter, this process will not be sufficient to explain the existence of definite values for the classical variables. For these mixtures are not ignorance mixtures, but the local projected states of a superposition in the combined system of the object and its environment. Hence, on von Neumann's account, they do not have a possessed value. We will return to the question of possessed values in our discussion of Everett's interpretation.[33] For now, we can say that the objective mixtures which a process of decoherence produces are locally indistinguishable from ignorance mixtures.

This is not the place to discuss in detail the research done on decoherence which is still a very new field and which has not yet furnished us with a general theory applicable to all macroscopic objects. But it is clear from the discussion so far what the question is that any theory of decoherence has to answer. While any interaction can transform the initial state of a system into a mixture, it needs to be shown that this mixture actually is a mixture of states that have a classical interpretation. This means: the basis that diagonalizes the density matrix is determined only by the interaction (and not the initial states of the systems) and it is approximately a basis of "classical" states, i.e., eigenstates of classical variables like position and momentum. In general, the basis determined by a decoherence interaction that diagonalizes the projected states is called a decoherence basis.

The central interaction to be considered for decoherence is the interaction of a system of charged particles with a thermal electromagnetic field. Unruh and Zurek (1989) have treated this case in a simple calculable model of a harmonic oscillator and a scalar field interacting with this oscillator. For this interaction, they derived an exact equation for the density matrix of the harmonic oscillator and solved this equation numerically for different initial conditions and strengths of interaction. The result they obtained is: there is very fast decoherence, that is, the decoherence is much faster than the relaxation time, i.e., the time of reaching the thermodynamic equilibrium. This means, decoherence happens long before any considerable amount of energy is exchanged between the oscillator and the scalar field. Hence, the interaction with the field turns the state of the harmonic oscillator into a classical state already before it affects the state in any way that is classically accountable. It is

[33] I consider it a strong argument for Everett's interpretation that it is able to accommodate the theory of decoherence, whereas this is not possible in a traditional interpretation. Most recent proponents of decoherence subscribe to some kind of an Everett interpretation, e.g. Unruh and Zurek (1989), Gell-Mann and Hartle (1991), Halliwell (1993).

plausible to assume that the same happens for more complex systems and the full electromagnetic field, even when the complexity of these problems has made it impossible, at least as of yet, to formulate equations of motion or even to solve them.

Another interesting question about decoherence is treated by Amann (1992): the question of delocalization in molecules, that is, the fact that in many molecules atoms do not have well-defined positions. On the other hand, macromolecules do have well-defined shapes, as is well-known from decades of biochemical research into the structure of organic molecules. Amann has shown that whereas in smaller molecules superpositions of position eigenstates often are energetically the most favorable states, such superpositions decohere in larger molecules because of the interaction with the thermal electromagnetic field.

In all cases of interaction with an electromagnetic field, it is important to note that the escaping photons travel at the speed of light and hence cannot be caught up with anymore after the interaction happened. This means: unless a measurement device is already in place before the decoherence happens, there is no experimental method to ever distinguish the local projected mixture of the system that has undergone decoherence from an ignorance mixture, because it is never possible even in principle to show the fact of a superposition of the combined system and electromagnetic field.

The effect of decoherence can be described as the introduction of *superselection* rules. Superselection rules say that certain states of a quantum mechanical system are never transformed into each other by any unitary time transformation matrix that is physically possible for the system. Consider, for example, a composite system consisting of two separate parts that are completely isolated from each other. Here, the Hamiltonian operator is simply the sum of the Hamiltonians for the subsystems:

$$H = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2 \tag{3.8}$$

Its eigenstates are the disentangled states $\phi_i \otimes \psi_j$ where ϕ_i, ψ_j are the eigenstates of H_1 and H_2. From this it follows that H never transforms a disentangled state into an entangled state and, on the other hand, that an entangled state

$$\sum_{ij} c_{ij} \phi_i \otimes \psi_j \tag{3.9}$$

and the corresponding mixture of projected states

$$\sum_{ij} \left|c_{ij}\right|^2 P(\phi_i) \otimes P(\psi_j) \tag{3.10}$$

will evolve in exactly the same way. Therefore, also the evolution of the system will never give us any clue whether the system is an entangled state or in a mixed state.

Proof of the statements above:

If $H = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2$, then the unitary transformation operator U of the combined system is:

$$U = e^{-itH/\hbar} = e^{-itH_1 \otimes \mathbf{1}/\hbar} \otimes e^{-it\mathbf{1} \otimes H_2/\hbar},$$

because H_1 and H_2 commute (they are operators on different Hilbert spaces.) But this is simply $U = U_1 \otimes U_2$, where $U_1 = e^{-itH_1/\hbar}$ is the unitary transformation operator on S1 and analogous $U_2 = e^{-itH_2/\hbar}$. Hence, for any disentangled state $\phi \otimes \psi$, the transformed state is $U(\phi \otimes \psi) = U_1\phi \otimes U_2\psi$, which is a disentangled state, again.

On the other hand, every state of the combined system can be represented in biorthogonal form as $\sum_i c_i \phi_i \otimes \psi_i$, where the ϕ_i and ψ_i are some sets of orthogonal states in H_1 and H_2. The projected states in this representation are $\sum_i |c_i|^2 P(\phi_i)$ and $\sum_i |c_i|^2 P(\psi_i)$. The time-transformed state is

$$U\left(\sum_i c_i \phi_i \otimes \psi_i\right) = \sum_i c_i \left(U_1\phi_i \otimes U_2\psi_i\right),$$

which again is in diagonal form, because the $U_1\phi_i$ and $U_2\psi_i$ are sets of orthogonal states, too (because U_1 and U_2 are unitary operators.) Therefore, the projected states of the transform are

$$\sum_i |c_i|^2 P(U_1\phi_i) = \sum_i |c_i|^2 U_1 P(\phi_i) U_1^{-1},$$

and analogously for H_2. But these are the time-transforms of the projected states, q.e.d.

Generally, if there is any interaction between the subsystems, the superselection rules break down and entangled states will develop differently than the corresponding projected mixtures. But if each subsystem is coupled with its environment in such a way that decoherence results (and moreover, that decoherence happens on a shorter time scale than the interaction between the different subsystems), then entangled states will decohere quickly into mixtures of disentangled states: assume that the interaction results in an entangled state $\sum_{ij} c_{ij} \phi_i \otimes \psi_j$. The interaction with the environment leads to a state $\sum_{ij} c_{ij} \phi_i \otimes \psi_j \otimes \omega_{ij}$ (assuming for simplicity that the interaction doesn't change the ϕ_i and ψ_j.) If all the ω_{ij} are pairwise orthogonal (i.e. if decoherence is complete), this has as a projected state for the systems without the environment the mixture $\sum_{ij} |c_{ij}|^2 P(\phi_i) \otimes P(\psi_j)$, which is a mixture of the disentangled states $\phi_i \otimes \psi_j$. Hence, the system will essentially behave like a classical Markovian stochastic system: the interaction between different subsystems will not create entangled states but mixed states of each subsystem. Whereas, properly speaking, this is a dynamic effect and not the result of a universal superselection rule, it leads to the same phenomenological results: disentangled states always evolve into disentangled states (possibly mixtures), and entangled states behave like the corresponding projected mixtures (they actually evolve into them.) Therefore, this effect of decoherence has been termed, somewhat sloppily, the introduction of superselection rules in decoherent systems.[34]

Superselection rules have first been used to address the measurement problem by Beltrametti and Cassinelli (1981). They did not introduce them as a dynamical effect, though, but simply gave an ad hoc postulate of superselection rules for classical systems. This way of introducing superselection rules causes problems for the account of measurement that dynamical superselection rules don't have (Hughes 1989, 285-87). The application of superselection rules caused by decoherence to the measurement process will be discussed in the next section.

Altogether, we can hope that decoherence is an answer to the question how classical behavior of macroscopic systems can be consistent with quantum mechanics: it gives a mechanism that makes the interference effects of quantum mechanics disappear and gives rise to a classical domain. But two problems still remain:

(1) So far the process has only been analyzed for very simple models. It is a problem for applied physics to treat decoherence for sufficiently realistic models of complex systems. This involves complicated questions of matter-field interactions and it may not be possible to give exact solutions for the more complex cases. But the research in this field is just beginning and with the help of experimental results and numerical methods we can expect further insight into the workings of decoherence. Nevertheless, it is plausible to assume that for normal macroscopic objects that are in constant interaction with particles and fields in their environment, decoherence will be pervasive and local superpositions are highly unstable.

(2) A question of a very different kind is how we are to interpret the projection mixtures of decoherence as reduced states in von Neumann's sense. As we have seen, a straightforward ignorance interpretation of objective mixtures will not do. This means:

---

[34] I am not aware of any research being done on the subject, but it seems to me that a study of complex decoherent systems could give us new insight into macroscopic chaotic systems like the human brain, where the combination of quantum effects and decoherence very well might play a substantial role in the dynamics.

although the local indistinguishability makes us expect that these projection mixtures will, for all practical purposes, look like ignorance mixtures of pure states, we still need the reduction postulate to attribute values to such projection mixtures. The fundamental question now is: can we eliminate this notion of a measurement altogether and give a strictly objective account of whether and how variables can have values in objective mixtures? This is a question about the interpretation of the formalism of quantum mechanics and it will be addressed in the next sections.

3.3 The Interaction Model of Measurement

Let us return to our discussion of the measurement problem. Any measurement involves an interaction between the system that the measurement is performed on, and the measurement apparatus ("apparatus" here can very well mean a human observer). Therefore, it is important to characterize measurement as an interaction if we want to find an answer to the measurement problem. This attempt has a long history: a quantum mechanical model of measurement as an interaction was first introduced by von Neumann (1932, 233-37) and was used by many authors in their attempt to give an explanation of the reduction postulate.[35] It has very much become the standard setting for discussions of the measurement process, and it will play a central role in this thesis. Nevertheless, it raises a lot of questions about its general applicability, especially because it extends the use of the quantum mechanical formalism to macroscopic measuring apparatus. Because of our lack of empirical confirmation for quantum mechanics as a theory of macroscopic systems, this extension is not trivial.

The first argument against such an extension is that quantum mechanics is so fundamentally different from classical mechanics that it gives obviously absurd descriptions of macroscopic systems. Macroscopic systems are never found in superpositions of different states. But rather than seeing this as an argument against the universal applicability of quantum mechanics, one can understand it as a more general form of the question of the measurement problem: why is it that quantum mechanics does not seem applicable to macroscopic systems if it is applicable to their microscopic parts?

The second argument against the interaction model of measurement is that it is too restrictive in its idealization of any actual measurement process, and that it cannot deliver what its proponents claim once these restrictions are given up. In this section, I shall try to

---

[35] E.g., Wigner (1976), Everett (1973), Jauch (1968), Daneri, Loinger, Prosperi (1962), Healey (1989), van Fraassen (1991).

defend the model from this suspicion by showing that it can be generalized to accommodate approximate measurement interactions. For this, I will draw on the discussion of interactions in the last section. First, let us discuss von Neumann's original model:

There are two systems, a measured system S and a measurement apparatus M. The initial state of S is any state $\phi \in H_S$, the initial state of M is some specified "ready" state ψ_o. This state ψ_o need not be unique or a pure state. Already von Neumann observed that the question whether ψ_o is mixed or pure is irrelevant for the discussion of the reduction.[36] For simplicity of notation, I shall assume that ψ_o is a pure state. Von Neumann imposes the following criterion on a physical interaction between S and M to be a measurement of an variable *A* on S:

If the initial state of S is an eigenstate ϕ_i of the measured variable *A*, then the interaction will lead to a state $\phi_i' \otimes \psi_i$, where ψ_i is an eigenstate of some variable B on M, which we will call an indicator state (because ψ_i indicates that S was in ϕ_i). For two orthogonal eigenstates ϕ_i, ϕ_j of *A*, also the resulting eigenstates ψ_i, ψ_j of *B* are orthogonal. The reason for this criterion is that it is necessary to guarantee that by a single observation of M we can say in which eigenstate of the variable *A* the system S was. As we have seen in section 2.4, only orthogonal states are distinguishable by single measurements.

There is an obvious circularity in the criterion. All it says is that, once a measurement interaction has taken place: if we are able to observe eigenstates of *B*, then we are also able to observe eigenstates of *A*. For this reason alone it is clear that the criterion by itself will not suffice as a definition of measurement. One might hope that in the case of a macroscopic measuring apparatus we can circumvent the circularity because it is clear what an observation of *B* means. But this cannot work: for the characterization of the measurement interaction we need a quantum mechanical description of the measuring apparatus, and with this come all the nonclassical states that are possible for a quantum mechanical system, like superpositions or objective mixtures. But once we allow these states, it is not clear any more what "observation" means—this was the problem we sought to explore, after all. Therefore, we need another postulate to tell us when a measurement is complete.

Von Neumann's addresses this problem by observing that every measurement consists of a chain of interactions that eventually has to lead to a human observer's consciousness. It

---

[36] Von Neumann (1932), p. 233 refutes the proposal that the transition from a pure state to an ignorance mixture could be explained by assuming that M was initially in an unknown state represented by an ignorance mixture.

is this endpoint that defines a measurement. He claims that it is the involvement of consciousness that causes the process of reduction (which we will discuss later in this section). We will return to this issue in the next section.

Besides this criterion, von Neumann imposes a second criterion on a measurement interaction: the measured eigenstates are not changed by the process of measurement:

$$\phi_i \otimes \psi_o \rightarrow \phi_i \otimes \psi_i. \tag{3.11}$$

This criterion is too restrictive and highly unrealistic, because many measurements change the measured state. This criterion is only plausible when we consider the reason why von Neumann introduced the model of measurement in the first place: he wanted to account for the possibility of repeatable measurements because repeatability was his main argument for postulating the reduction process (see footnote 20 in section 2.4). I propose to take this criterion as a definition and call a measurement repeatable if it fulfills this criterion.

What follows from the definition of a measurement process for a general state of the measured system? Because of the linearity of any unitary transformation matrix, criteria (1) and (2) suffice to tell us the final state of S. If the initial state of S is expressed on the basis of eigenstates of A as $\sum_i c_i \phi_i$, then the interaction with M will lead to:

$$\left(\sum_i c_i \phi_i\right) \otimes \psi_o \rightarrow \sum_i \left(c_i \phi_i \otimes \psi_i\right). \tag{3.12}$$

The final state is entangled. The projected states on the subsystems are

$$\sum_i |c_i|^2 P(\phi_i) \text{ and } \sum_i |c_i|^2 P(\psi_i) \ . \tag{3.13}$$

Because the ϕ_i and the ψ_i are pairwise orthogonal sets of states, each ϕ_i and ψ_i are relative states of each other. The biorthogonal decomposition can simply be read off from the mixtures: With probability $|c_i|^2$, S is in state ϕ_i and M is in state ψ_i.

That is: the measurement interaction turns a superposition of eigenstates of the measured variable into a mixture of these states, with a probability distribution $|c_i|^2$ that obeys Born's statistical interpretation of quantum mechanical amplitudes. On the other hand, the measuring apparatus is in a mixture of indicator states ψ_i that are exactly correlated to the ϕ_i (and hence have the same probability distribution).

This looks very much like everything we need to explain the measurement process: the interaction turns superpositions into mixtures, and even gives them the right probability distributions. So what is the problem? To be sure, there are two:

1. The mechanism only works if the interaction exactly fulfills criteria (1) and (2). If the measurement is not repeatable or if it is not exact, the biorthogonal decomposition rule

does not even approximately represent the state of S after the measurement as a mixture of eigenstates of *A*.[37] I will argue that this problem can be addressed by considering the decoherence of measurement devices.

2. As we already discussed in section 3.1, the mixtures resulting from an interaction are not the right kind of mixtures: while we need an ignorance mixture as the outcome of a measurement, which is really a pure state (we just don't know which), the interaction model gives us an objective mixture, which is objectively different from any pure state.

In the remainder of this section, I will generalize von Neumann's model to accommodate approximate measurements and return to the second question in section 3.4. Let us begin with a more detailed analysis of the measurement interaction.

Von Neumann assumes for simplicity that the ϕ_i form a complete basis. This is equivalent to the requirement that the measured variable *A* be nondegenerate on H_S. This requirement is not necessary, though. It is obvious that if our model allows us to measure any nondegenerate variable, we can also measure any degenerate variable (there is less information required for the latter). The requirement then is: Be Φ_i the subspace spanned by the eigenvectors of the *i*-th eigenvalue. If the initial state of S is some $\phi_i \in \Phi_i$, then the resulting state ψ_i of M must be an element of a subspace Ψ_i corresponding to the *i*-th eigenvalue of some variable *B* on M. Obviously, the subspaces Ψ_i have to be pairwise orthogonal to assure that measuring *B* will be sufficient to tell which subspace Φ_i the system S was in.

So far, a measurement is treated as exact: the final states or subspaces of the measuring apparatus are perfectly correlated with the initial states of the system. There is no mistake possible. Let us try to generalize the model to accommodate approximate measurements: if we keep the repeatability criterion, the most general form the interaction can take is $\phi_i \otimes \psi_o \rightarrow \phi_i \otimes \psi_i$. Therefore, all that criterion (1) amounts to is the requirement that the ψ_i are pairwise orthogonal. Hence, this is the only requirement that we can relax. And this is indeed what we need: if the ψ_i are not orthogonal, the state of M will not give unambiguous information about the state of S. It is only possible to give probabilistic statements about the correlation between about the value of *B* on M and the value of *A* on S.

---

[37] Albert (1992),194-6, who credits the argument to Yakir Aharonov.

In this way, the Hilbert space formalism can elegantly account for approximate measurements.[38]

If we don't use pure states but subspaces, the account of approximate measurements is analogous: all interactions can be seen as a transformation of a given set of initial subspaces Φ_i on S into a set of final density operators σ_i on M in the following way: if S is initially in some unknown state in the subspace Φ_i, this fact can be modeled by attributing to it the density operator

$$\rho_i = \frac{1}{N_i} P(\Phi_i), \tag{3.14}$$

where $P(\Phi_i)$ is the projection operator onto Φ_i and N_i the dimension of Φ_i. (If Φ_i is of infinite dimension, we have to use unnormalized density operators and can only give conditional probabilities.) This assignment of a density operator to an unknown state is called the von Neumann rule and corresponds to the maximum entropy principle in statistical mechanics. We will discuss it further in chapter 5. Again, $\rho_i \otimes \psi_0$ gets mapped to some density operator Σ_i of the product space $H_S \otimes H_M$ by the measurement interaction. This operator is projected onto a density operator σ_i in H_M.

From that, we can give a general characterization of an approximate measurement: an interaction is a measurement for a set of orthogonal subspaces Φ_i if the final projection mixtures σ_i are sufficiently distinct. To define distinctness of density operators, we can use a scalar product for operators defined by $\langle A,B \rangle = tr(AB)$. If the scalar product is zero, we call the density operators orthogonal, and the measurement is exact.

Notice, however, that in the case of an approximate measurement the final states do not project to a mixture that is diagonal in the original eigenstates ϕ_i, because in the final states $\sum_i (c_i \phi_i \otimes \psi_i)$, the ψ_i are not orthogonal. As Albert has noted (Albert 1992), even approximately orthogonal ψ_i can completely change the decomposition of the mixed state of S. As an example, take a measurement of the variable A having an eigenbasis ϕ_1, ϕ_2 on S. Ideally, an initial state $(\phi_1 + \phi_2)/\sqrt{2}$ should lead to a final state

$$\xi = (\phi_1 \psi_1 + \phi_2 \psi_2)/\sqrt{2}, \tag{3.15}$$

[38] The only alternative to this that I can see is to assume that the interaction is not unique because of external disturbances. But if the external disturbances are modeled quantum mechanically, we will again have a unique interaction, now on the combined system of S, M and environment and the result will be as above: the final projected states of M are not orthogonal.

where ψ_1, ψ_2 are orthogonal states on M. Now think of the measurement interaction introducing a small error Δ, so that the final state is

$$\xi' = \left(\phi_1(\psi_1 + \Delta\psi_2) + \phi_2(\Delta\psi_1 + \psi_2)\right) / \sqrt{2} \tag{3.16}$$

(where I ignore an overall normalization constant). This has as biorthogonal decomposition:

$$\xi' = \left[(1+\Delta)(\phi_1 + \phi_2)(\psi_1 + \psi_2) + (1-\Delta)(\phi_1 - \phi_2)(\psi_1 - \psi_2)\right] / 2\sqrt{2} \tag{3.17}$$

which, for Δ as small as you want to choose it, represents $\rho_{\xi'}^{(S)}$, the projected state of S as a mixture of orthogonal pure states of S in the following way:

$$\rho_{\xi'}^{(S)} = \frac{1+\Delta}{2} P\left[\frac{\phi_1 + \phi_2}{\sqrt{2}}\right] + \frac{1-\Delta}{2} P\left[\frac{\phi_1 - \phi_2}{\sqrt{2}}\right] \tag{3.18}$$

But the states $(\phi_1 + \phi_2)/\sqrt{2}$ and $(\phi_1 - \phi_2)/\sqrt{2}$ are states that are maximally incompatible with the basis we had set out to measure (ϕ_1 and ϕ_2). This means: we cannot say that the measurement interaction turns the state of S into a mixture of states that even approximately are eigenstates of the measured variable. By the same argument, also M will not be in a mixture of eigenstates of B.

What would be a physically more plausible outcome of an approximate measurement? Consider an approximate position measurement:[39] a plausible result from the model should be that the apparatus is in a mixture of (orthogonal) eigenstates ψ'_i of some macroscopic variable B (which can be read off by an observer and tells her the approximate value of A), and that the system is in a mixture of states that are correlated to the ψ'_i, but not necessarily orthogonal or eigenstates of any particular variable. Rather, we expect wave packets of a certain width (depending on the error of the measurement) centered around the approximate value for A. Obviously such wave packets are neither pairwise orthogonal nor are they defined as eigenstates of some variable on S. Can we, instead of the biorthogonal decomposition rule, find some other rule for decomposing the final state that makes more sense physically?

This can be done by assuming that the measuring apparatus is decoherent. In this case, the final state of the measuring apparatus gets quickly correlated with the environment, and therefore the final superposition $\sum_i c_i \phi_i \otimes \psi_i$ of the combined system turns into a mixture. Let us call the decoherence basis of the measuring apparatus a *pointer* basis ψ'_i. Obviously, if M is a decoherent system, then the final states ψ_i must be elements of the

---

[39] Von Neumann has treated this in a simple model (von Neumann, 1932, 236-7). This model is discussed in more detail in Everett (1973), 56-60 and 100-103

pointer basis, otherwise the superselection rules for M will have the result that the information about S is not retrievable by any measurement on the measuring apparatus alone.

In the case of an approximate measurement the ψ_i will only be approximately identical to the pointer basis ψ'_i. That is, we can express the ψ_i as

$$\psi_i = \sum_j b_{ij}\psi'_j \text{ with } b_{ij} \approx \delta_{ij}. \tag{3.19}$$

Then the final state of the measurement interaction is

$$\xi = \sum_i c_i\phi_i \otimes \psi_i = \sum_{ij} c_i b_{ij}\phi_i \otimes \psi'_j = \sum_j c'_j\phi'_j \otimes \psi'_j \tag{3.20}$$

with

$$\phi'_j = N_j \sum_i c_i b_{ij}\phi_i \tag{3.21}$$

where N_j is a normalization constant, and $c'_j = 1/N_j$. In the terminology we introduced in section 3.1, ϕ'_j is the relative state to ψ'_j, given ξ.

Now this final state decoheres because of the interaction of M with its environment, the decoherence basis being ψ'_j. That is, the state develops into a state

$$\xi' = \sum_j c'_j\phi'_j \otimes \psi'_j \otimes \omega_j \tag{3.22}$$

where the ω_i are a set of orthogonal states of the environment. To calculate the projected states after formula (3.4), we use the bases ϕ_i, ψ'_i, and ω_i in the systems S, M, and the environment (remember that it doesn't matter which bases we use to calculate the projected state) and obtain for S:

$$\begin{aligned}
\rho^{(S)}_{\xi'} &= \sum_{ijkl} \left\langle \phi_i\psi'_k\omega_l, \xi' \right\rangle^* \left\langle \phi_j\psi'_k\omega_l, \xi' \right\rangle \phi_i^\dagger \otimes \phi_j \\
&= \sum_{ijklmn} \left\langle \phi_i\psi'_k\omega_l, c'_m\phi'_m\psi'_m\omega_m \right\rangle^* \left\langle \phi_j\psi'_k\omega_l, c'_n\phi'_n\psi'_n\omega_n \right\rangle \phi_i^\dagger \otimes \phi_j \\
&= \sum_{ijk} \left|c'_k\right|^2 \left\langle \phi_i, \phi'_k \right\rangle^* \left\langle \phi_j, \phi'_k \right\rangle \phi_i^\dagger \otimes \phi_j \\
&= \sum_k \left|c'_k\right|^2 \phi'^\dagger_k \otimes \phi'_k = \sum_k \left|c'_k\right|^2 P\left(\phi'_k\right)
\end{aligned} \tag{3.23}$$

where I have left out the tensor product sign ⊗ within scalar products. This result means: unlike using the biorthogonal decomposition rule, using decoherence will give us a projected state of S that depends bilinearly on the states ϕ'_i and therefore approximates a state diagonal in the ϕ_i as $b_{ij} \to \delta_{ij}$. This is the result we found physically plausible in our considerations above: whereas the biorthogonal decomposition gave us as the final state a mixture that was diagonal in some basis that depended on the initial state and could basically be any basis on S, decoherence gives us a state that is a mixture of the relative states ϕ'_i to the pointer basis ψ'_i.

This means: if we, for example, have a position measurement with an error Δx, then the resulting state of S will be a state of width Δx around the position that the pointer state indicates.

The final state of M is:

$$\begin{aligned}
\rho_{\xi'}^{(\mathrm{M})} &= \sum_{ijkl} \langle \phi_k \psi'_i \omega_l, \xi' \rangle^* \langle \phi_k \psi'_j \omega_l, \xi' \rangle \psi_i'^\dagger \otimes \psi'_j \\
&= \sum_{ijklmn} \langle \phi_k \psi'_i \omega_l, c'_m \phi'_m \psi'_m \omega_m \rangle^* \langle \phi_k \psi'_j \omega_l, c'_n \phi'_n \psi'_n \omega_n \rangle \psi_i'^\dagger \otimes \psi'_j \\
&= \sum_{ik} |c'_i|^2 \langle \phi_k, \phi'_i \rangle^* \langle \phi_k, \phi'_i \rangle \psi_i'^\dagger \otimes \psi'_i \\
&= \sum_{ik} |c_k b_{ki}|^2 \psi_i'^\dagger \otimes \psi'_i = \sum_k |c_k b_{ki}|^2 P(\psi'_i)
\end{aligned} \tag{3.24}$$

That is: $\rho_{\xi'}^{(\mathrm{M})}$ is *always* diagonal in the pointer basis ψ'_i, and the probabilities approximate the correct probabilities as $b_{ki} \to \delta_{ki}$. Again, this is exactly the desired result. An analogous treatment can be given to an approximate measurement of degenerate operators.

Of course, the question still remains whether measuring apparatus really are decoherent. Because, as I have said, we don't have a universally applicable theory of decoherence yet, this claim remains an hypothesis. But in the case of measuring apparatus, there is an argument why this should be the case: remember that it is sufficient for the pointer basis to be decoherent that the states of the pointer basis will interact in such a way with the environment that the resulting states of the environment are orthogonal without changing the pointer states. But this means simply that the interaction of the environment with the measuring apparatus is again a measurement interaction. Now remember von Neumann's observation about the chain of measurement interactions: the measurement apparatus itself is never the endpoint of a measurement, it must at least allow the possibility of an observation, i.e., the different pointer states must interact differently with the environment so that, if an observer is present, the environment can transmit the information about the pointer states to the observer.

In the most common case, the transmission from the measurement apparatus to the observer will be optical, that is, the measurement apparatus must look differently in its different pointer states, which is to say it must emit or reflect orthogonal states of the electromagnetic field. Consider, for example, a photographic plate: if different spots on the plate are blackened, this results in different patterns of absorption of incident light. In the case of a measuring apparatus with a material pointer, this pointer will interact differently with the electromagnetic field depending on which position it is in. In all these cases, it is sufficient that the state of the environment (electromagnetic field) after the interaction is different for different states of measuring apparatus. And of course it is the purpose of a measuring apparatus that its different states are simply discernible (generally just by looking at it). But this means that different apparatus states lead to clearly different states of the electromagnetic field around the apparatus. Otherwise we would not be able to *see* any difference between the apparatus states. Of course, the measurement apparatus might be read by other means than by looking at it (printout, electrical signals, or sound, for example). But then, whatever the states of the environment are that carry the information about the state of the measuring apparatus, those states have to be orthogonal.

Note that in case of approximate measurements, which of course may happen at any stage of the chain of measurement interactions, the argument above ensures that as long as the final states of the last measuring system in the chain are decoherent, the final states of the measured system will be the relative states of these and therefore approximate the eigenstates of an exact measurement. The question of the endpoint of the chain is of course still open.

Let us sum up what an analysis of von Neumann's model of the measurement interaction can tell us. First of all, the disappearance of coherent states of the measured system can be explained from the interaction itself. Any measurement will transform superpositions of eigenstates of the measured variable on the system into mixtures. If, furthermore, the measuring apparatus itself is decoherent, the combined system (object plus apparatus) will decohere into a mixture of states of the pointer basis for the measuring apparatus and the corresponding relative states for the system. In this case, only the observed system plus the measuring apparatus plus the environment will be in a superposition.

It follows from this discussion that the transition from an initial superposition to a final mixture of eigenstates in the process of measurement can be explained, not by invoking an ad hoc biorthogonal decomposition, but from the decoherence of the measurement apparatus. This makes it possible to say why just the eigenstates of the measured variable decohere: this is exactly the purpose for which measurement apparatus are built. A measuring apparatus only serves its purpose if its decoherent states get correlated with the eigenstates that it is supposed to measure, and therefore it serves to decohere the eigenstates of the measured observable (at least approximately.) But as we found in section 3.1: the resulting mixture is not an ignorance mixture and hence cannot be interpreted as a probability distribution over reduced states. It is only at this point that reduction becomes necessary. Without it, as von Neumann argues, we cannot explain the repeatability of measurements:

If all the measurement interaction does is producing an objective mixture of eigenstates then nothing tells us that repeated measurements will result in the same measurement outcome. We would rather expect that there is a probabilistic distribution of all possible outcomes. Is there anything in quantum mechanics that tells us otherwise? This problem often has not been noticed by proponents of an interaction model of reduction.[40] They implicitly assume that the resulting mixture can be understood as an ignorance mixture, that is, that the system is really in one eigenstate and we just don't know in which one. Once we know this from observation, of course, the system will be in this eigenstate whenever we perform a measurement again. But why should we expect the same behavior in the case of an objective mixture? Here we know that the system really is not in any pure state.

3.4 The Reduction Postulate

Von Neumann bridges the difference between the objective mixture attainable by decoherence and the ignorance mixture demanded for the repeatability of measurement by postulating that there is a separate process of reduction to pure states that happens upon measurement. His reason for this postulate is that a physical process governed by a unitary transformation matrix can never produce a reduction to pure states (or, we might add, an ignorance mixture of these) (von Neumann 1932, 202-6). Hence, there must be some fundamentally different process that explains that measurements give unique values for variables even if the initial state of the system did not have such a value.

While von Neumann does not explicitly distinguish between objective and ignorance mixtures, we can rephrase his treatment in the following way: the interaction between a system and a measurement apparatus leads to an objective mixture of eigenstates of the measured variable, the separate and parallel process of reduction turns this objective mixture into an ignorance mixture, i.e., some unpredictable pure state.

That is: if the initial state of the object system is $\sum_i c_i \phi_i$, where the ϕ_i are eigenstates of the measured variable A, the state of the object system after the measurement is some pure state ϕ_i, but we don't know which. All we know is that the probability for the outcome ϕ_i is $|c_i|^2$. We can represent this fact by attributing to the system the state $\sum_i |c_i|^2 P(\phi_i)$, where this state has to be understood as an ignorance mixture. As we have seen in the section 2.4, this

40 See, e.g., Nancy Cartwright's criticism of the model of Daneri, Loinger, and Prosperi (Cartwright, 1983, p.169-171), and Hughes's criticism of Jauch (Hughes, 1989, p.283)

state cannot be obtained through a unitary transformation from the initial state of the system, and we found in this chapter that also the interaction with some other system cannot produce this state. Therefore we have to postulate that there is some fundamentally different process in nature which happens when a measurement is performed and which, unlike all other physical processes, cannot be described by a unitary transformation. This process is called the process of reduction.

Von Neumann justifies the existence of a process of reduction that cannot be explained within quantum mechanics with the necessary distinction between subject and object in our description of nature. (Von Neumann, ibid., 222-23) Consciousness is fundamentally different from any phenomenon that can be modeled in a physical theory. Any theory of nature presupposes a conscious subject that formulates and entertains this theory. Therefore, the conscious subject always stands outside of the model.

Hence, it is possible that a process involving a conscious subject has a fundamentally different description than any other physical process. Von Neumann's position here is subjectivist: saying that the process of conscious observation is fundamentally different from all other physical processes is equivalent to saying that the conscious subject stands outside of the physical world. If we were to take human observers as physical systems like others, this distinction would lose all plausibility. But von Neumann does not want to refute the physicality of the human mind completely. Rather, he invokes what he calls the "principle of psycho-physical parallelism." Usually, this is taken as the thesis that mind and physical world are causally independent as in Leibniz's theory of preestablished harmony. But von Neumann's claim is rather epistemological than metaphysical: He says that it must be possible to describe the process of subjective apperception—which, in truth, is outside of the model—*as if* it was happening within the physical world. This statement is yet another indication of the subjectivist position underlying von Neumann's argument: The physical world is not the objective substratum of every phenomenon (including conscious observers), but the construct of our consciousness. Therefore, consciousness itself is outside of the physical world.

If apperception is a process that is fundamentally different from ordinary physical processes, then we are faced with the following dilemma: if we are to be able to give a physical description to the process of apperception, then how can we postulate that a fundamentally different process of reduction is consistent with a general description of physical systems given by quantum mechanics, which entails that all processes are unitary? If on the other hand, measurement involves something that is outside of physics as described by quantum mechanics, then how should we be able to give a description of this fact within quantum mechanics?

Von Neumann's answer to this dilemma is to give the process of reduction a strange ambiguity. It is so to speak not fully physical, but as far as its physical effects are concerned, it can be shown to be consistent with unitary evolution in a certain sense. The "not-quite-physicality" is explained by the fact that it only happens in processes involving consciousness, which is beyond the reach of any theory of nature. About the issue of consistency, von Neumann remarks that measurement doesn't just involve the interaction between an object system and a measuring apparatus. It always involves a whole chain of such interactions eventually leading to a conscious observer.

Von Neumann cites the example of a measurement of temperature (von Neumann 1932, 223): one could say that a thermometer measures the temperature of its environment. But one could also describe the interaction of the mercury with its environment and its expansion by a suitable physical model and say that it is the length of the mercury column that is measured. Further, one could physically describe the interaction of the mercury with the electromagnetic field and say that its state is measured. Or, one could give a description of the refraction of the light and absorption in specific parts of the retina and claim that it is the position of the absorption that is measured. One could go on to describe the excitation of cells of the optical nerve and certain areas of the brain and say it is this that the observer measures. But, as he finishes: ". . . however far we calculate . . . at one point we will have to say: and this is perceived by the observer." This means, however far we extend our physical model, we cannot get rid of a conscious subject perceiving a measurement outcome at the end of any measurement process. This is the endpoint in the chain of interactions that actually defines a measurement and without which our definition from the last section would be incomplete. But at the level of the conscious subject are we forced to require that variables actually have single values. It is not possible to give a meaning to a conscious observer being in a superposition of conscious states. But we are free to move the border between what we consider the observed object and the observer: at any point in the chain we can say that this is the point where the observation happens, and therefore apply the reduction postulate. Hence, if we want to avoid contradictions in the model, we have to show that it doesn't matter for the predicted measurement outcome at what point we choose to apply the reduction postulate.

The important argument for this proof of consistency is that because of the linearity of the equations of motion it does not matter at which point the process of reduction takes place. Every measurement in the chain only is an interaction with the measurement apparatus before, not with the complete composite system. Therefore, only the projected mixture plays a dynamical role and not the overall superposition. Hence, all the reduction postulate amounts to is the reinterpretation of the projection mixture as an ignorance mixture, and it is obviously irrelevant at which stage this is done: the equations will look exactly the same regardless. Von Neumann concludes from this that while it is necessary to postulate a distinct process of reduction, it is not necessary to worry about when exactly in the chain of interactions between the object system and human consciousness it happens. It is sufficient to know that somewhere on the way from the system to the human consciousness a reduction takes place. It follows from this argument that the only way to test the reduction postulate empirically would be to perform a measurement on the total state of the composite system: Only there could we see a difference between the superposition and the ignorance mixture. But as we have seen in our discussion about decoherence, such a measurement would be extremely difficult to perform if any decoherent system is involved.

Von Neumann's subjectivism is of a different kind than Bohr's (Bohr 1935): Bohr takes the whole categorical system of classical physics as given and thinks that quantum mechanics can only be formulated within these categories. First, we have to describe an experimental setup in terms of classical physics and only then can we predict the behavior of a quantum mechanical system in this setup.

Von Neumann, on the other hand, has no problem admitting that quantum mechanics is a fundamentally new and, in principle, self-sufficient way of describing nature. Therefore, also measurement apparatus and even human observers must be describable by quantum mechanics. But any physical theory does not give us nature as it is but only a model that the subject constructs. Therefore, his subjectivism is not the same as Bohr's operationalism but more of the kind of Kant's transcendental idealism. He acknowledges the demand for a universally applicable physical law (here for the case of the physicality of human observers), but not in the form of a direct postulate of physicality. Instead, he formulates it as a methodological principle that we should be able to describe processes of human perception as if they were physical. This is reminiscent of Kant's regulative principles or maxims,[41] which are not to be understood as statements about how the world is, but what form our inquiry into nature should take. Kant writes (Kant 1929, B694):

---

41 I take it that Kant uses "maxims" and "regulative principles of reason" as synonyms, at least in the Critique of Pure Reason. For a more detailed analysis of his use of regulative principles, see (Lehner 1986).

> I entitle all subjective principles, which are derived, not from the constitution of an object but from the interest of reason in respect of a certain possible perfection of the knowledge of the object, *maxims* of reason. There are therefore maxims of speculative[42] reason, which rest entirely on its speculative interest, although they may seem to be objective principles.
>
> When merely regulative principles are treated as constitutive, and are therefore employed as objective principles, they may come into conflict with one another. But when they are treated merely as *maxims*, there is no real conflict, but merely those differences in the interest of reason that give rise to differing modes of thought. In actual fact, reason has only one single interest, and the conflict of its maxims is only a difference in, and a mutual limitation of, the methods whereby this interest endeavors to obtain satisfaction.

Let us try to adapt Kant's ideas to our situation: our interest in a systematic and coherent theory of nature compels us to extend the application of a mechanical model as far as ever possible. From this "interest of reason" results a postulate of physicalism. But we have no reason to take this postulate as an objective law: all it tells us is to try to find a mechanical explanation for any phenomenon we encounter. It does not make any statement about matters of fact independently of our modeling. This is the content of von Neumann's postulate of psycho-physical parallelism, and like Kant, he stresses that it should be taken as a statement about our making of models.

Conflicting with this postulate of physicalism in von Neumann's account is the postulate that once we have performed a measurement, the system (and the measuring apparatus and any further element in the chain of measurement) is in a pure state. Von Neumann justifies this claim by theory-immanent reasoning: there is no other quantum mechanical state that we could ascribe to the system which would give us the right empirical predictions. He obviously sees this as a sufficient reason to postulate the reality of the reduced state. But notice that this argument goes from empirical knowledge (the value of a measurement outcome) to some statement within the model (i.e., a state description). Hence, it is not a logical deduction but rather a form of inference to the best explanation. Or, to put it in another way: the theory itself does not tell us how to relate its elements to our perceptions.

---

[42] i.e. theoretical

Kant invokes a different faculty of reason for the making of such connections between theory and empirical knowledge: this is the faculty of judgment. He claims that in all such instances of judgment we use a regulative principle: nature is suitable for the formulation of systematic empirical laws (Kant 1952, B XXV - XXXIX). This principle again is only justified by an interest of reason, namely the abstraction of theories from specific phenomena. Therefore, postulating a pure state as the outcome of a measurement process should also not be seen as a statement about a matter of fact independent of the given theoretical framework, but merely as a methodological rule to ensure the applicability of quantum mechanics to our empirical knowledge.

There is a further parallel: von Neumann's solution to the conflict between unitary evolution and reduction resembles very much Kant's treatment of the antinomies: there are two conflicting methodological principles that give the appearance of an objective contradiction. But this is an illusion resulting from taking methodological rules as laws of nature existing objectively and independently of the perceiving subject. If one was to be a realist about quantum mechanics, the coexistence of the two conflicting processes would clearly pose a problem. But von Neumann claims that on a subjectivist account it is sufficient to show that within the theoretical account of quantum mechanics the application of the two different postulates is consistent. This consistency proof is von Neumann's proof of the irrelevance of the moment of reduction discussed above.

It is not surprising that many people were not convinced of von Neumann's defense of the consistency of the two processes. It is undoubtedly true that for all practical purposes the coexistence of the two processes does not lead to any inconsistencies in the application of quantum mechanics. But even if we accept von Neumann's subjectivist standpoint, his solution of the measurement problem is not very satisfying. First of all, if we are to take his notion of psycho-physical parallelism serious, it should be possible in principle to model a conscious observer in quantum mechanics. This would be all the more important because, as we remarked earlier, the definition of a measurement in von Neumann's model is incomplete: it only makes the conditional statement that if a quantity is measurable on the apparatus, then a correlated quantity is measurable on the object. If we don't give a quantum mechanical account of what constitutes the endpoint of such a chain of interactions, the definition is spurious. Von Neumann does not say anything about how this endpoint is to be expressed in quantum mechanics. This leads to a fundamental ambiguity in the status of the reduction process: Are we to think of reduction as the subjective appearance of a quantum mechanical superposition and hence not as a physical process, but rather a redescription of the same fact;

or are we to take reduction as an objective physical process on the same footing as unitary evolution which is caused by a conscious act of observation? As I have argued in the introduction, this distinction makes sense independently of whether we ultimately hold a subjectivist or objectivist view about the epistemology of scientific theories. In principle, this question is decidable by experiment, as we have seen before, because it is equivalent to the question whether the state of the composite system is a superposition or a mixture.

Notice that also on the Kantian account of von Neumann's position, his proposed solution is not admissible. Compare the situation to Kant's solution of the third antinomy: Kant accepts that the principle of causality cannot exclude the existence of transcendental freedom (that is, absolute spontaneity of an act). Nevertheless, he insists, such transcendental freedom necessarily lies outside of the reach of all possible experience. The fact that universal causality is postulated by a regulative principle does not mean we can ever exempt a phenomenon in our experience from it. More general: regulative principles are not limited in their scope, so that there could be a phenomenon that the principle does not apply to, but in the status of their claim: they do not give us a categorical description of an object, but they tell us how we are to find a representation of the object within the categorical framework (Kant 1929, B537-38).

Von Neumann, though, attempts the former use: he tries to limit the applicability of the dynamical law of unitary evolution within the range of possible experience, and wants to express the transcendental nature of the observing subject by a different phenomenal law of causality. This must lead to contradictions within the theory. These contradictions appear in von Neumann's conflicting characterizations of the reduction process: his consistency argument seems to be an argument for the subjective nature of reduction, because it assumes that there should be no objective matter of fact when the reduction takes place. On the other hand, his argument from the repeatability of measurement for a process of reduction is an argument for the objectivity of this process: because an ignorance interpretation of objective mixtures is impossible, there must be an objective process of reduction.

A subjective account of reduction seems quite implausible: how could an observer be in an objective mixture of states (i.e. there is no matter of fact about which measurement outcome she observed) and still have a definite belief about having observed one outcome? This problem led E. Wigner to assume that the dynamical evolution of conscious observers must be fundamentally different from the unitary evolution of other systems (Wigner 1961). This enables them to cause a physical process of reduction. But not only is there no empirical evidence for such a difference, it even seems hard to imagine what kind of a nonunitary process could be at work here and how this (quantum mechanical) process would manage to reliably produce pure classical states.

J. A. Wheeler has discussed a fundamental problem about an objective interpretation of the reduction postulate (Wheeler 1957 and 1979-81): how are we to apply the reduction postulate to cosmology, where we don't have a defined experimental setting and an observer who is external to this setting? Should we assume that without the presence of a conscious observer no reduction of the state has ever taken place? Then we will have to assume that the state of the universe evolved into more and more complicated superpositions for billions of years until the first conscious observer in the history of nature appeared. Even worse, the appearance of a conscious observer itself cannot be supposed to happen simultaneously for all the components of the state of the universe, which presumably is an immensely complex superposition of eigenstates for any macroscopic variable. But then, it becomes fully unclear when the reduction would happen. This cosmological argument turns the ambiguity of von Neumann's postulate of psycho-physical parallelism into a serious paradox: if a conscious observer is to be the agent of reduction as von Neumann postulates, then it has to be outside of the physical universe. But if it is supposed to be a physical system (or at least can be described as such), then it must be part of the physical universe.

The gravity of the dilemma becomes fully visible if we consider measurement on entangled states. This was first pointed out by Einstein, Podolsky, and Rosen (1935) in their famous thought experiment: Consider a composite system in an entangled state, and let the two subsystems separate to some large distance. If we now perform a measurement on one of the subsystems, this measurement will not only reduce the state of the measured subsystem, but also the state of the distant partner.

This fact alone puts adherents of an objective process of reduction into an uncomfortable position: they have to assume that the process acts nonlocally, i.e., directly from one system to the other without any intermittent medium. It seems much more plausible (and that was Einstein, Podolsky, and Rosen's intention) to assume that we simply learned about a fact about the second subsystem that objectively existed already before our measurement (although it was not described by the quantum mechanical state). But there is little reason for a subjectivist to feel confirmed either: we can, after all, measure any variable on the first subsystem, and from this it follows with a suitable entangled state that we can measure any variable on the second subsystem. This is especially clear in David Bohm's (1957) version of the thought experiment. Here, the entangled state is a singlet (spin zero) state, and we measure different components of spin on each subsystem. Because the total spin of the composite system is zero, if we measure the spin in any direction on one subsystem, we know that in the other system the spin in that direction has the opposite value.

A subjectivist now would have to assume that there are objective facts about any variable that is measurable in this way (in Bohm's example values for spin in any direction). John Bell (1964) was able to derive from this assumption an inequality of measurement outcome correlations that is in conflict with the values predicted by quantum mechanics. But this means that there cannot be any local hidden variable theory that could explain away the nonlocality of the reduction process. The result can be generalized to the assumption that the measurement outcomes depend only statistically from the hidden variables (Wigner 1970). The inequalities have been tested experimentally, and the quantum mechanical predictions have been vindicated.[43]

The Bell-Wigner inequalities have given rise to an enormous body of literature, because they seem to make metaphysical assumptions empirically testable without the need of a prior theoretical model. An important result was found by Suppes and Zanotti (1976) and by Jarrett (1984): The assumptions of the Wigner-Bell inequalities can be analyzed in two postulates:

- The probabilities of measurement outcomes on one side are independent of what measurement was performed on the other side. (Jarrett calls this assumption 'locality'. It is not equivalent to the concept of locality I introduced in 2.2, although we would expect it to follow from locality as defined there, if no interaction through fields is present.)
- The probabilities of measurement outcomes on both sides, given a complete description of the state, are statistically uncorrelated (completeness).

The quantum mechanical predictions fulfill the first postulate. Therefore it seems plausible to require that also the theoretical model of the measurement process is local. This conclusion has been disputed (Jones, 1991) because of the fact that locality is a notion that is only defined in field theory (see Section 2.2). Nothing in point mechanics a priori excludes the possibility of action at a distance. Nevertheless, locality seems a well-confirmed empirical principle and it plays a central role in experimental practice: One fundamental assumption of experimental physics is that you can shield systems from their environment. This would be impossible if there are non-local interactions. Furthermore, there is nothing in the derivation of the Bell-Wigner inequality that could not be applied to a modeling of the process in a

[43]Best known is the experiment by Aspect (1976) which also proved that if there is a reduction, it has to propagate faster than the speed of light.

relativistic quantum field theory. The nonlocality, after all, does not stem from the non-relativistic character of the Schrödinger equation, but from the reduction postulate, which is logically independent from any assumption about the dynamics of the theory and supposed to apply to any measurement on a superposition.

Because of the no hidden variable theorems, there is no way to get rid of some form of nonlocal process if we assume the reduction postulate. It has often been pointed out that the nonlocality of the process cannot be used to send signals, so that the violation of relativistic locality is 'hidden' and one can assume that reduction is at least at the phenomenal level not inconsistent with theories that assume locality, such as the theory of relativity (Shimony, 1980). But this is of little comfort if one tries to give an interpretation of quantum mechanics in the sense discussed in the introduction: then we don't look merely for empirical consistency, but for a joint model of both theories.

To sum up: we have two weighty reasons to look for a replacement of the reduction postulate: its conflict with desiderata of scientific methodology and its conflict with the assumption of locality. Everett's formulation promises just that: it is a theory of quantum mechanics without an objective process of reduction, and it does not prohibit us from treating conscious observers as quantum mechanical systems.

Chapter 4

The Interpretation of Everett's Formulation

4.1 Everett's Relative State Formulation of Quantum Mechanics

The fundamental idea of Everett's model is that the repeatability of measurement can be accounted for within quantum mechanics.[44] It is only for the interpretation of quantum mechanical states in terms of observations (i.e., the Born rule) that we need to introduce the consciousness of the observer. The starting point of Everett's treatment of the question of repeatability is that we need some recording device that keeps track of an earlier measurement outcome if we want to be able to decide whether this outcome agrees with a later measurement or not. Everett calls such a recording device an observer. Note that despite this anthropomorphic term all we need at this point is a system that allows for some kind of permanent trace of a measurement outcome (e.g., a photographic plate, a written recording, or a computer memory). I will call such systems memories.

A quantum mechanical model for a memory is not difficult to construct. A memory should be a system whose states can permanently represent propositions about other systems. As we have found in Chapter 2, mechanical propositions contain four components that specify:

(1) a system

(2) a variable (or property) on that system

(3) the value of the variable

(4) the time

We will call such a specification a memory entry, and a (finite) set of entries a memory list. Everett represents an entry by a symbol $\omega_i^{(j)}$, where j designates the measured system, the choice of Greek letter ω the measured variable, and ω_i the i-th eigenvalue of this variable. In Everett's model, the specification of time is not given explicitly, but by a linear ordering of the entries. As long as we are only concerned with the temporal order of observations and not their exact time, this is sufficient. Otherwise, we could easily add a fourth parameter to each entry specifying the time of the observation. A memory list therefore will be written [A, B, ... , N], where each letter stands for a memory entry and the order of the letters represents the temporal order of the entries.

[44] Throughout this section, I will rely on (Everett 1973)

We can now define a quantum mechanical memory: it is a quantum mechanical system which has states in its state space that represent memory lists.[45] Such a memory state representing the list [A, B, ... , N] is written as ω[A, B, ... , N]. As in the case of a macroscopic measurement (section 3.3), it is more appropriate to say that a memory sequence is represented by a subspace of states of the memory, because generally a memory will be a macroscopic system and a lot of different quantum mechanical states will correspond to one memory sequence. But again this generalization doesn't substantially change the argument, so I will generally use single states for simplicity of notation.

As in von Neumann's model of measurement, we require that two memory states that represent different lists be orthogonal, for the same reason as there: it guarantees the distinguishability of different memory states. Let us now turn to the dynamics of a measurement process for a memory, which is a simple analogue to von Neumann's model: the memory states must be stable over time, i.e., they must be stationary states of the system. The only change they can undergo is the acquisition of new information through a measurement. This must not disturb the information already contained in the memory. Hence, the measurement interaction of a memory with an object that is in the eigenstate φ_i of a measured variable A with eigenvalue α_i will look like

$$\varphi_i \otimes \omega[\ldots] \rightarrow \varphi_i \otimes \omega[\ldots \alpha_i], \qquad (4.1)$$

where $[\ldots]$ stands for any memory list that was recorded before the measurement, and $[\ldots \alpha_i]$ stands for the same list with the entry α_i added at the end.

Here, we have assumed that the measurement interaction is exact (for simplicity), repeatable (because that is the case we are interested in), and only involves the object and the memory, without an intermediate measuring apparatus (which could easily be written in, but is irrelevant for our purposes). In the case of a measurement on a superposition of eigenstates, linearity dictates, as we have seen, that the interaction will lead the initial state

$$\Psi_0 = \sum_i c_i \varphi_i \otimes \omega[\ldots] \qquad (4.2)$$

into the final state

$$\Psi_1 = \sum_i c_i \varphi_i \otimes \omega[\ldots \alpha_i] \qquad (4.3)$$

[45]The meaning of the memory states of course is not an intrinsic physical property of these states, but given by the form of the interaction given in equation 4.1.

To represent a second measurement, all we have to do is replace the memory list $[\ldots]$ by $[\ldots \alpha_i]$ in (4.2), and we obtain for the state after the second measurement:

$$\Psi_2 = \sum_i c_i \varphi_i \otimes \omega[\ldots \alpha_i, \alpha_i] \tag{4.4}$$

While this state is still a superposition, the state of the memory now is an objective mixture of memory states in which the earlier and the later result agree. According to the Born rule, only records in which both measurement outcomes agree have a nonzero probability. This argument has a long history, it actually predates Everett's theory: its original form was Mott's argument for the emergence of particle trajectories in quantum mechanics.[46] Hence: while there still is no objective matter of fact about the measurement outcome, it is an objective matter of fact that there are no outcomes which represent different values for A in the two measurements. To make this claim somewhat more precise, let us think of a second measurement apparatus that measures on the memory the property 'the last two entries agree'. That is, all $\omega[\ldots \alpha_i, \alpha_i]$ are eigenstates of eigenvalue 1, all other $\omega[\ldots \alpha_i, \alpha_j]$ with $\alpha_i \neq \alpha_j$ eigenstates of eigenvalue 0. Then, Ψ_2 is an eigenstate of this measurement, i.e. it also has the property 'the last two entries agree' although it is not an eigenstate for any single outcome $\omega[\ldots \alpha_i, \alpha_i]$. We will discuss this type of argument in more detail in the next chapter: it is of fundamental importance for the interpretation of Everett's model.

This argument throws a new light on von Neumann's reason for postulating the reduction process. Maybe the postulate is not necessary to ensure the repeatability of measurements. But before we accept Everett's argument, there is still a major problem to be solved: How are we to understand this claim: that the state Ψ_2 shows the measurement to be repeatable, although there is no matter of fact about a value for the observable? It is at this point that Everett invokes the consciousness of observers. He says (Everett 1973, 63):

> ... we do not do justice to the theory of pure wave mechanics until we have investigated what the theory itself says about the appearance of phenomena to observers, rather than hastily concluding that the theory must be incorrect because the actual states of systems as given by the theory seem to contradict our observations.

---

[46]Nevill F. Mott, "The Wave Mechanics of α-Ray Tracks," Proceedings of the Royal Society A126 (1929): 79–84

That is: assume the memory in question is a conscious observer. Instead of trying to find a physical process that gives A a value on a state like Ψ_1, we could explain the observation of a value as a subjective appearance to an observer in these states. It is obvious that any physical theory of a conscious observer has to distinguish between the physical state of this observer and the phenomenal quality that her state has for the observer herself. This is the fundamental idea of Everett's interpretation: that the difference between a superposition of observer states given by the unitary evolution and the observation of a single outcome has to be explained as the difference between an objective physical fact about an observer and its appearance to the observer.

Everett proposes the following interpretation: we know that in a state Ψ_1 we cannot say that the variable A has one value. But instead of saying that it has no value, let us assume that it has all eigenvalues given by nonzero components of Ψ_1 simultaneously. The state Ψ_1 describes not a single fact about the system, but a coexistent ensemble of facts, each element of the ensemble being represented by a component of $\sum_i c_i \varphi_i \otimes \omega[\ldots \alpha_i]$. If the memory in Ψ_1 is instantiated by a conscious observer, each component $\omega[\ldots \alpha_i]$ describes the observer as having perceived a single measurement outcome. That is: for each component $\omega[\ldots \alpha_i]$ the observer, as described by that component, is not conscious of any other component but $\omega[\ldots \alpha_i]$.

To be precise, Everett does not say quite that much. The better part of his explanatory effort is contained in one footnote:

> At this point we encounter a language difficulty. Whereas before the observation we had a single observer state afterwards there were a number of different states for the observer, all occurring in a superposition. Each of these separate states is a state for an observer, so that we can speak of the different observers described by the different states. On the other hand, the same physical system is involved, and from this point of view it is the same observer, which is in different states for different elements of the superposition (i.e., has had different experiences in the separate elements of the superposition) (Everett 1973, 68).

This is about everything that Everett gives us as an interpretation of his understanding of a superposition. And clearly, it is not quite enough. There are at least three problems for an interpretation of Everett's statement:

(1) It is quite obvious that we are not simply dealing with a 'language difficulty' here: one should think that there is a clear answer to whether we are dealing with a single observer or many of them. If it is one, how possibly could she be in many different states simultaneously, and if there are many, does that not mean that we have introduced a new physical process—call it observer fission—that makes von Neumann's reduction process look perfectly harmless? This means: we are either left with a serious metaphysical or a serious physical problem.[47]

(2) In whichever way we answer question (1), why would we as observers never be conscious of the process of splitting or the multiplicity of states? This is an epistemic problem about Everett's model. It has attracted less attention than problem (1), presumably because it seems hard to imagine how it could be addressed as long as problem (1) is not settled. The first to seriously attempt to address this question were Albert and Loewer (1988) with a thought experiment that we will discuss in section 5.1.

(3) Even if the observer herself is not conscious of the superposition, there could be interference between different components of the superposition. Everett seems to assume that there are no possible interference effects between different observer states, but he does not give an argument for this assumption.

Before we turn to these problems, though, I will review how Everett argues that his model can substitute von Neumann's reduction process. First of all, it is to be applied to the case of several observers: we should demand that if Everett's model is to represent an apparent reduction, different observers measuring the same variable should agree on the outcome of the measurement. If in the situation of equation (4.3) a second observer (with memory states $\omega'[\ldots]$) observes the object, the total state after the second observation is

$$\sum_i c_i \left(\varphi_i \otimes \omega[\ldots\alpha_i] \otimes \omega'[\ldots\alpha_i] \right) \tag{4.5}$$

If then the first observer measures what eigenvalue the second has found (if the second observer is honest, she can simply ask her), and the result "observer two has found eigenvalue α_i" will be recorded as a value α_i' in the first observer's memory, the state is

$$\sum_i c_i \left(\varphi_i \otimes \omega[\ldots\alpha_i, \alpha_i'] \otimes \omega'[\ldots\alpha_i] \right) \tag{4.6}$$

This means: for every possible component of the first observer's state, her observation of α_i and the second observer's report about *her* observation (α_i') agree. Again, the agreement is

[47] This problem has been widely discussed in the literature. See (Healey 1984) for a detailed treatment.

not an objective matter of fact, but only established 'within' the components of the superposition.

This argument together with the argument for repeatability seems sufficient to account for the appearance of reduction (of course only if we accept the possibility of a consistent interpretation): neither any further observations of the observer herself nor any observations of other observers will contradict the value originally found. That is: they are arguments for the stability and the intersubjectivity of an appearance of a measurement outcome, although there is no objective matter of fact about it.

But to recover the reduction postulate we need not only a statement about the repeatability of an observation, we also need a statement about the probability of each outcome. Only these two statements together will give us the Born rule. Everett addresses this issue by considering a sequence of measurements on an ensemble of N identically prepared systems. If the initial state of each object system (numbered by the index j) is again $\sum_i c_i \varphi_i^{(j)}$, the initial state for the composite system is:

$$\Psi_0 = \sum_{i_1,\ldots i_N} c_{i_1} \cdots c_{i_N} \varphi_{i_1}^{(1)} \otimes \ldots \otimes \varphi_{i_N}^{(N)} \otimes \omega[\ldots] \tag{4.7}$$

After the first k systems are observed (k < N), the state is

$$\Psi_k = \sum_{i_1,\ldots i_N} c_{i_1} \cdots c_{i_N} \varphi_{i_1}^{(1)} \otimes \ldots \otimes \varphi_{i_N}^{(N)} \otimes \omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_k}^{(k)}\right] \tag{4.8}$$

All possible sequences of observation results $\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_N}^{(N)}\right]$ occur in this superposition. Everett defines the probability for a specific sequence as a measure M defined over the components $\omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_N}^{(N)}\right]$. He then postulates two requirements for this measure:

(1) It is a function of the coefficients c_i and invariant under a change of the phase of the c_i.

(2) It is additive for the sum of components. Everett justifies the additivity requirement with the conservation of probabilities over subsequent states Ψ_k (being steps in a sequence of measurements). The probability for a component $\omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_N}^{(N)}\right]$ of Ψ_k should be the sum of the probabilities for all components of Ψ_{k+1} that have memory sequences that begin with $\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_N}^{(N)}\right]$, i.e.

$$M\left(\omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_k}^{(k)}\right]\right) = \sum_{i_{k+1}} M\left(\omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_k}^{(k)}, \alpha_{i_{k+1}}^{(k+1)}\right]\right) \tag{4.9}$$

This requirement ensures that the probabilities of later observations can be correctly conditionalized on the probabilities of earlier observations.

He then shows that the only measure that fulfills these requirements is Born's square amplitude measure.

It follows from this definition that for long sequences (large N) the frequencies of the α_i within the memory lists will approximate the Born probabilities $|c_i|^2$ for *almost all* component states, i.e. for a subset of the component states with a measure close to one, using the measure defined above. Everett claims in the introduction to his paper (Everett 1973, 9) that he is able to deduce the probabilistic statements of the Born rule from his model. Apparently, what he has in mind is this argument linking probability to observed frequencies, which is completely analogous to the attempts to define probability from frequencies in frequentist theories of probability (mentioned in 2.3). We see from the foregoing that this claim is exaggerated. The law of large numbers for long sequences cannot be used to *define* the concept of probability in Everett's model, because the law of large numbers itself is a probabilistic statement: it requires a notion of probability already defined.[48] The introduction of probabilities is part of the empirical interpretation of the model and logically independent of the model itself. But this is not a problem for Everett's model as long as the interpretation does not lead to inconsistencies. From this point of view, Everett's discussion of long sequences of observations can be seen as a check of consistency for his definition of probability: however probability is defined in his model, the definition must imply that observed frequencies approximate the probabilities.

Everett's introduction of probability has been criticized on several counts:

(4) It has been argued widely[49] that the measure that Everett defines cannot be interpreted as a probability if one is to assume that all outcomes of a measurement coexist: Because then all outcomes occur with *certainty* and are not probabilistic events.

(5) A decomposition of a quantum mechanical state as a superposition of states is in no way unique: we can represent it as a superposition of any basis of the Hilbert space. Why should we, for the sake of the probabilistic interpretation, prefer one basis (the memory states) over any other? This problem is known as the basis problem of Everett's interpretation.[50]

---

[48]This mirrors the problem of a frequentist definition of probability in classical mechanics, see sect. 2.3.

[49]For example, (Healey 1984) or (Hughes 1989)

[50](Cartwright 1974) raised this problem as a critique of an interpretation of Everett proposed by van Fraassen.

This problem leads to a host of puzzling questions if the measurement is not exact and produces a state that is not diagonal in the memory states.

(6) Everett is silent about whether and how the measure M is defined for states other than the memory states. But it follows immediately from his definition that the measure is also defined for the states of the observed systems $\varphi_{i_1}^{(1)} \otimes \ldots \otimes \varphi_{i_k}^{(k)}$ (being the relative states of the memory states) and hence for any element of the basis

$$\chi_I = \varphi_{i_1}^{(1)} \otimes \ldots \otimes \varphi_{i_k}^{(k)} \otimes \omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_k}^{(k)}\right] \tag{4.10}$$

of the Hilbert space H_k of the composite system of observer and observed systems (I is the collective index). Now we are faced with a dilemma: if the measure M is a normal quantum mechanical measure, it is the defined for any state in H_k. According to Gleason's theorem, this means that there is exactly one quantum mechanical state (mixed or pure) that represents this measure. It is easy to see that this state is not the superposition $\sum_I c_I \chi_I$ but the mixture $\sum_I |c_I|^2 P(\chi_I)$. (The measure given by the superposition is not additive.) But this means that the assumption of the additivity requirement simply begs the question of the measurement problem: how it is that a superposition of measurement outcomes can be treated like a mixture. If on the other hand, we treat M as a classical measure only defined over the elements of the basis χ_I considered as a classical state space, then we have aggravated the basis problem (5) further: we have excluded any superposition of these states from ever occurring, that is we have introduced an explicit notion of a 'classical realm' (in form of an a priori superselection rule) for which the rules of quantum mechanics don't hold.

Notice, though, that if we were to find an answer to problem (5) that justified the existence of a preferred basis independently from the state Ψ_k, it would be possible to correct problem (6) if we accept that the concept of probability is not derived from the theory but explicitly defined: instead of the square amplitude measure, we can simply interpret the overall state Ψ_k itself as defining the measure over the final observer states. While Ψ_k does not give a classical measure on H_k, it does define a classical measure over the basis $\omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_N}^{(N)}\right]$ of the observer's Hilbert space even if the projection of Ψ_k on this space is not diagonal in the $\omega\left[\ldots \alpha_{i_1}^{(1)}, \ldots \alpha_{i_N}^{(N)}\right]$ (simply because it does so for any basis). Notice on the other hand that even if we had made sure that the projection is diagonal in the memory states, this would not answer problem (5): a mixture can be interpreted as a measure over any basis.

Before we try to interpret Everett's model, it makes sense to give another reason why we should even try. Such a reason is the fact that Everett's model can account for the EPR paradox without invoking nonlocality (Everett 1973, 82-83). Assume a particle pair S_1, S_2 in a singlet state Φ_0, separated by a large distance. Now two observers O_1 and O_2 perform measurements on S_1 and S_2, respectively. If O_1 measures a variable A whose eigenstates are φ_1, φ_2, we can write Φ_0 in this basis as

$$\Phi_0 = 1/\sqrt{2}\left(\varphi_1^{(1)}\varphi_2^{(2)} - \varphi_2^{(1)}\varphi_1^{(2)}\right) \tag{4.11}$$

O_2 measures a variable B with eigenstates η_1, η_2. The φ_i can be expressed in the basis of the η_i as

$$\varphi_i = \sum_j^2 b_{ij}\eta_j\,, \tag{4.12}$$

where b_{ij} is a 2×2 unitary matrix. The two measurements result in the state

$$\begin{aligned}\Phi_1 = 1/\sqrt{2}\Big(&\omega^{(1)}\left[\alpha_1\right]\varphi_1^{(1)} \otimes \sum_j^2 b_{1j}\eta_j^{(2)}\omega^{(2)}\left[\beta_j\right] \\ &- \omega^{(1)}\left[\alpha_2\right]\varphi_2^{(1)} \otimes \sum_j^2 b_{2j}\eta_j^{(2)}\omega^{(2)}\left[\beta_j\right]\Big)\end{aligned} \tag{4.13}$$

If O_1 now goes on to find out what results O_2 has obtained (in the way we discussed above), the resulting state is

$$\begin{aligned}\Phi_2 = 1/\sqrt{2}\Big(&\sum_j^2 b_{1j}\omega^{(1)}\left[\alpha_1,\beta'_j\right]\varphi_1^{(1)} \otimes \eta_j^{(2)}\omega^{(2)}\left[\beta_j\right] \\ &- \sum_j^2 b_{2j}\omega^{(1)}\left[\alpha_2,\beta'_j\right]\varphi_2^{(1)} \otimes \eta_j^{(2)}\omega^{(2)}\left[\beta_j\right]\Big)\end{aligned} \tag{4.14}$$

Following Everett's interpretation, we can give this state a probabilistic interpretation for observer O_1. The probability for a state $\omega^{(1)}\left[\alpha_i,\beta'_j\right]$ is the square of its coefficient, i.e., $1/2\left|b_{ij}\right|^2$. This is the probability that quantum mechanics predicts. But no assumption of a nonlocal physical process was necessary to derive this result: all that happened were two local interactions between O_1 and S_1, and O_2 and S_2, and the temporal order of the two measurements is irrelevant. Because all the possible outcomes are represented in state Φ_1, the choice of one observer which observable to measure does not affect the state of the distant particle and observer at all (and hence lead to a nonlocal process), but only determines how the observer's state is correlated to the state of the particle pair—and that is a perfectly local process. Of course, this answer is not satisfying if we are looking for an explanation of the phenomenon in classical terms, because the states Φ_0, Φ_1, and Φ_2 are entangled states: they are not separable in our definition of Section 2.2. What is important here is that there is a good sense in which the dynamical evolution is local: the measurement interaction on one side has no effect on the interaction on the other side.

Let us sum up the results of this section: Everett's model of a memory can account for the repeatability of quantum mechanical measurement. But it remains unclear how this model is to be interpreted: Everett neither gives a satisfying metaphysical interpretation of the superpositions resulting from the measurement process, nor does he give a sufficient explanation of his claim that the reduction can be understood as a subjective appearance of observers objectively in a superposition. Finally, his derivation of the Born rule is problematic. Nevertheless, his derivation of repeatability and his ability to give a local account of the EPR phenomenon seem important enough to explore his formulation of the measurement problem further. Is it possible to supply an interpretation of Everett that answers the open questions? In the following sections, I will review some attempts at such an interpretation.

4.2 The Interpretation of Everett's Theory

Everett’s model of observation had a peculiar fate. This is already evident in the name it has become known under, the ‘Many-Worlds Interpretation,’ a term that Everett himself never used. The name was coined by Bryce DeWitt who was one of several physicists who in the early seventies rediscovered Everett’s work (DeWitt and Graham 1973) which in the years before had found little attention. But where Everett’s interpretation of his model was noncommittal in metaphysical terms, DeWitt offered a bold statement of ontology:

> . . . the real universe is faithfully represented by the state vector [the superposition of measurement outcomes]. This universe is constantly splitting into a stupendous number of branches, all resulting from the measurement-like interactions between its myriads of components. Moreover, every quantum transition taking place on every star, on every galaxy, in every remote corner of the universe is splitting our local world on earth into myriad copies of itself.

This very much is a straightforward answer to our first question about the interpretation of Everett's formalism: if the universe splits upon an act of measurement, then a forteriori also the observer. After the measurement, we have for every outcome a separate universe, inhabited by an observer who is convinced that she has found that outcome -- and no other. So far, so good. But it is obvious that this interpretation is not satisfying as an answer to our questions about the measurement process. If the universe splits upon each measurement, then this is a physical process and as such is just as obscure as a physical process of reduction. It actually is nothing but a fancy way to describe the reduction process. And if we were worried about the subtle nonlocality of the reduction process as manifested in the EPR paradox, then the instantaneous splitting of the whole universe sure won't make us sleep better. Philosophers ever since have been worrying about how such a split could happen -- is it space-time that splits or rather the objects within space-time, are the splits reversible, and how can there be one state for many universes (e.g., Healey 1984)? It seems that the many-worlds phenomenon shows but one thing: that physicists have a quite different attitude towards words than philosophers, and that DeWitt, writing for a popular scientific journal, might have used a bold metaphor -- mostly for shock value.

Let us return to Everett. The closest to a mention of 'splitting' in Everett is his remark in the footnote (quoted in the last section) that after a measurement we could talk about many observers instead of one. But should we conclude from this that we have to understand Everett as postulating a physical process of 'observer splitting'? Although not as blatantly nonlocal, such a process still would not be something you would like to find in your physics textbook. Many criticisms have been leveled at the violation of conservation of energy or particle number that such a process would imply. But there is an even more straightforward argument against any numerical increase in observers: Everett's claim that he does nothing but extending the scope of quantum mechanics. And of course, in standard quantum mechanics, a system's being in a superposition does not mean in any way that there are several systems. Therefore, we would have to reintroduce the distinction between observers and 'normal' quantum mechanical systems at the level of the interpretation of a superposition. But this means that Everett would not have satisfied his own requirement.

Moreover, Everett himself states (Everett 1957, 459):

> Throughout all of a sequence of observation processes there is only one physical system representing the observer, yet there is no single *state* of the observer (which follows from the representation of interacting systems).

The remark in parentheses refers to his earlier discussion of composite systems (ibid., 456):

> There does not, in general, exist anything like a single state for one subsystem of a composite system. Subsystems do not possess states that are independent of the states of the remainder of the system, so that the subsystem states are generally *correlated* with one another. … It is meaningless to ask the absolute

state of a subsystem—one can only ask the state relative to a given state of the remainder of the subsystem.

What Everett refers to here is the fact, discussed in section 3.1, that subsystems of composite systems cannot be ascribed pure states in quantum mechanics. In 3.1, I called such states objective mixtures: they are represented mathematically as mixed states (i.e. distributions over states), but this distribution cannot be given an ignorance interpretation, that is, be understood as representing our incomplete knowledge of the true state. Everett's proposal, then, is to take this puzzling feature of quantum mechanics seriously: objective mixtures describe a system that is simultaneously in several—mutually exclusive—states or, equivalently, that has several mutually exclusive properties.[51]

This reading of Everett has long been overshadowed by the many worlds interpretation, at least in the discussions of philosophers. Here its only serious defender has been Simon Saunders.[52] Plausibly, it is also this reading of Everett that underlies the work of many of the physicists that have worked on the basis of Everett's model, even if they typically have been reluctant to engage in an analysis of the metaphysical presuppositions of their work. Noteworthy in this group is especially Wojciech Zurek in his work on decoherence (e.g. Zurek 1993) who has been the most willing to discuss 'philosophical' issues of interpretation. The decoherent histories approach[53] in quantum cosmology rests on Everett's model, and while its proponents have mostly been silent on the issue of interpretation, the approach clearly presupposes an understanding of quantum mechanics as I have sketched above: it starts from a complex quantum mechanical state of the universe that evolves unitarily and tries to find conditions on the emergence of classical behavior through decoherence of nonclassical components of the state.

It is undoubtedly true that this interpretation of Everett amounts to a radical revision of our fundamental beliefs about the logical structure of nature. But this fact in itself cannot be a reason to reject the interpretation. It has to be shown, though, that the metaphysical assumptions are self-consistent and that Everett's model under this interpretation is consistent with our experience as described by standard quantum mechanics. I will address the latter

---

[51]Notice that it still makes good sense in quantum mechanics to call states of properties mutually exclusive: according to von Neumann's identification of properties with Hilbert space projection operators, the conjunction of two properties is identified with the product of the two operators. If this product is zero, then the properties are mutually exclusive, i.e. we can never *observe* a system having both. For states, the same is true if the states are orthogonal.

[52]See (Saunders 1996) for a concise exposition of his views.

[53]For example, (Griffiths 1984) or (Gell-Mann and Hartle 1990)

question in chapter 5. The issue of self-consistency is not easy to address, especially in a theory that is so at odds with our everyday way of thinking. It can basically only be shown that certain seeming inconsistencies can be resolved and that the theory can be successfully used in a variety of situations. I will return to this issue in chapter 6. In the remainder of this chapter, I will make a preliminary attempt at clarifying some of the fundamental consequences of the interpretation and try to make plausible that they do form a coherent picture.

How can we think of a system having several incompatible properties simultaneously? There is an instructive parallel in the case of contingent properties: an object can have incompatible properties at different times. One way to think about this (that I mentioned in section 2.2) is that the object is *extended* in time, i.e. that it has temporal parts, and that it has different properties at different parts of its extension. Even simpler, a spatially extended object can have different local properties at different spatial parts of its extension. Similarly we can think of quantum mechanical systems as extended, not only in geometrical space and time, but also in *logical space,* and as having different properties at different parts of their "logical extension".[54] The logical space of a mechanical system is its state space—in quantum mechanics the Hilbert space. Therefore another way to explain this interpretation is to say that it is a realist interpretation of the Hilbert space of quantum mechanical systems, just as a geometrical interpretation of time is a realist interpretation of e.g. the Minkowski space of special relativity. This means: the ensembles that superpositions represent in Everett are not *physical ensembles* of objects (like worlds or observers) but *logical ensembles* of properties of one and the same object. The do not describe a singular matter of fact about many things, but a range of *possibilities* for one thing. We have already encountered this difference in our discussion of statistical ensembles in section 2.3: There it was the difference between a physical ensemble of particles in some piece of matter and the Gibbs ensemble of possible states for that whole piece of matter (a logical ensemble).

If a quantum mechanical state is understood as describing a range of possibilities, then the foremost question arising is: How does this interpretation represent the difference between the one state in the ensemble (for example, of measurement outcomes) that is actual and all the others that are not? Everett's answer is simple and radical: there is no objective difference whatsoever. He writes (ibid. 459, note):

---

[54]The analogy between the status of time in the theory of relativity and the status of possibility and actuality in quantum mechanics has been pointed out by Saunders (1996)

> The whole issue of the transition from "possible" to "actual" is taken care of in the theory in a very simple way—there is no such transition, nor is such a transition necessary for the theory to be in accord with our experience. From the viewpoint of the theory *all* elements of a superposition (all "branches") are "actual," none any more "real" than the rest.

Actuality is not an objective "metaproperty" that picks out certain states in the state space and distinguishes them from counterfactual possibilities, but it is an *indexical,* i.e. a relational concept that is only defined once a specific "position" in logical space is given, just as the word 'here' is only meaningful if a spatial position is given. Again, there is a close analogy to the case of time: because in special relativity the notion of simultaneity depends on the rest frame of the observer, it is quite implausible to assume that there is an objective fact about which events happen *now* in Minkowski space-time. Instead, it is natural to interpret 'now' as an indexical that is defined only in relation to a given point on a specific trajectory.[55] Therefore, it is not surprising that Everett stressed the notion of *relativity* in his model and called it "relative state interpretation". This interpretation of actuality is known in metaphysics as *modal realism.*

I will return to the discussion of the metaphysical issues connected with Everett's interpretation in chapter 6. For now, I will only remark that a consequence of the modal realism and relativism of Everett's model is that there is plenty of opportunity for terminological confusion about concepts of reality of actuality. The Everett quote above is an example for this: saying that all elements of a superposition are actual does (strictly speaking) not make sense even for a modal realist—objectively the term 'actual' is not defined, and for any given reference frame, i.e. a definite component of the state, only one element is actual. I believe that such confusions are to blame for the suspicion that Everett's interpretation (and modal realism in general) is inconsistent or incomprehensible. I will attempt to minimize the risk of misunderstanding by distinguishing a concept of *objective reality* comprising the totality of facts described by the quantum mechanical state (e.g. "All the outcomes of a measurement process are equally objectively real.") and a concept of *actuality* defined relative to a given state of a physical system (e.g. "After a measurement has been performed, only one outcome is actual for any state of the observer.")

All of this is to say: Everett does give an answer to question (1), even though this answer has been widely eclipsed by the notion of a "many worlds interpretation":

---

[55]Simon Saunders has explored this analogy in detail (ibid.)

Superpositions describe logical ensembles of possible states of one physical system. But he does not attempt to justify this answer. And it certainly would be desirable to have an argument for it before we accept such a fundamental revision of the metaphysical model of physics. Furthermore, Everett's interpretation makes question (2) even more pressing: Why should a conscious observer not be aware of this multiplicity of states if they all are to be *her* states? This means: the fundamental question of empirical adequacy is whether Everett's claim is correct that it follows from quantum mechanics (in this interpretation) that the *phenomenal* reality of conscious observers is given by the indexical 'actuality' (i.e. the relative state) and not by the concept of objective reality (the complete quantum mechanical state).

4.3 Saving Reality: Hidden Variables

The radical assumptions of Everett's theory are a good reason to wonder if we cannot make do with a more conservative metaphysical picture, but still construct a theory that avoids the troubling reduction postulate. And it looks like we have come rather close to that goal: maybe what Everett's theory shows is that Einstein's suspicion was correct, after all, that quantum mechanics cannot be a complete theory of nature. It only describes the *possible* trajectories of systems, like statistical mechanics. We have found that the simplest strategy, namely treating quantum mechanical states as statistical distributions, leads to conflicts with the predictions of quantum mechanics. But to save the objectivity of actuality, a more modest move would be sufficient: we can admit the objective reality of the quantum mechanical state but stipulate that there is an additional theoretical entity that marks what is real in the range of possibilities described by the quantum mechanical state.

Such a theory will have two kinds of state: the quantum mechanical state, evolving according to a Schrödinger equation without collapse, and a *value state* which assigns values to certain variables (the pseudoclassical or hidden variables) to which the quantum mechanical state only assigns a distribution. In a measurement situation, the value state then should single out one outcome as real—all the others are counterfactual. Even then, additional value assignments are problematic in quantum mechanics: we would get a conflict with Kochen and Specker's no-hidden-variable proof if we assumed that the value state specified values for all variables *and* that these values are revealed in a measurement. The extent to which we can assign values to variables (in addition to eigenvalues) without getting into conflict with the minimal requirement that these variables have consistent functional relations for commuting variables has been explored by Bub and Clifton (1995). Roughly speaking, their result is that only one additional maximal variable (or a commuting set of nonmaximal variables) can have a value for a given state.[56] Two basic strategies have been explored to pick this pseudoclassical variable: Either one variable can be singled out a priori, or the quantum mechanical state itself determines which variables have values. Both approaches have been used by well-known no-collapse interpretations.

The first approach is exemplified by Bohm's hidden variable interpretation.[57] Here the hidden variable is position: Every mechanical system (particle) always has an exact position and a deterministic trajectory. The quantum mechanical state represented in position variables is considered as a physical field (the *wave field*). This field evolves under the Schrödinger equation without collapse. The particles move under classical forces and an additional 'quantum force' that the wave field exerts on them. Probability in this interpretation is to be understood purely subjectively, that is statistical distributions describe of lack of exact knowledge of the particle positions. Because the Schrödinger equation is formally equivalent to a form of the classical equations of motion (the Hamilton-Jacobi equation), it turns out that one can consistently regard the square of the wave field as a probability density for the particle position: If the initial probability distribution for the position of a particle is given by the square of the wave function, then the equation of motion for the particle implies that any later probability distribution will be given by the square of the wave function at that time.

The value state in Bohm's theory is a classical state, i.e. it specifies values for all mechanical variables. The reason why this fact does not lead to conflicts with quantum mechanical predictions is that these values do not determine the measurement outcomes for variables other than position (unless the system is in an eigenstate of the variable).[58] Rather, such measurement outcomes are *contextual,* i.e. they are determined by the particle position and by how the measurement is set up. Hence a measurement of any variable other than position does not reveal anything about the true value of that variable. Nevertheless, after the measurement the particle will be in the component of the wave function that is an eigenstate of the measured variable with the measured eigenvalue. Hence the realist could take comfort

---

[56]The exact result is more complex. Simultaneous value assignments are possible in noncommuting subspaces as long as these subspaces are orthogonal to the given state. I will not discuss these results in detail here, because no attempt has been made to exploit such value assignments for the purpose of interpretation. (Obviously what is of interest for the interpretation are the *non*-orthogonal subspaces.)

[57](Bohm 1952). See (Albert 1992) for a concise presentation in the context of other no-collapse interpretations.

[58]See (Albert 1992) for a detailed discussion of this fact.

in the idea that what we are really measuring in measurements of such variables are field variables of the wave field.

But is the wave field a comfortable object to be realist about? First of all, it is not a field in geometrical space proper but in phase space. This means that for many-particle systems of N particles, it is defined not in three-dimensional space, but in 3N-dimensional space. (After all, it simply is the quantum mechanical state in a specific representation.) And it is this feature of the wave function (that it is defined on the tensor product space) that is responsible for its nonclassical behavior. Therefore that all the weirdness of entangled states is still with us, represented by the wave "field". And of course we know from our discussion of Everett what it means that the wave field represents the total uncollapsed state of the world: It describes the totality of all possible histories of the universe. All the "ontological baggage" of Everett's interpretation has not gone away—it just has been relegated to a realm of shadows, the wave field, which represents all these possible histories—and Bohm's theory explicitly requires from us to take this as an objectively real physical object. Just like in Everett's theory we need an argument why there are no interference effects from other components of the wave field, i.e. why after a measurement all other components of the wave field will have no effect on the subsequent motion of the particle.

Other than Everett's theory, Bohm's has to use nonlocal effects to explain EPR measurements. These effects result from the fact that the wave field is defined in phase space: Any change in the wave function on one side of the particle pair can have an immediate effect on its interaction with the particle on the other side. If the reason why we were unhappy with the reduction postulate was its assumption of action at a distance, Bohm's theory is not an improvement.[59]

Lastly, there is a deeper problem with Bohm's theory: As we have seen, the consistency of its probabilistic predictions with standard quantum mechanics depends on the assumption that all we know about the initial particle position is the probability distribution given by the initial wave function. Because the particle position determines any measurement outcome, if we knew any more about the initial particle position, we could predict measurement outcomes with more accuracy than quantum mechanics admits. In Bohm's theory, this assumption is justified with the fact that any measurement (also measurements of position) has to rely on physical interactions, and these interactions always are interactions with the wave function, not with the particle itself. The particle only gets carried along

---

[59]See the discussion of this issue in section 3.4.

passively with the wave function, but its position has no effect in turn on the wave function. Hence, if we perform a position measurement, the final pointer position will be correlated to the particle position only insofar as the pointer's wave function is correlated to the particle's wave function. And this means that even if we could know the pointer position exactly, all that we knew about the particle's position was its wave function relative to the pointer position (i.e. Everett's relative state) but not where within this wave function the particle is located. Put simply, through position measurements we can localize the wave function of a particle, but never localize the particle within the wave function. But this means that the only element that distinguishes Bohm's theory from Everett's, namely the particle positions, have no physical effects and are unobservable in principle.

This leaves us with a strange tension. Because in ones sense, the particle positions *are* observable: namely by introspection. After all, they are supposed to explain why after a measurement we experience one outcome as real: because it is the one designated by the hidden variables. But this means that different from any other process in nature, for which the particle positions are completely irrelevant, for consciousness they do matter. Hence, consciousness is fundamentally different a process than any other in nature. And on the same token, this means that conscious phenomena do not matter for whatever else happens in nature: They are purely epiphenomenal.

To sum up: Bohm's theory does not get rid of the "ontological ballast" of counterfactual possibilities existing in nature, because it treats the full wave function as existing objectively. Rather, it *adds* a new theoretical entity (The particle position vector) to explain why we experience only one of the possibilities as actual. The decisive question is now: Can we explain our experience of reality, namely the actuality of only one component of the quantum mechanical state, from Everett's theory alone (without invoking an additional value state)? If this is so, then the introduction of a value state serves no explanatory purpose (besides being unobservable as a physical quantity) and should fall prey to Ockham's razor. Even worse for Bohm, introducing such a value state would leave us with an inextricable sceptical problem: If the pure wave function is enough to account for the experience of conscious observers, then even introspection cannot reveal any information about the value state. The fact that I consciously experience one measurement outcome (or any state of affairs) does not imply that this is the outcome that is picked out by the value state. I might just as well be part of the "realm of shadows", the vast expanses of the wave function that are not inhabited by the value state.

The second approach for picking the pseudoclassical variable is taken by the so-called *modal interpretations.* This is a group of interpretations of quantum mechanics whose best known representatives are Richard Healey (1989) and Dennis Dieks (1989).[60] In these interpretations, the overall quantum mechanical state defines which variables have values. The pseudoclassical variables are defined for subsystems of composite systems through the biorthogonal decomposition rule discussed in section 3.1. As we have seen, given a pure state of the composite system, each of the subsystems is in a mixed state and there always are bases of the subsystem spaces that diagonalize these mixed states (eqs. 3.6 and 3.7). I already remarked then that it is very tempting to interpret these mixed states as ignorance mixtures for some pure state (which in the case of a von Neumann measurement would just be the desired result). Modal interpretations do just that: Those variables on the subsystems have values which have eigenstates that are elements of such a basis. As we have seen, this would lead to deviant predictions for variables defined on the composite system (such as correlations). Therefore, modal interpretations have to assume that the total quantum mechanical state objectively exists and that it determines the predictions for measurements on irreducible variables of the composite system. But in addition to this, there are value states for the subsystems which assign values to the additional (pseudoclassical) variables. The value states are quantum mechanical states defined on the Hilbert space of the subsystem or (as in Healey's interpretation) subspaces of this Hilbert space called system representatives.

Modal interpretations have the advantage over Bohm's interpretation that in the case of an exact von Neumann measurement they give a straightforward explanation why we always find one well-defined value: because in such a measurement both the state of the object and the state of the apparatus are mixtures diagonal in the eigenstates of the measured variable, and therefore this variable *has* a value, given by the value state. But as we have seen in section 3.1, the biorthogonal decomposition rule runs into problems in the case of non-exact measurements.

On the other hand, modal interpretations have problems that Bohm's interpretation doesn't have: It is not clear how the overall state is to be decomposed if we have composite systems of more than two parts, because in this case there is no possibility to have subsystem states that both are elements of a basis and are perfectly correlated with each other (which is what we need to explain a measurement outcome). And of course, in nature we will always

---

[60]Bas van Fraassen's interpretation (van Fraassen 1991) shows some marked differences in how the value state is assigned. I will not discuss his interpretation here. But my main criticism is aimed at the introduction of a value state in general, and therefore also addresses van Fraassen's interpretation.

encounter systems that have interacted with countless other systems in the past. Even worse, it seems a completely intractable problem to define explicit equations of motion for the value states as soon as there is any interaction between the subsystems (Albert 1992: 191–97). This means that modal interpretations don't actually give us a physical theory, but only promissory notes for one.

But finally and in my view, most fundamentally, modal interpretations share with Bohm's theory all the problems caused by the fact that they simply add a new theoretical entity to Everett's quantum mechanics without changing the empirical predictions of quantum mechanics:

– As in Bohm's theory, there are explicit nonlocalities in EPR measurements: Upon measurement on one particle, the value state of the other particle will change immediately.[61]
– As in Bohm's theory, the quantum mechanical state of the universe exists objectively, representing all possible histories of the universe. The irreversibility of measurement has to be shown:[62] Why can we ignore the total quantum mechanical state after a measurement for all practical purposes?
– Again, all interactions have to be described by the full quantum mechanical state, the value states are just being carried along. This fact is of even more central importance for modal interpretations, because we don't even know the equations of motion for value states. As in Bohm's theory, this means that we can measure value states only insofar as we can establish correlations in the quantum mechanical state. And this, like in Bohm's theory, leaves us with a picture of consciousness that is somewhat unphysical and fully epiphenomenal.

For these reasons we are led to the same conclusion as with Bohm's theory: If Everett's theory can explain the subjective appearance of single measurement outcomes, then the introduction of additional variables is simply superfluous and burdens our theory with completely unobservable entities. In the next chapter I will turn to this question, the central problem of this thesis.

---

[61]Both Healey (1989: 137–79) and van Fraassen (1991: 338–74) discuss this issue extensively.
[62]Both Dieks and Healey (1995) have appealed to decoherence for this.

Chapter 5

Quantum Mechanics and Consciousness

5.1 Supervenience and the Interpretation of Superpositions

In this chapter, I will return to the questions about Everett's model enumerated in section 4.2. In this section, I will treat Everett's claims about the appearance of superpositions to conscious observers (problems 1–3), in the next section I will discuss his claims about probability (problems 4–6).

The fundamental question about the consistency of Everett's theory with our experience is: If I am the observer who has performed an observation as described in eqs. (4.2) and (4.3), is there a contradiction between my *belief* that I have observed an outcome α_f, and the description of me as an observer given by the state Ψ_1 in eq. (4.3)? It is at this place that the consciousness of the observer comes into play. We have to assure that our "belief states" can also be described by Everett's quantum mechanical model. The novel feature in our discussion (compared to discussions of self-knowledge in the philosophy of mind) is the possibility of superpositions of conscious states. The question we have to answer is, put somewhat sloppily: "What does it feel like to be in a superposition?"[63]

First, it is necessary to precisely formulate the notion of supervenience of mental states for a quantum mechanical system. The traditional definition is that for a conscious subject, every mental state corresponds to a set of physical states of that subject (or her brain—this distinction need not worry us here), so that whenever the subject is in one of the physical states from the set, the corresponding mental state obtains.[64] Mental states are not states in the physical (mechanical) sense of the word, because they don't give a complete description of the system. It is perfectly consistent that I see a tree *and* feel a pain *and* have a longing for pickled herring. Therefore, mental states are properly speaking properties of the

---

[63]Albert 1992, 112

[64]See, e.g., (Kim 1984). Note that I don't assume that this correspondence be the same relation for different subjects. In the terminology of the philosophy of mind, I do not assume a type-type identity of mental and physical states, just a token-token identity. It is often contended that the definition of mental states also depends on matters of fact about the environment ("wide content"). But what is meant with that is the definition of types of mental states across individuals. It may very well be that the physical constitution of different individuals varies enough to make the definition of mental state types by purely intrinsic properties impossible, i.e. different mental states might supervene on the same physical state in different individuals. But this is not relevant for our discussion. I do take it as implied by the assumption of supervenience that my belief "This is a glass of water" does not become a different mental state if a mischievous naturalist, without me noticing, replaces the water with vodka.

subject and supervenience means that they are *physical* properties. I will also refer to them as *mental qualities*. An important condition for the present purposes is the requirement that mental states supervene on the *present* physical state of the system, not on its whole history. If we were to give up this requirement, there would be no need for a physical memory, because a remembrance of a past event could supervene on this event directly, without any physical memory. Then the core element of Everett's formulation, the physical model of memory would lose its function.

In quantum mechanics, the analogous statement of supervenience is that to every mental state corresponds a subspace of quantum mechanical states of the subject, so that whenever the state of the observer is within that subspace, the corresponding mental state obtains. As I mentioned in section 2.4, the reason for identifying properties with subspaces instead of subsets is the linearity of quantum mechanics: this amounts to the postulate that if two states have a property L, then any linear combination of these states also has the property L. Therefore, one has to identify the subsets of the classical state space with subspaces of the Hilbert space in quantum mechanics.

Now consider the following thought experiment, a variation of a thought experiment devised by David Albert and Barry Loewer (Albert and Loewer 1988): we have an observation modeled by equations (4.2) and (4.3). We prepare the observed system so that its initial state is an eigenstate φ_i of the observed observable A. Then the observation process happening approximately at some time t_0 will in our model look like this:

$$\phi_i \otimes \omega[\ldots] \to \phi_i \otimes \omega[\ldots\alpha_i], \tag{5.1}$$

where α_i is the eigenvalue corresponding to φ_i, and, naturally, in the final state there will be a well-defined result α_i for the observation of A. The observer's state $\omega[\ldots\alpha_i]$ corresponds to the observer's belief "I have made an observation at t_0 (as noted earlier, this information is not explicit in our notation) and observed a unique and well-defined outcome of α_i."
Equivalently, we can say that the observer's belief is characterized by the following qualities:

(1) It is a belief that the observer O has made an observation at t_0.

(2) It is a belief that O has observed a unique and well-defined measurement outcome.

(3) It is a belief $B(\alpha_i)$ that the outcome was α_i.

According to our supervenience assumption, each of these qualities defines a subspace of the space of possible quantum mechanical states of O, and $\omega[\ldots\alpha_i]$ is an element of each of these subspaces. This, of course, is true for any $\omega[\ldots\alpha_i]$. Let us call the subspaces corresponding to qualities (1), (2), and (3), S_1, S_2, and S_3^i respectively, where S_3^i depends on the value of α_i.

Now, assume that the initial state of the system is a superposition of eigenstates $\sum c_i \varphi_i$. Then, the final state will be the state Ψ_1 from eq. (4.3), which, of course, is not contained in any of the subspaces S_3^i, so on the standard account of superpositions, there is no matter of fact about the belief $B(\alpha_i)$. Nevertheless, because Ψ_1 is a linear superposition of the outcomes that would have resulted if the initial state had been an eigenstate, it still is an element of the subspaces S_1 and S_2. This means: Even in state Ψ_1, the observer believes that she has made an observation and found a definite outcome. The importance of the thought experiment is that it questions an assumption that most interpreters of quantum mechanics have made implicitly: that it is impossible that an observer be in a superposition of conscious states, because she would surely notice if she was. But the thought experiment shows that there can be no feeling of fuzziness (or blankness, or what have you) that comes with being in a superposition of conscious states if we require that feelings (as mental states) supervene over physical states. Nevertheless, the fact that Ψ_1 is an element of S_2 leaves us with a puzzle: how can we attribute a belief to O that she has observed a unique and well-defined measurement outcome if we don't attribute any one of the beliefs $B(\alpha_i)$ to her?

Albert and Loewer take the paradoxical result as an argument against supervenience. They use the thought experiment to justify their "many-minds" interpretation, the dualist picture of the brain as physical system (which is in a superposition) and different minds (corresponding to the different subjective beliefs about the measurement outcome). They take the objective state Ψ_1 to describe an ensemble of real minds which somehow coexist in the physical brain of a conscious observer. Every mind has a well-defined belief $B(\alpha_i)$, i.e. the beliefs supervene on the state of the mind, not the physical state. This is how statement (2) can be understood despite the fact that overall, there is no matter of fact about which $B(\alpha_i)$ is true. This picture suggests that the outcome of the experiment is caused by a special relation between mental states and brain states (hence the failure of supervenience).

I do not think that the experiment forces us to this conclusion. Note that the result is not due to any special property of consciousness or introspection (all we assumed was supervenience), but to the nondistributivity of quantum logic: An object can have the property $S_1 \cap S_2 \cap (S_3^1 \cup S_3^2 \cup \ldots \cup S_3^n)$ without having the property

$$(S_1 \cup S_2 \cup S_3^1) \cap (S_1 \cup S_2 \cup S_3^2) \cap \ldots \cap (S_1 \cup S_2 \cup S_3^n).$$

We would have arrived at a similarly paradoxical result if we had used a mechanical measuring apparatus and measured a set of exclusive measurement results and their disjunction. For example, assume we wanted to build an apparatus that measures whether the spin of a spin-$\frac{1}{2}$ particle is a pure eigenstate of spin in a certain direction or a superposition of these: it only beeps if the particle is in a spin eigenstate with the eigenvalue of either $+\frac{1}{2}$ or $-\frac{1}{2}$ in the given direction. If we feed this apparatus with a particle in a superposition of these two spin states (i.e., not an eigenstate), we can conclude by an analogous argument that the apparatus will beep, anyway. This means: the inability to distinguish a superposition from the logical disjunction of its components is not a special property of consciousness, but applies to all quantum mechanical measurements. Rather, it is based on the fact that properties are represented by subspaces of the Hilbert space (from which it follows that there can be no property that two states have but a superposition of them doesn't have).

I agree with Dennis Dieks who argues against Albert's and Loewer's interpretation that if we were to accept that the state after measurement is a superposition,

> it would not be obvious a priori that this would be in conflict with a physical description in terms of one single recorded spin value. Everything depends on the precise link between the mathematical formalism on the one hand and physical properties on the other; on what it means on the physical level to be represented by a superposition in the theoretical formalism (Dieks 1991).

Rather than an argument against supervenience, the thought experiment is an argument for the interpretation of superpositions as describing ensembles of properties. Again, the analogy with time is instructive: there is nothing mysterious in claiming that a (classical) particle always has a well-defined position, even when it doesn't have the same position at all times. Equivalently, we can say that an observer in state Ψ_1 has a definite belief (everywhere in her logical extension, i.e. in all her possible states) without having the same belief in all her possible states. Notice that this argument does not work if we take the superposition as describing an ensemble of objects (as in the many-worlds interpretation), because it doesn't make sense to claim that an ensemble of objects has a well-defined property because the property is well-defined for each of its elements: Properties of physical ensembles are different from the properties of their elements.

Interpreting superpositions as describing logical ensembles, then, is justified because it can explain the nondistributivity of quantum logic. The formal similarity between quantum logic and modal logic has been noted for a long time (Dalla Chiara 1986). The interpretation shows a way to understand this similarity. Of course, this argument is an inference to the best explanation: It doesn't *follow* from quantum logic that superpositions are descriptions of logical ensembles. But an inference to the best explanation, I believe, is as good as we can do. As I discussed in the introduction, the task of the interpretation of quantum mechanics is finding a metaphysical model for quantum mechanics. And like any theoretical model, this will not be a logical consequence of the data, but merely a consistent way to account for them.[65]

Notice, though, that the thought experiment by itself does not give us reason to accept the second part of Everett's interpretation, namely the modal realism about the logical ensemble. We can explain its result just as well if we assume that all but one possibility is counterfactual. Then the fact that Ψ_1 is an element of S_1 and S_2 but not of the S_3^i is simply due to its incompleteness: like a statistical state, it is not a full description of the physical situation. Of course, this assumption leads us to the hidden variable theories described in 4.3.

We still have to address Albert and Loewer's conclusion from the thought experiment: If we claim that to an observer in state Ψ_1 this state can have the appearance of a definite measurement outcome, does this not contradict the assumption of supervenience? This question can be answered if we consider the definition of supervenience. As I remarked in the last section, mental states are, properly speaking, properties of the subject and supervenience means that these properties supervene on the subject's physical *properties*. What the thought experiment shows is not that supervenience breaks down, but that the relation between quantum mechanical states and properties is more complex than in the classical case. While in classical physics, every state either has a certain physical property or does not, in quantum mechanics there is the possibility that a state is neither in the subspace corresponding to the property or in its orthogonal complement (corresponding to the negation of the property). Commonly, it is said that in this case there is no matter of fact about the system's having the property or not. The thought experiment gives us a reason to qualify this statement: We can understand it by representing such a state as a statistical ensemble of a state in which the system has the property and a state in which it doesn't. What the no-hidden-variable proofs show is that such an ensemble representation is not sufficient to determine other properties of the system (and hence also not to predict its evolution). But this only means that there is *more* to a quantum mechanical state than its being an ensemble state,

---

[65]If we were to have more than one model that does this job equally well, then all the better. But I don't think that the measurement problem is still with us because too many satisfying interpretations for quantum mechanics have been found.

not that it cannot be understood as an ensemble state at all.[66] This interpretation helps to understand the peculiar nature of improper mixtures: The formal equivalence between proper and improper mixtures is not a superficial coincidence. Improper mixtures, like proper mixtures, describe logical ensembles, and they have exactly the same form because the nonclassical nature of the total state does not affect properties that are defined on one subsystem alone.

Let us return to question (2) about the consciousness of being in a superposition. As I have argued, Albert and Loewer's thought experiment does not teach us anything new about consciousness itself. Nevertheless, it shows that if we assume supervenience, the logical ensemble interpretation has an important consequence for mental states: Being in an ensemble of mental states does not imply that we are conscious of being in such an ensemble. It might seem (at least as long as we think of the ensembles as counterfactual ensembles like in a hidden variable interpretation) that this is all we need to answer question (2). But this is not the case as long as we haven't shown that all our mental states supervene over *commuting* subspaces: While the thought experiment tells us that a superposition of mental states would not by itself carry consciousness of the fact of the superposition, it leaves open the possibility that this superposition (which is, after all, just another pure quantum mechanical state) could be a mental state in its own right, i.e. belong to the subspace that some other mental state supervenes over.

Consider the following example: assume a conscious (but not very bright) being whose brain consists of a single spin-$\frac{1}{2}$ particle, and its belief "It is raining" supervenes on some state φ_1 while its belief "It is not raining" supervenes on the state φ_2 orthogonal to φ_1. From the argument so far it follows that if this being is in any superposition of these states, it will perceive itself as having a definite belief about the weather outside. Nevertheless, it could be the case that its brain state $1/\sqrt{2}\,\varphi_1 + 1/\sqrt{2}\,\varphi_2$ comes with a belief "I am hungry." Whatever it would be like to be such a creature, it would be radically different from what we know: Its beliefs definitely cannot follow a Boolean logic like ours do (at least as long as our beliefs follow any logic). This means: our mental states are classically describable only if they supervene on commuting subspaces (and if they weren't classically describable, we wouldn't be mystified by quantum mechanics). But if we are serious about the assumption that an observer is a quantum mechanical system, how can there be anything that singles out a

[66]This is the intuition behind the path integral method in quantum mechanics: We can represent the evolution of a superposition of states by the evolution of the ensemble of these states if we allow for interference between the different states.

particular basis (or a particular set of orthogonal subspaces) as the relevant memory states, while superpositions of these states have no meaning?

There is a simple answer to this question which does not work: the observer in state Ψ_1 is not in a superposition, she is in an objective mixture. And this mixture is diagonal in the memory states, therefore it singles out these states as the relevant possible states. We have already discussed why this answer is not sufficient in section 3.3: in general, we cannot expect the mixture being exactly diagonal in the memory states if the observation process is not ideal; and, more importantly, the fact that the mixture is diagonal in one basis does not mean that the quantum mechanical state can only be understood as a measure over that basis: the objective mixture (as any quantum mechanical state) attributes a probability to any state in the Hilbert space.

But we have also found a remedy for this problem in 3.3: the assumption of decoherence. If we assume that an *observer* is a decoherent system, then any superposition of states from different coherent subspaces will quickly evolve into a mixture of these states and therefore cannot have any independent functional role in the dynamics of the observer system. But it is plausible to assume that a state that has no independent functional role will also not have an independent conscious quality. (This does not amount to the assumption of functionalism, i.e. that mental states are defined by their functional properties, but merely to the assumption that consciousness itself is some kind of dynamical process that supervenes on a physical process.) This means that all mental states will supervene over the decoherence basis (defined by the interaction with the environment) and all the subspaces representing mental qualities will commute. And that justifies Everett's assumption that the memory states supervene over a set of pairwise orthogonal states and completes our answer to question (2): Our mental states obey a Boolean logic, not because of some special status of consciousness, but because our conscious processes supervene over processes of a decoherent physical system.

Of course, the assumption of decoherence also answers question (3): Decoherence ensures that the transition to a mixture of observer states is irreversible in practice (although the overall state is still a superposition) and therefore the different components will evolve completely independently of each other in the future. Notice also that all no-collapse interpretations of quantum mechanics have to find an answer to this question and that decoherence so far has emerged as the only plausible and physically realistic scenario to give such an answer. (See the discussion in section 4.3.) The reference to decoherence for the purpose of interpreting quantum mechanics here is in no way unique.

Why should we assume that memory states decohere? As I remarked already in 3.3, it is very plausible to assume that all macroscopically different position states decohere, unless we very carefully shield the system from its environment. A more substantial argument for the decoherence of memory states of course depends on how the memory is physically realized. In the case of inanimate recording devices, the same argument as for measurement apparatus holds: the requirement that records are easily discernible already implies that different memory states interact differently with their environment. In the case of conscious observers we would need to assume some functionalist theory of mental states to make the same argument. But this is not necessary, if we have a more detailed theory of memory in conscious beings: The physical processes in the brain that are most probably relevant for our conscious processes, the firing of neurons in the case of perception and the building of synaptic connections in the case of memory,[67] both involve irreversible exchanges of energy and matter with the environment that are large on a quantum mechanical scale. Therefore, classically defined neurophysiological states should decohere just like classically defined pointer states of a measuring instrument. Decoherence therefore can address the question of a preferred set of conscious states without invoking a nonphysical "phenomenal perspective" (Lockwood 1989) or the existence of nonphysical minds (Albert and Loewer, 1988).

Notice that given our supervenience assumption, there is no need that the decoherence of observer states be perfect, that is that the set of subspaces that decoherence singles out is exactly defined, because supervenience itself need not be defined absolutely exactly. It is certainly more realistic to assume that a mental state is defined not by some specific variable on our brain (or whatever the mental state supervenes over) having an exact value, but that it has a certain 'fuzziness around the edges', i.e. that it is defined by a characteristic density operator that approximates a projection operator of a subspace, especially if we are dealing with continuous operators. This is analogous with von Neumann's treatment of continuous variables like position: Particles never have an exact position, but always a certain spread in position: their states are to be represented by wave packets, not by delta functions.

Decoherence, therefore, plays an important role in the justification of Everett's model. But the relevant assumption is not decoherence in the *objects* , but decoherence in the *subject* (the observer). Even if the state of the rest of the world was quantum mechanically coherent, a decoherent observer would find definite measurement outcomes (only the dynamical law

---

[67]According to the theory of neural networks, which is the standard neurophysiological model of memory (Lockwood 1989).

describing the outcomes would not be classical mechanics, but quantum mechanics.) Of course, this doesn't change the fact that the assumption of decoherence in macroscopic objects explains their classical behavior. (This issue is logically independent from the question discussed here.) But critics of the decoherence approach have remarked, the mere fact that the states of macroscopic systems decohere does not guarantee that observations lead to determinate results.[68] We need an argument that tells us how the objective mixtures that result from decoherence relate to our experience. It is such an argument that I have attempted to give in this section.

5.2 The Subjective State

Let us return to the situation of the thought experiment. The argument so far was supposed to show that an observer's belief $B(\alpha_i)$ that she has found a definite measurement outcome is consistent with the fact that she objectively is in the superposition Ψ_1. But this is not enough to make any predictions about the future, because it follows from the argument that the belief is consistent with a great number of different states of the observer. In von Neumann's formulation, it is postulated that the state after the measurement is an eigenstate of the measurement. If we want to show that Everett's model can replace the reduction postulate, we have to show that the observer could consistently ascribe to herself the quantum mechanical state $\omega[\ldots\alpha_i]$ corresponding to the belief $B(\alpha_i)$ and that this state ascription can be used to give correct (probabilistic) predictions for future observations. In the rest of this section I will argue that this is indeed the case.

Because the state ascription is based on a belief of the observer, I will call the state $\omega[\ldots\alpha_i]$ the *subjective state.* Note that "subjective state" here does not simply mean "state of the subject," but that the state ascription is based on a belief of the observer (and is therefore epistemologically subjective). Let us furthermore, for the moment, assume the realist position that the state (2) is the true state of the observer and the object and call this state the *objective state,* because it represents an objective matter of fact about observer and object. (I will say more about the issue of realism and the reality of quantum mechanical states later.)
But so far, we do not have the means to check equivalence:
(A) The objective state has a well-defined equation of motion (the Schrödinger equation), so we can predict future objective states. But so far, we do not know how an objective state

---

[68](Healey 1995). See also (Zurek 1991)

describes *events* , because we do not have a rule that analyzes superpositions like state Ψ_1 in terms of empirical events. (We know from the thought experiment, that there is no distinct subjective quality connected to being in state Ψ_1.)

(B) The subjective state, on the other hand, gives us a well-defined empirically accessible event, namely having the corresponding belief. But we do not have a dynamical law for subjective states which would enable an observer to predict her future from her present subjective state. Even worse: we don't even have a notion of identity over time for these states: What does it mean for an observer in some subjective state to claim that another subjective state was her past state of will be her future state, if a multitude of such states objectively exist?[69]

Let us consider problem (A) first: As we already found in section 4.2, the objective state describes a range of possible subjective states for an observer. Because the subjective states describe events, that implies that the objective state describes a range of possible events. Therefore, the obvious way of interpreting Ψ_1 in terms of actual events is understanding Ψ_1 as giving a *probability measure* for the observer having a belief $B(\alpha_i)$. Notice that probability, so introduced, is not part of the dynamics of the theory (which is still given by the unitary evolution of Hilbert space states), but of the interpretation of the theory, and logically independent of the theory itself.

But because the probability measure relates the theory to observation, we are not free in our choice of the measure if the interpretation is to be empirically equivalent with von Neumann's interpretation of quantum mechanics. The Born rule predicts that in the process of our example the probability for a measurement outcome α_i is $|c_i|^2$. Because in our interpretation the measurement outcome is not part of the objective description but is represented by the belief $B(\alpha_i)$, we have to postulate the following *correspondence rule:*

> If the objective quantum mechanical state of an observer is a superposition
>
> $$\sum_I c_I \Phi_I \otimes \omega[A_I],$$
>
> where A_I is any memory sequence, I a collective index, and $c_I \Phi_I$ the product of the corresponding states of the observed objects, then the probability that the observer is in the subjective state $\omega[A_I]$ is $|c_I|^2$.

[69]This problem has been widely discussed in the literature for the case of future states, see e.g. (Healey 1984). Healey refers to Bell noticing that the same is true about the past.

Notice that it is not trivial that we can use the quantum mechanical state to define a probability measure for the subjective states. A quantum mechanical state generally defines a measure on the subspaces of its Hilbert space. But, as we remarked earlier, the subspaces do not form a Boolean algebra (but a so-called partial Boolean algebra). Therefore the measure is not a probability measure in the usual sense (i.e. fulfills the Kolmogorov axioms).[70] I our case, though, the subspaces S_3^i defined by the observer's beliefs are pairwise orthogonal (according to the orthogonality postulate of Everett's model). But a set of pairwise orthogonal subspaces forms a Boolean subalgebra of the partial Boolean algebra of subspaces of the Hilbert space[71] Therefore, if we restrict the measure on this subalgebra, it does fulfill the Kolmogorov axioms and forms a traditional probability measure for the beliefs $B(A_i)$. But as we have seen from our discussion of projected density matrices in section 3.1, the probabilities $|c_I|^2$ are simply the probabilities given by the objective state Ψ_1 for the basis $\omega[A_I]$ of observer states. Therefore, we have answered questions (5) and (6): we can take Ψ_1 itself as defining the probability measure for the possible subjective states. This means that the correspondence rule is simply:

> The probability for a subjective state is given by the measure that the objective state defines on the subspace (of the observer's Hilbert space) corresponding to the subjective state.

Of course, this rule is perfectly natural, because, first of all, quantum mechanical states are *defined* as measures on the state space. Further, the rule explains the formal equivalence between the two kinds of mixtures we get in von Neumann's interpretation: the objective mixture of measurement outcomes attained by unitary evolution and the ignorance mixture of outcomes after the reduction process. Notice again that all of this is completely independent of the issue of modal realism: The probability measure is simply defined on a set of possible subjective states. Whether on of them is actual as an objective matter of fact or whether all of them equally exist is perfectly irrelevant. Lastly, notice again that the reason why the objective state defines a classical measure on the subjective states is *not* because it is diagonal in these states (although this is the case for our simple model of an exact measurement). Rather, it is because the subjective states form a set of orthogonal subspaces defined by the requirement of stability under decoherence.

---

[70](Suppes 1966)
[71](Hughes 1989 192-96)

As the objective state evolves, it defines a time-dependent probability distribution on the subjective states. Consider this in a simple example: a two-dimensional system with eigenstates φ_1, φ_2 and eigenvalues α_1, α_2 of the observable A, and a "model observer" with a three-dimensional state space: $\omega[\ldots]$ (the "ready" state), $\omega[\ldots\alpha_1]$, and $\omega[\ldots\alpha_2]$ (the states corresponding to a definite outcome for the observation of A). Then, in Everett's model, the observation process is described by a unitary transformation

$$(c_1\varphi_1 + c_2\varphi_2)\otimes\omega[\ldots] \to c_1\varphi_1\otimes\omega[\ldots\alpha_1] + c_2\varphi_2\otimes\omega[\ldots\alpha_2] \tag{5.2}$$

A simple time-dependent unitary evolution with the initial and final states as in (9) is given by:

$$\begin{aligned}\Phi(t) = {} & c_1\varphi_1\otimes\left(\cos t\cdot\omega[\ldots] + \sin t\cdot\omega[\ldots\alpha_1]\right) + \\ & + c_2\varphi_2\otimes\left(\cos t\cdot\omega[\ldots] + \sin t\cdot\omega[\ldots\alpha_2]\right)\end{aligned} \tag{5.3}$$

where t is a time parameter with $0 \le t \le \pi/2$.

The probability distribution induced by (5.3) on the subjective states of the observer is:

$$\begin{array}{lll} \cos^2 t & \text{for} & \omega[\ldots] \\ |c_1|^2\sin^2 t & \text{for} & \omega[\ldots\alpha_1] \\ |c_2|^2\sin^2 t & \text{for} & \omega[\ldots\alpha_2] \end{array} \tag{5.4}$$

This means: the continuous unitary evolution (5.3) of the objective states induces a time-dependent probability distribution (5.4) on the possible subjective states. This distribution is *consistent* with the observer originally being in state $\omega[\ldots]$ and at some point of time t changing into state $\omega[\ldots\alpha_1]$ or $\omega[\ldots\alpha_2]$ (with the probability $dP = 2|c_i|^2\sin t\, dt$) in which state she remains from there on.[72]

But notice that neither the evolution of the objective state nor the time-dependent probability distribution defined by it suffice to define a unique time evolution of subjective states (as in statistical mechanics at time-dependent distribution does not define an equation of motion for the micro states): All we have is a definition of global transition probabilities. The subjective state could as well flip-flop many times between the possible states. This is our problem (B). It is at this point that the ways part between a relativist (modal realist) and

[72]Notice in this example the coexistence of the objective, continuous, and deterministic process of unitary evolution, and the discontinuous and probabilistic transitions between subjective states. It is interesting to compare this model to the discussion of the temporal structure of experience in the philosophy of mind, for example with William James's lucid remarks on substantive and transitive parts of the 'stream of thought.' in (James 1981, 236-40)

an objective realist interpretation: For an objective realist, the actuality of one subjective state must express an objective matter of fact very much like the value states discussed in 4.3. Hence she has to construct objective dynamics for these states and compare the dynamics to the evolution of the objective state. As in the modal interpretations discussed in 4.3, this leads to rather intractable problems (besides the nonlocality imposed by the Bell inequalities). A relativist on the other hand only needs to show that the *beliefs* of an observer about her history are consistent with the evolution of the objective state. This means that it must be possible to *conditionalize* the objective probabilities (i.e. the probabilities given by the objective state) on a given subjective state and arrive at the probabilities given by von Neumann's theory.[73] This means that the objective probabilities conditionalized on an earlier subjective state are the Born probabilities for later measurements calculated on the basis of the subjective state.

Let us consider this in an example, too: Take an observer performing two subsequent observations on an object system. First, she observes an observable A with eigenstates φ_1 to φ_n and eigenvalues α_1 to α_n, then she observes another observable B with eigenstates ψ_1 to ψ_n and eigenvalues β_1 to β_n. A and B in general will not commute, the transformation between the φ_i and the ψ_j is given by the unitary transformation

$$\varphi_i = \sum_j b_{ij}\psi_j \text{ with } \sum_j b_{ij}^* b_{jk} = \delta_{ik} . \tag{5.5}$$

If the initial state of the object is

$$\eta_0 = \sum_i c_i \varphi_i , \tag{5.6}$$

then the objective state develops according to Everett's model:

$$\eta_0 \otimes \omega[\ldots] \rightarrow \sum_i c_i \varphi_i \otimes \omega[\ldots,\alpha_i] \rightarrow \sum_{ij} c_i b_{ij} \psi_j \otimes \omega[\ldots,\alpha_i,\beta_j] \tag{5.7}$$

First of all, there is the question of how we can say that any subjective state given by a component in the final state of (5.7) has a history, i.e. has unique predecessors under the earlier subjective states. The subjective state, being a *memory* state, makes definite claims about past events. Specifically, being in a memory state $\omega[A_1, A_2, \ldots A_n]$ implies the belief that I was in any of the states $\omega[A_1, A_2, \ldots A_k]$ with $k < n$ at some earlier time. But this means that for any subjective state there is a well-define sequence of predecessor states that represents the "subjective history" of this state. On the other hand, any subjective state has several

---

[73]Of course, an objective realist needs this argument, too, if she does not want to arrive at the conclusion that we are typically misguided about our objective history.

possible successors and the probability for each of these should be given by the Born rule. The following diagram shows the possible successor relations (as arrows) for the subjective states in our example:

$$
\begin{tikzcd}[column sep=small]
 & & & & \omega[\ldots] \arrow[dlll] \arrow[d] \arrow[drrrr] & & & & & \\
 & \omega[\ldots\alpha_1] \arrow[dl] \arrow[dr] & & & \omega[\ldots\alpha_2] \arrow[dl] \arrow[dr] & & \ldots & & \omega[\ldots\alpha_n] \arrow[dl] \arrow[dr] & \\
\omega[\ldots\alpha_1,\beta_1] & \ldots & \omega[\ldots\alpha_1,\beta_n] & \omega[\ldots\alpha_2,\beta_1] & \ldots & \omega[\ldots\alpha_2,\beta_m] & \ldots & \omega[\ldots\alpha_n,\beta_1] & \ldots & \omega[\ldots\alpha_n,\beta_n]
\end{tikzcd}
\qquad (5.8)
$$

In this case, according to the Born rule, the probability for the transition from $\omega[\ldots]$ to $\omega[\ldots\alpha_i]$ is $|c_i|^2$, the probability for the transition from $\omega[\ldots\alpha_i]$ to $\omega[\ldots\alpha_i,\beta_j]$ is $|b_{ij}|^2$.

The question of consistency in this case is whether the measures given by objective states at different times lead to a conditionalization which agrees with the transition probabilities given by the Born rule. The transition probabilities can then be understood as conditional probabilities determined by the objective probability measure. Although the answer to this question seems rather obvious, it is important to see that it is not trivial. It is well known that in general conditionalization is not possible in quantum mechanics at all, because of interference between different "histories" or "paths" of a system.[74] The reason why conditionalization is possible in our case is because every subjective state has a unique predecessor in the hierarchy. Therefore, between any two states, there is maximally one path connecting the two, and there are no interference effects. Hence, we only need to check whether the transition probabilities agree with the quotients of the total probabilities. For example, in the case of the states $\omega[\ldots\alpha_i]$ and $\omega[\ldots\alpha_i,\beta_j]$, we have for the transition probability $\mathrm{P}(\beta_j \mid \alpha_i)$:

$$\frac{\mathrm{P}(\omega[\ldots\alpha_i,\beta_j])}{\mathrm{P}(\omega[\ldots\alpha_i])} = \frac{|c_i b_{ij}|^2}{|c_i|^2} = |b_{ij}|^2 = \mathrm{P}(\beta_j \mid \alpha_i) \qquad (5.9)$$

Hence, the dynamics defined on subjective states in terms of transition probabilities given by the Born rule are consistent with the dynamics defined on the objective state given by unitary

---

[74]The standard textbook example for this is the double slit experiment. See e.g. Hughes (1989), 232-237

evolution. The different cases of observation on several systems and by several observers that Everett discusses[75] are straightforward applications of the model from equations (5.5-8) to composite systems. We have already mentioned the two most salient examples in section 4.1: the repeatability of measurement and the agreement of different observers about the value of the same observable.

We can see now that Everett's argument for the introduction of probability (section 4.1, equations 4.7-4.9) is just another example of this argument. As we have seen, Everett does not (as he seems to claim) derive the probabilistic dynamics from his model. Rather, he shows that if we interpret the measure defined by the objective state as an absolute probability, we get the usual Born probabilities as conditional probabilities for transitions between subjective states. This also clarifies the status of his argument about the law of large numbers: it shows that if we interpret the measure defined by the objective state as probability, we can say that it is probable that observed frequencies approximate the Born probabilities, because of all the subjective states possible under the objective state, such states that represent frequencies close to the Born probabilities will have the greatest objective probability.

Together, these arguments show that any future observation (including observations of other observers) will be consistent with the original assignment of the subjective state and the assumption of a probabilistic evolution according to von Neumann's reduction postulate state. This concludes our argument for the consistency of subjective state assignments. An observer can always represent her quantum mechanical evolution by a probabilistic transition law on a hierarchy of subjective states. But notice that the consistency argument relies on restricting the interactions of the observer to one specific type, namely observations as modeled by Everett's theory. For any other interaction (except those that leave the subjective states invariant) the dynamics of subjective states are not defined, nor is the notion of their identity over time. Such interactions also would allow an outside observer to measure superpositions of subjective states and show that the subjective description is incorrect. But notice that a measurement of an observable on subjective states which does not commute with the memory states would mean interfering with the function of the memory, that is, it would change the content of the memory. Therefore, it is not surprising that a consistent subjective state ascription based on the observer's memory is not possible.[76]

---

[75](Everett 1973, 68-83).

[76]Such interactions are discussed in (Albert 1992) and (Albert and Putnam 1995).

Lastly, let us return to question (4): It should have become sufficiently clear by now that the worry whether a concept of probability can even be defined in Everett's interpretation results from the confusion about the concept of reality in this interpretation. The fact that all measurement outcomes are described by the objective state does not mean that they are all actual (much less certain). I will discuss the meaning of modal statements in the next chapter, but we have already seen in the last section that all the interpretation implies about components of the objective state is that they are all *possible* as subjective states and that (when an observer performs a measurement) there is no objective matter of fact about which of these possibilities is actual for her. This does not mean that we cannot have a concept of probability, just that this concept (like all modal concepts) will have to be defined differently than in an objective realist framework: Probability is simply defined as a measure on a set of possible events (subjective states). What we don't have is the stipulation that one of these events is eventually going to be real not just as a phenomenon to me, but as a matter of objective truth—but this stipulation is completely unnecessary to understand a concept of probability. What we do need is that probabilities so defined hook up in the right way with the frequencies of events (which are the only observable element). And that this is so is shown by Everett's argument about the law of large numbers. Apparently, an important reason for the confusion about probabilities is that Everett himself did not make sufficiently clear what constitutes an event in his model (and that events are not part of the objective description). It is this gap that the introduction of subjective states is supposed to close.

We can now understand the difference between the interpretation discussed here and Albert's many-minds interpretation. Albert postulates a new kind of theoretical entities, namely minds, as correlates to our subjective appearances. These minds are *substances* in the sense of our discussion of section 2.1: they have individual histories through time. In this sense, Albert's interpretation is quite similar to the hidden-variable interpretations discussed in 4.3: as they do, Albert attempts to solve the measurement problem by introducing new theoretical entities that are thought to exist objectively. And, like in their case, my main argument against Albert's interpretation is that we can do without such new entities. (And of course, the introduction of non-physical substances into a physical theory is not a move that many philosophers would want to do lightly.) Furthermore, as in the case of hidden-variable interpretations, the introduction of individual minds as carriers of subjective perceptions necessitates postulating dynamical laws for these minds that are consistent with the overall probability distributions given by the objective state and with the no-hidden-variable theorems. Albert, like the defenders of modal interpretations, doesn't offer such a theory; and while it is easy to postulate an appropriate dynamical evolution for normal von-Neumann type observations (this is what the consistency argument of this section has shown), it is not clear at all what its general form should be.

In contrast, the subjective states discussed here are not new individual substances, but physical properties of observers. Hence there is no objective matter of fact about their identity over time, i.e. they do not have objective histories in all cases. Nevertheless, we can ascribe histories to subjective states if the overall evolution allows us to do so consistently. As we have seen in this section, this is the case for von-Neumann type observations. But also for a much wider class of interactions will it be possible to ascribe histories to subjective states if we invoke the assumption that such states are decoherent, because decoherence ensures the existence of consistent histories (Zurek 1993). Under the assumption of decoherence, Albert's and my interpretation will therefore not disagree in practice about the existence of histories, only about their status.

5.3 The Empirical State

So far, we have given an argument that Everett's formulation of quantum mechanics is empirically adequate by showing that it is consistent with a subjective appearance of well-defined measurement outcomes. It would of course by possible to claim that this is all that needs to be said about the model. But there is a central feature of the model that is unsatisfactory from a philosophical point of view: it seems natural to think of the objective state as the element of the model that describes the physical system "as it is" independent of any observation, while the subjective state is merely an erroneous description of the system brought about by the physical limitations of the process of observation. This picture leads us to the conclusion that we can never have knowledge about the state of a physical system: any observation will only create a new illusion but can never reveal the real state.[77] While Everett's argument shows that we necessarily agree in our illusions, this only means that the illusion is collective.

Notice that the argument about the observability of the objective state resembles a classical sceptical argument. We give a model of observation within our theory and from that conclude that we cannot have knowledge about objective matters of fact. I propose to deal with this problem in a way that is analogous to Kant's answer to the sceptical argument. To

---

[77]A consideration of the hierarchy (5.8) shows that every subsequent subjective state will be "further away from the truth", if we take the scalar product between subjective and objective state as the measure of distance.

put it in the simplest possible way, the answer is that the absolutely unobservable objective state cannot be what we *mean* by reality. Reality is what corresponds to our observations in the object, not the absolutely objective cause of our observations.

Kant distinguishes two concepts of reality: transcendental reality, that is, the existence of things in themselves, absolutely independent of our perception, and empirical reality, that is, the lawful existence of things as we perceive them.[78] Transcendental reality cannot be perceived although we have to assume it to explain our perceptions. What we mean by reality in our normal use of the word is empirical reality. Far from being a lawless aggregate of illusions, it is ordered by the conceptual framework that our mind imposes on it. We will return to the comparison with Kantian metaphysics in the next section. For now, regard this allusion to Kant merely as an analogy supposed to motivate the search for a correlate of empirical reality other than the objective state (which is obviously unsuitable for this purpose).

The sceptical problem is reflected in our interpretation by the fact that the transition probabilities (5.9) of the subjective state are given by the objective state. The question therefore is: can the observer in a given subjective state ascribe a quantum mechanical state to the object that leads to the correct probabilities for future observations conditional on the observer's subjective state? If there is such a state, we can regard it as representing empirical reality for an observer in the given subjective state. We have already found this state in section 3.1: It is the *relative state* to the subjective state. For any subjective state ω of the observer, the objective state ψ uniquely defines a relative state of the object. This relative state gives the correct conditional probabilities for future observations (that is, their resulting subjective states) under the condition that the observer initially was in the subjective state ω. Let us call this state of the object that corresponds to the subjective state of the observer the *empirical state*.

Notice that the correspondence is a purely physical and objective relation and is independent of any notion of meaning of the subjective state. This is important in the case of an imperfect measurement when the belief $B(\alpha_i)$ can be false. Then the empirical state is not simply the eigenstate of α_i, but a state of the object that gives the objectively correct predictions for future measurements. Consider the model of an approximate measurement with a decoherent measuring apparatus from section 3.3: The pointer basis ψ_i' there now is to

[78]For a discussion of this distinction, see Allison (1983), especially chapter 1.

be understood as the subjective states $\omega[\ldots\alpha_i]$. An observer in the subjective state $\psi_i' = \omega[\ldots\alpha_i]$ has found the measurement result α_i and hence (implicitly) believes that the object is in the corresponding eigenstate φ_i. But because the measurement is not exact, the relative state to $\omega[\ldots\alpha_i]$ (the empirical state) is not φ_i but $\varphi_j' = N_j \sum_i c_i b_{ij} \varphi_i$ (eq. 3.21) which is a "wave packet" of finite width (depending on the error of the measurement) around φ_i. (And if the observer performs a second measurement of the same variable, we can show by using the empirical state that there will be a finite probability that the second measurement does not agree with the first, just as it follows from the objective state.) This means: the empirical state does not represent an absolutely objective matter of fact independent of the process of observation, but it does not merely reflect the belief of the observer, either. As we would expect from a notion of reality, the empirical state gives us a stable criterion that the truth or falsity of an observer's beliefs can be judged by.

The dynamics of the empirical state are given by the evolution of the subjective state and the relative state formalism. This leads to a well-known picture of the evolution of the empirical state: as long as object and observer don't interact, the state undergoes the usual unitary evolution given by the Hamiltonian of the object. When object and observer interact, the subjective state changes indeterministically, and therefore the same happens to the empirical state: it changes into one of the eigenstates of the measured observable; the transition probabilities are given by the transition probabilities of the subjective state. In the example of equations (5.3-4): the initial empirical state φ_0 in this example is $c_1\varphi_1 + c_2\varphi_2$ (the relative state to $\omega[\ldots]$). When the subjective state changes at t, also the empirical state changes to either φ_1 or φ_2. The total probability for a change to φ_1 is $|c_1|^2$, for a change to φ_2, it is $|c_2|^2$. The evolution can be represented in the following diagram:

$$
\begin{array}{ccccc}
\text{empirical state} & \text{subjective state} & \text{evolve into:} & \text{entangled state} & \\
\sum c_i\varphi_i \quad \otimes & \omega[\ldots] & \rightarrow & \sum c_i\varphi_i \otimes \omega[\ldots\alpha_i] & \\
 & & & \Downarrow \ \text{equivalent to} & (5.10) \\
 & & & \varphi_f \quad \otimes \quad \omega[\ldots\alpha_f] & \\
 & & & \text{empirical state} \qquad \text{subjective state} &
\end{array}
$$

The dynamical rules for empirical states obviously are von Neumann's two processes of quantum mechanical evolution: as long as no observation happens, the evolution is unitary (governed by the Schrödinger equation); if an observation takes place the evolution is probabilistic (von Neumann's reduction process). Hence, the empirical state is equivalent to the quantum mechanical state as given by von Neumann's theory. This shows the equivalence of Everett's theory (in the interpretation offered here) and von Neumann's.

But as we can see from diagram (5.10), the empirical states themselves do not form a closed model: in any measurement interaction they evolve into an entangled state which is only subjectively equivalent to the reduced state. Therefore, we cannot simply treat them as objective states or our account of the observation process will be incomplete: it does not give us a physical model of the observation process, nor does it tell us how our subjective experience relates to quantum mechanical states. This is the most fundamental deficit of von Neumann's model. It is a deficit on his own postulate of psycho-physical parallelism that it must be possible to describe the process of subjective perception as a physical process.[79] But the process of reduction he postulated is *not* a physical process in the strict sense: it differs fundamentally from the usual unitary evolution of quantum mechanics. The project of this interpretation can be seen as a completion of von Neumann's unfinished project of psycho-physical parallelism in quantum mechanics.

The new and seemingly paradoxical feature of this interpretation is that probabilities here have a fundamentally different epistemological role than in a traditional ignorance interpretation of statistical physics: there, the values of the observables are objective, and it is the probability distributions that are subjective, resulting from incomplete knowledge of the objective state of the system. In our case, the probability distribution is objective (given by the objective state), and the possessed values are subjective. But the attempt to salvage a more traditional metaphysical model results in one of the hidden variable theories discussed in 4.3. And, as the argument in the present chapter should have shown, their additional theoretical entities (the value states of hidden variables) serve no explanatory purpose: We can account for our experience of single measurement outcomes without them. This means: if one wants to, one can of course introduce these additional variables and struggle with the host of theoretical problems they bring with them: the nonlocality, contextuality of measurement, obscure dynamics, and dualism of the physical and mental. But then one should admit the fact that their introduction does not serve any theoretical purpose (i.e. the explanation of empirical phenomena) but solely is motivated by metaphysical prejudice: the unwillingness to accept that the world *could* be the way Everett's model describes it.

---

[79](Neumann 1932, 223). See the discussion in section 3.4

Chapter 6

Reality in Quantum Mechanics

6.1 Kantian Epistemology

As we have seen, there are different possible readings of Albert's thought experiment. While Albert himself takes it as an argument against supervenience of mental states on physical states, I propose to understand it as an argument against the possibility of complete introspective knowledge of our objective physical state. In this reading, Albert's thought experiment therefore can be seen as a sceptical argument: starting from an analysis of how we arrive at beliefs about certain objects, it concludes that our beliefs are not truthful representations of these objects.[80] In the interpretation offered here, I use this argument in a way that is similar to Immanuel Kant's use of the sceptical arguments of eighteenth-century epistemology: Kant uses skepticism as a reason to reject a traditional model of epistemology which often is called the Cartesian model and to replace it with his model, which he calls transcendental idealism. With 'similarity' I mean an analogous fundamental conceptual structure, which I will sketch in this section.

It is often said that the central point of Kant's metaphysics is the distinction between appearance and the thing in itself: our knowledge is not about things in themselves, but about things as they appear to us under the conditions of our forms of intuition, space, and time, and of the pure concepts of understanding, the categories. While this point is generally well known, another equally important point is frequently misunderstood: the distinction between appearance and thing in itself does not only hold for external objects in space but also for our own representations. Also these are not known as they are in themselves but only as they appear to us. Whereas they are not subject to space as the form of outer intuition, they still fall under the condition of time because we only perceive them as they are represented by our 'inner sense,' the Kantian term for consciousness. Therefore, we also perceive ourselves as appearances and not as we are independent of our form of sensibility.

Kant discusses this distinction less prominently than the distinction for external objects but there are several places in the *Critique of Pure Reason* (Kant 1929) where it is stated unambiguously: quite obviously, in the discussion of time in the transcendental

[80]This is actually stronger than most classical skeptical arguments in epistemology, like Descartes' or Hume's, which only assert that we cannot know whether representation and object agree.

aesthetic, especially B66-69 but maybe even more pointedly in his discussion of transcendental idealism within the Antinomy of Pure Reason (B518-525) where he says:

> Even the inner and sensible intuition of our mind (as object of consciousness) which is represented as being determined by the succession of different states in time, is not the self proper, as it exists in itself—that is, is not the transcendental subject—but only an appearance that has been given to the sensibility of this, to us unknown, being. This inner appearance cannot be admitted to exist in any such manner in and by itself; for it is conditioned by time and time cannot be a determination of a thing in itself (Kant 1929, B520).

This shows that the distinction between appearance and thing in itself is independent of the distinction between object and subjective representation. The former distinction can be applied on either side of the latter. Henry Allison calls the failure to distinguish between these two distinctions the 'standard picture' of Kant interpretation. It results in a confusion of Kant's transcendental idealism with a subjective idealist position like Berkeley's (which Kant calls empirical idealist) and therefore disputes Kant's claim that he can salvage empiricism.[81]

The distinction between representation and object is an empirical distinction between different kinds of things whereas the distinction between appearance and the thing in itself distinguishes the mode of our knowledge of objects: an appearance is an object known through the representation of our senses (be this object external or an internal representation itself), and therefore according to Kant subject to the conditions a priori of our sensibility; a thing in itself is the object thought of as independent of our sensibility and its conditions. While it is impossible to know anything about the object independent of these conditions, it is still possible to think of an object in this way (even necessary, as Kant claims, once we realize that we know objects only as appearances).[82] This means, appearance and thing in itself are not two different objects (like representation and its object), but two ways of considering an empirical object.

Kant's double distinction can be applied to any epistemological model even if it does not attempt to be a priori, i.e., formulated without the use of empirical knowledge about objects, representations, and their relation. Any epistemological theory will model objects on one side and epistemic subjects and their beliefs or representations on the other side. The interpretation of such a theory is more complex than the interpretation of a theory that does

---

[81](Allison 1983, 4-5 and 247-254).
[82](Kant 1929, B306).

not contain a model of the observing subject. First of all, the model of the subject and its states (or representations) is not identical to us as the 'true' subject—as little as in any theory the model of an object is identical to the actual object. Therefore, the theoretical description of the subject is in need of an interpretation, which relates the description to our actual experience. But this interpretation in turn determines the interpretation of the model for the objects in the theory, because according to the theory, the objects are given through their relation to subjective representations as modeled by the theory. Especially, once the interpretation of the subjective representations determines which representations are observable, then this will in turn determine which features of the model for objects are observable. We therefore generally have a distinction between observable (phenomenal) and non-observable (noumenal) elements of the model, which is given by the *interpretation* of the model and is independent of the distinction between objects and their subjective representations given by the *theory* itself.

In Kant's case, the theory concentrates almost exclusively on the faculties of reason. This, of course, has to be understood from the anti-sceptical intention of his argument. Kant tries to justify the assumption of a fundamental logical structure of nature under the categories and therefore cannot start with any substantial assumptions about this structure. On the other hand, he assumes that we can have knowledge about our mental faculties by means of logical analysis, which is not vulnerable to sceptical arguments. This is not the place to assess whether Kant's theory of faculties is immune to sceptical questioning. The interesting point in our context is his version of what I call the interpretation of the epistemological theory. This argument is given in the transcendental deduction of the categories. In the following brief outline of the argument, I will not attempt to analyze this famously difficult argument, but merely compare its overall structure with the general schema given in the last paragraph.[83]

Kant starts with a fundamental principle (his 'principle of interpretation') that for a representation to be conscious it is necessary that it be unified by an act (the synthesis) of our mental faculties, because any conscious representation must carry the consciousness of its unity which cannot be part of what is given through the senses. Therefore, any conscious representation will obey certain formal conditions imposed by these functions of unity. These criteria are represented by a set of formal concepts, the categories. The second step of Kant's argument translates these formal conditions on conscious representations into conditions on

---

[83]This structure can be seen more clearly in the version of the transcendental deduction given in the second edition. At the same time, my reading of this version adds plausibility to the peculiar two-step argument offered there, which many commentators have found puzzling.

objects of our experience. This step is described by Kant's famous statement: 'the conditions of the possibility of experience in general are likewise conditions of the possibility of the objects of experience.'[84] An object can only be observable if its subjective representation fulfills the conditions of unity of consciousness, therefore any observable object will fall under certain universal rules that represent the conditions of possible experience. We can think an object independent of these conditions, but because we don't have a universal theory (ontology) of objects that would describe objects independent of the conditions of our experience, we can know nothing about such an object.

In the interpretation of Everett's theory, other than in Kant's case, we start with an empirical model of the observation process. Albert's thought experiment forces us (at least in the interpretation given here) to distinguish the state of the observer as given by the model (the true state) from its appearance to us as conscious observers of ourselves (the phenomenal state). This distinction is not part of the 'bare' theory but of its interpretation, and is analogous to Kant's transcendental distinction between appearance and thing in itself. Unlike in Kant's case, there is a formal way to represent the noumenon, because the theory is not a priori, but starts with empirical assumptions about the character of its objects. But like in Kant's case, the argument from the interpretation shows that the noumenon (the true state of the observer) cannot be known. And, like in Kant's case, the theory itself gives us the means to infer from the phenomenal state of the subject to a phenomenal state of the object (the empirical state). The dynamical rules inferred for the empirical state are the familiar von Neumann rules. This demonstrates the empirical adequacy of Everett's model, assuming that the von Neumann rules reflect our empirical knowledge about quantum mechanical systems.

The parallels between Kant's and my conceptual distinctions can be represented in the following two diagrams:

Interpretation of Everett:

$$\begin{tikzcd}
\text{phenomenal state of subject} \arrow[r, leftrightarrow, "\text{relation given by quantum mechanics}"] \arrow[d, leftrightarrow, "\text{distinction given by the interpretation}"] & \text{empirical state of object} \arrow[d, leftrightarrow, "\text{distinction given by the interpretation}"] \\
\text{true state of subject} \arrow[r, leftrightarrow, "\text{not fundamentally distinct}"] & \text{true state of object}
\end{tikzcd}$$

[84](Kant 1929, B197).

Kant's metaphysics:

$$\begin{tikzcd}
\text{representations under the conditions of consciousness} \arrow[r, leftrightarrow, "\text{empirical relation}"] \arrow[d, leftrightarrow, "\text{transcendental distinction}"] & \text{object under the conditions of experience} \arrow[d, leftrightarrow, "\text{transcendental distinction}"] \\
\text{transcendental subject} \arrow[r, leftrightarrow, "\text{not materially distinct}"] & \text{transcendental object}
\end{tikzcd}$$

Here, the columns describe the distinction within the theory between subject and object while the rows describe the distinction given by the interpretation of the theory (in Kant's case, the transcendental distinction) between appearance and thing in itself.

There is an important remark to be made about the lower row: the distinction between subject and object is itself part of the level of appearances and cannot be maintained as a fundamental distinction on the noumenal level. In quantum mechanics, this is reflected by the fact that if we assume an absolutely true state, it will be an immensely complex state of the whole universe, not a state of distinct subject and object systems. Presumably, both the observer and the object exist only in certain components of this state (certain 'branches of the wave function of the universe'). This point has been stressed in the discussion of the many-histories interpretation of quantum mechanics.[85] Here, it is interesting to see the parallel with Kant's assertion in 'The Paralogisms of Pure Reason' that on the transcendental level we cannot assume a material distinction between transcendental subject and transcendental object:

> The something which underlies the outer appearances and which so affects our sense that it obtains the representations of space, matter, shape, etc., may yet, when viewed as noumenon (or better, as transcendental object), be at the same time the subject of our thoughts.[86]

6.2 Reality in Quantum Mechanics

A Kantian epistemology in the sense sketched in the last section necessitates a reformulation of the concept of reality. Obviously, once we accept an epistemology that implies that the form of our knowledge is fundamentally determined by the way how we experience objects and not simply by the form of the objects of our experience in themselves,

[85]See, for example,(Zurek1993, 307-8).
[86](Kant 1929, A358).

we cannot anymore identify phenomenal reality (the totality of facts of our immediate experience) and objective reality (the totality of facts independent of our experience). The question is if we can have in such a theory a concept of *empirical reality* which is both epistemically accessible (which the transcedentally objective reality is not) and epistemically objective (unlike phenomenal reality which is merely an aggregate of subjective experiences). We expect epistemic objectivity to imply that empirical reality is stable over time, intersubjective, and suitable to predict future events.

Kant defines his concept of reality in the analytic of principles (B272-74) and discusses it in detail in the A version of the paralogisms (A366-80). He defines: "That which is bound up with the material conditions of experience, that is, with sensation, is actual." This means: the reality of an object is not its (noumenal) existence as a thing in itself, i.e. its independence of the conditions of our knowledge, but the fact that it is given through an actual perception. With 'bound up' Kant means that the object does not have to be the immediate object of the perception:

> What we do, however, require is the connection of the object with some actual perception, in accordance with the analogies of experience, which define all real connection in an experience in general.[87]

The analogies are the principles a priori of persistence of the substance, of causality, and of interaction of the simultaneous. They allow us to infer the reality of objects indirectly from our perceptions through empirical laws (of stability, causality, or interaction) which are granted objective validity by the a priori principles they stand under. Reality, therefore, is not given by disconnected subjective perceptions either, but is the objective correlate to the whole of our experience (the totality of perceptions through time), united by general laws.

The epistemic accessibility of reality is ensured by the fact that it is constituted of appearances and therefore given directly to our perception (A370-71), its epistemic objectivity on the other hand is to be guaranteed by the necessity of its general laws (which reflect the necessity of the synthesis of objects in the transcendental apperception, A106-10). Kant's argument for epistemic objectivity of his concept of reality is problematic: It remains questionable whether the argument in the transcendental deduction can give us a sufficient notion of objectivity or, more general, whether such an argument can be given on purely internalist grounds at all.

---

[87](Kant 1929, B272).

In the interpretation of quantum mechanics, we don't have this problem. The model itself comes with a concept of objectivity (at the cost of not being a priori in Kant's sense), so we can have an externalist argument for the epistemic objectivity of the empirical state, and this argument was given in chapter 5: The empirical state is independent of the content of the beliefs of the observer, it is intersubjectively defined, it has a well-defined dynamics and it gives objectively correct predictions for future observations. On the other hand, we have not shown yet that the empirical state is epistemically accessible. Given that it is not necessarily the state the observer believes the object to be in, how can the observer know what the empirical state of an object is? According to our definition of the empirical state as a relative state, it depends both on the subjective state and on the objective state (eq. 3.2) But the objective state cannot be perceived: this follows directly from Albert and Loewer's thought experiment. It seems that our knowledge will be forever confined to our own subjective states and our interpretation leads to the subjective idealism that realists have seen as the inevitable consequence of the Kantian program.

Let us consider the problem in our earlier example from equations (4.2-3) and diagram (5.10): There, the objective state Ψ_1 develops out of the state Ψ_0 that is a conjunction of the state of the observer before the observation and the initial state of the object. But this means: if we know the initial state of the object and have a physical description of the observation process (or at least some general information of the form: if the initial state of the system is..., then the state of the observer after the observation is...), then we can infer the objective state Ψ_1 (although it cannot be perceived). Hence, we can know the empirical state from an earlier observation if we know *its* initial empirical state. Because then, our consistency proof has shown that we can consistently use the initial empirical state for any predictions of the future behavior of the system. This means, we can use the prior empirical state in place of the objective state, and, through unitary evolution, arrive at a composite state Ψ_1, which, although it is not the true objective state, is an empirically adequate stand-in for it and especially allows us to infer the final empirical state. But now we are led to a regress: How are we to know the initial empirical state?

Somewhere we must reach a point where we don't have any prior information about the observed system. Von Neumann has shown (1932, p. 179-84), that in this case we can use the identity operator as the prior density operator of the object system, representing the fact that we have no information about its state (this is the so-called von Neumann rule). Here, the identity operator represents a mixed state that is equally distributed over the whole Hilbert space. Because the identity operator is invariant under any unitary transformation, the notion of equidistribution is independent of the basis we choose to express it in. Von Neumann shows that if the observation is exact (i.e., there is no error) and complete (i.e., the eigenstates of the observed observable are non-degenerate), the relative state after the observation will be the same as the relative state obtained from the objective state. This means: an exact and complete measurement fully determines the empirical state—the objective state plays no epistemic role.

If the measurement is approximate, we can use our model from section 3.3 again. Instead of a wave packet $\varphi'_j = N_j \sum_i c_i b_{ij} \varphi_i$ as in our discussion in section 5.3 we now obtain a mixed state $P'_j = \sum_i |b_{ij}|^2 \varphi_i \otimes \varphi_i^\dagger$. This state represents a statistical distribution that (like the wave packet) approximates the exact state $\varphi_i \otimes \varphi_i^\dagger$. If, on the other hand, the observation is not complete (that is, if the observed eigenvalue only specifies a subspace, not a single eigenstate), the relative state is an equidistributed density matrix over that subspace. In this case the relative state describes how the measurement has narrowed down the possibilities for the empirical state.

Von Neumann's argument has been criticized in the case of incomplete observation (Lüders 1951). It is true that if we assume a pure state as the objective initial state of the system, then the von Neumann rule does not yield the true empirical state. But remember that there is no way of having knowledge about the objective state independent of acts of observation. We can express this fact by postulating that the objective state is represented generally by the identity operator. The issue here is closely related to the question in statistical mechanics about the status and the justification of a maximum entropy principle.[88] Note that unlike in the classical case where we don't have a unique notion of equidistribution over the phase space (which makes the concept ill-defined as an a priori measure), the Hilbert space structure allows for a unique definition of this concept (because the identity operator is invariant under any unitary basis transformation).

We can interpret these results in two ways: The straightforward reading is that the empirical state gained through the von Neumann rule is a reflection of our ignorance of the objective state (and hence of the true empirical state): There is one true objective state of the universe, incredibly complex and completely unknown to us, that determines the correct empirical state. The empirical state gained from the von Neumann rule is an approximation

---

[88]Guttmann, forthcoming: chap. 1

to the true empirical state of the object which we can improve by refining our measuring instruments of by repeated measurements. The von Neumann rule then describes (basically like the Bayesian method in statistics) a way of updating our description when we gain new information. The true empirical state is still epistemically accessible, although only approximately in real-life situations. But a realist would not expect more than that, anyway.

But there is a second, more radical interpretation that is more in the spirit of the Kantian analogy: We can renounce the realism about the objective state altogether. As there is no absolutely objective matter of fact about the reality of one possible state of affairs, there is also no absolutely objective matter of fact about the *probabilities* of possible states of affairs. Then the empirical state given by the von Neumann rule simply *is* the empirical state of the object. Any quantum mechanical state ascription is not the description of some "fact in itself," but a description of where in the logical space of possibilities I (the conscious subject) am situated. Notice that also on this interpretation, the empirical state is epistemically objective, because its relation to the subjective state is still physical and objective, only that now the empirical state is fully defined by this relation and does not depend at all on a prior probability distribution (represented by the objective state). Borrowing from Kant's terminology, we can call the absolutely objective state a *transcendental state*. Like Kant's transcendental object, we can think of it as describing the absolutely objective matters of fact standing behind every concrete process of observation, but we are never able to determine it. On our first (realist) interpretation, this state is some unknown pure state of the object, on the second (idealist) interpretation it is an equidistributed density matrix *representing* our ignorance of the state prior to any measurement.

For any practical predictions we make, the choice of interpretation is completely irrelevant because the most we can ever know is the empirical state gained by the von Neumann rule, which, as we have seen, is independent of the choice of the prior (transcendental) state of the object.. The difference only shows in contexts like cosmology, where the non-realist interpretation forces us to abandon the notion of the objective state of the universe as a basis for explanations in cosmology. Rather, we can only speak about the state of the universe "from our point of view," i.e. conditional on the matters of fact here and now. This is analogous to the methodological approach advocated in cosmology as the *anthropic principle* (Harrison 1981): It has been noted that complex biological processes and hence consciousness in the universe seems only possible if certain cosmological constants have narrowly defined values. The anthropic principle postulates that the values of these constants should be *explained* by the fact that conscious beings like us exist. Similarly, on a non-realist interpretation we cannot explain our present observations from a (transcendental) total state of the universe. Rather, we deduce an empirical state of the universe from our present observations (which is all we do in practice, anyway). It is interesting to note that our interpretation could be a basis for the anthropic principle, if dynamical processes could be found that determine the values of cosmological constants.

Let us return to the concept of empirical reality: Reality should be more than a momentary empirical state: it should comprise a notion of history that is as epistemically objective as the present empirical state of the world. We have already seen that because subjective states are memory states, they contain information about past subjective states of an observer. This means that they represent a subjective history of the observer. For each stage of this history, there is a corresponding empirical state of some object. But this is still not all there is to empirical reality: We also consider facts as real that we have only inferred knowledge of. But this means that we extrapolate the empirical state from our knowledge about the dynamics of a system. This is just Kant's definition of reality quoted in the beginning of this section: Real is every fact that is lawfully connected to our observations.

The new phenomenon in quantum mechanics is that these extrapolations will not always match seamlessly: If we measure noncommuting observables, we will have a situation as in diagram (5.10), where the state inferred from an earlier observation and the state inferred from a later observation are not identical. This means that in quantum mechanics, reality is not a seamless whole: there can be breaks between the "patches of reality" defined by different observations. These seams are nothing but the instances of state reduction in von Neumann's formulation. Now we can understand von Neumann's observation (which we discussed in section 3.4) that the point in time at which reduction happens can be moved at will: this reflects the fact that the different states (before and after reduction) don't represent an objective change in the object, but merely a change in perspective, namely which subjective state the description is made from. Von Neumann's proof for the mobility of the time of reduction is identical to our argument that we can conditionalize on subjective states.[89]

Everett's model therefore does not only reproduce the empirical predictions of non Neumann's. It leads to exactly the same concept of empirical reality. The difference is that in Everett's model this reality is intertwined with countless possibilities that exist beside it. And this brings us back to the question about Everett's model that probably has vexed philosophers

---

[89]Sect. 5.2

the most: How can it be said that possibilities exist? I believe that the answer to this has to be pragmatic: They are (so I have tried to show) necessary elements of a highly successful theory of nature. Neither the attempt to cut them out of the picture (by postulating a physical process of reduction) nor the attempt to make them part of reality (for example, by reinterpreting them as a field as in Bohm's theory) leads to satisfying results. In either case, we will cone into conflict with the no-hidden-variable proofs of quantum mechanics. Although these conflicts can be solved by brute force, the resulting theories have an air of desperation: They need a huge theoretical apparatus with unobservable and unphysical interactions, variables, and states. On the other hand, as I have tried to argue in this chapter, admitting possibilities into existence does not mean that we lose a coherent concept of reality. So why should we not admit them into our scientific ontology, just as we have admitted atoms, fields, or the curvature of spacetime?

6.3 Conclusion

Let me briefly recapitulate the assumptions made in the argument of this thesis that Everett's theory can account for the phenomena of quantum mechanics:

(1) The metaphysical model.
Quantum mechanical states describe coexistent ensembles of possible facts. Moreover, there is no objective matter of fact which of these possible states is actual. Actuality is only defined relative to a given possible state.

(2) The decoherence assumption.
Conscious observers, like all other macroscopic objects under normal circumstances, are decoherent systems. They are in constant interaction with their environment. This interaction singles out a set of (pairwise orthogonal) subspaces which contain the only dynamically stable states of the system.

(3) The memory assumption.
Conscious observers have memories, i.e. physical systems whose states contain stable records of past states.

(4) The supervenience assumption.
The mental qualities (mental states) of conscious observers supervene over their physical properties.

As I discussed in the introduction, the metaphysical model gives an interpretation of the mathematical formalism of quantum mechanics. This interpretation implies an interpretation in terms of observations (namely, the Born rule), but it is more general than such an interpretation: It gives an interpretation in terms of "what the world is like", i.e. in terms of a set of fundamental (metaphysical) entities. The model may seem rather extravagant. But it is interesting to note that it is not really new in physics. A model of interfering possibilities is implicitly contained in the path integral method. In this method, the quantum mechanical amplitude of a state is calculated by adding contributions from all possible paths (histories) leading to that state (Feynman 1948). Historically, the intuition behind the many histories interpretation was based on this method. This is a good example for the claim from the introduction that a metaphysical model grows out of the application of the theory. One might conjecture that, if Feynman had not been so afraid of getting drawn into metaphysics, he could have seen that the intuition behind the development of his path integral method leads to a radically new interpretation of quantum mechanics.

Assumption (4), on the other hand, is an embedding assumption: It shows how to embed the metaphysical model of classical mechanics into the metaphysical model (1). But (4) uses a fundamentally new concept of subjective appearance. Therefore, it does not simply reduce classical mechanics to quantum mechanics. (4) has a similar status as the ergodic principle in statistical mechanics or the supervenience assumption in philosophy of mind: All these principles make us understand how two different theories can be consistent with each other by constructing models of the theoretical entities of one theory out of the theoretical entities of the other.

The decoherent histories interpretations have concentrated on criterion (2), and to lesser extent, on criterion (3). But this is not sufficient for an interpretation of quantum mechanics. Without assuming a metaphysical model like (1) it is not clear why and how decoherent mixtures of states should be understood as probability distributions. Assumption (4) is necessary to establish a concept of reality. Describing measurement outcomes as mixtures is not sufficient for this end, as I discussed in section 3.3. Zurek acknowledges the insufficiency of (2) and (3):

> This issue of the "collapse of the wave packet" cannot be really avoided: ... we perceive outcomes of measurements and other events originating at the quantum level alternative by the alternative, rather than all of the alternatives at once.
>
> An exhaustive answer to this question would undoubtedly have to involve a model of "consciousness", since what we are really asking concerns our

> (observers) impression that "we are conscious" of just one of the alternatives. Such model of consciousness is presently not available (Zurek 1993, 310).

Zurek addresses the problem by assuming that consciousness is to be understood as a kind of information processing; and in physical information processing devices the physical state must in some way reflect the information content. He sees these assumptions as "considerably more speculative than the rest of the paper" (ibid.), which is concerned with decoherence. In section 5.1 I have attempted to show that speculation on the nature of consciousness is not necessary for the purpose at hand: The supervenience assumption is all we need. (Notice that supervenience is also hinted at by Zurek's account.) And it had better be this way: for if our concept of reality depended on a substantive assumption about the nature of consciousness, there would be little hope of extending it beyond the state of the subject, and we would be left with a subjective idealism of Albert's kind.

This concludes my argument that the metaphysical model for quantum mechanics offered in this thesis can give an account of measurement. But of course, this is not all that needs to be done. On the account of a metaphysical model given in the introduction, the decisive criterion for a metaphysical model is its usefulness for concrete applications of the theory. Lewis makes a similar argument for his modal realism: "Why believe in a plurality of worlds?—Because the hypothesis is serviceable, and that is a reason to think that it is true (Lewis 1986, 3)." I will hold off the issue of truth until later, but usefulness should all the more be the decisive criterion for the worth of a model. Although a theory of measurement is one such application and although it has played a prominent role in our attempts to understand quantum mechanics, it certainly is not the only application of quantum mechanics we want. The exploration of other possible uses for my model is beyond the scope of this thesis. But we have encountered one substantial example here: The research about decoherence, which is based on the intuitions of this model. The theory of decoherence has practical impact that is independent of any metaphysical model, and which goes far beyond a model for the measurement process.

I believe that the kind of "modal realism"[90] advocated here has many more fruitful applications. This is a promissory note, not more. It should be plausible, however, that if

[90]I would rather not use this label, which even Lewis himself considers somewhat of a misnomer (Lewis 1986, viii). For one it seems to imply the categorical mistake to assume that possibilities are real objectively where I want to say that reality is not an objective matter. Secondly, this position has nothing to do with scientific or metaphysical realism which assumes that it is facts about the world "in itself" that make scientific theories true or false (Lewis's opinion notwithstanding). Therefore, I much prefer the association with transcendental idealism described in section 5.4.

Lewis's possible worlds are "a paradise for philosophers (ibid., 4)," the assumption of causal efficacy of possibilia implied in our model will make this paradise all the more relevant to our world. We may expect consequences for our accounts of causality, intentionality, and the like. The discussion of section 5.2 hints at an account of our experience of time that is much richer than what can be had from the classical model.

Lewis admits that the attractivity of his metaphysics depends crucially on the impossibility of getting its benefits "more cheaply elsewhere (ibid., 136-91)." For this I believe to have a good case: The persistence of the measurement problem itself shows that already the modest benefit of having a well-defined reality in quantum mechanics is not to be had cheaply elsewhere. Nevertheless, one might be reluctant to let go of a metaphysics so deeply engrained in our thinking. Against that I can only hold that there are good reasons not to feel all too nostalgic about the classical model. It surely has had its many good uses. But it also has had quite a few failures. In Chapter 2, we encountered some examples: the problems of a classical account for probability or for the tense structure of time. Mere metaphysical conservatism should not stop us from trying new roads exploring such issues.

And finally, what about truth? Does the usefulness of our model entail its truth? We have seen that Lewis thinks so. But, one might object, metaphysics is not (like science may be) about the construction of useful models, it is about how the world truly is. And the usefulness of a model has no bearing on whether it is true of the world. The reader may expect by now that I align myself with Kant and consider this dispute as misguided, in science as well as in metaphysics. Neither metaphysics nor science are about how the world is in itself, but about how it appears to us. And we cannot anymore in metaphysics than in science, step outside of any conceptual framework to decide whether a conceptual framework is true or not. I do not share Kant's optimism that we can turn this limitation into a method to settle the problems of metaphysics once and for all by constructing a metaphysics from the inside, i.e. from the conditions of our experience. But this, I believe, means that metaphysics, like other sciences, is about constructing models of what the world is like.

Reference List

Albert, David Z. 1992. *Quantum mechanics and experience* . Cambridge, Mass.: Harvard UP.

Albert, D. and B. Loewer. 1988. Interpreting the Many-Worlds Interpretation. *Synthese* 77, 195-213.

Albert, D. and H. Putnam. 1995. Further adventures of Wigner's Friend. *Topoi* 14, 17-22.

Allison, H. E. 1983. *Kant's transcendental idealism*. New Haven: Yale University Press.

Amann, A. 1992. Must a molecule have a shape? *South African Journal of Chemistry* 45: 29-38.

Aristotle. 1908. *Metaphysica*. Translated by W. D. Ross. The Works of Aristotle, eds. J. A. Smith and W. D. Ross, vol. 8. Oxford: Clarendon Press.

Aristotle. 1928. *Categoriae*. Translated by E. M. Edghill. The Works of Aristotle, ed. W. D. Ross, vol. 1. London: Oxford University Press.

Bell, J. S. 1964. On the Einstein Podolsky Rosen paradox. *Physics* 1: 195-200.

Beltrametti, Enrico G., and Gianni Cassinelli. 1981. *The logic of quantum mechanics*. Reading, Mass.: Addison-Wesley.

Benthem, Johan van. 1983. *The logic of time*. Dordrecht: Kluwer.

Birkhoff, G. and J. von Neumann. 1936. The logic of quantum mechanics. *Annals of Mathematics* 37: 823-43.

Bohm, David. 1951. *Quantum Theory*. Englewood Ciffs: Prentice-Hall.

________. 1952. A suggested interpretation of the quantum theory in terms of "hidden variables", part I and II. *Physical Review* 85: 166-93.

Bohr, Niels. 1935. Can quantum-mechanical description of physical reality be considered complete? *Physical Review* 48: 696-702.

Bub, J. 1996. Complementarity and the orthdox (Dirac-von Neumann) interpretation of quantum mechanics. In *Perspectives on Quantum Reality,* ed. R. Clifton, 211-26. Dordrecht: Kluwer.

Cartwright, Nancy. Van Fraassen's modal model of quantum mechanics. *Philosophy of Science* 41:199-202.

________. 1983. *How the laws of physics lie*. Oxford: Clarendon Press.

________. 1989. *Nature's capacities and their measurement*. Oxford: Clarendon Press.

Cushing, J. 1982. Models and methodologies in current theoretical high-energy physics. *Synthese* 50, 5-101.

Dalla Chiara, M. L. 1986. Quantum Logic. In *Handbook of Philosophical Logic,* vol. III, eds. D Gabby and F. Guenthner, 427-69. Dordrecht: Reidel.

Davidson, D. 1970. Mental events. In *Experience and Theory*, eds. L. Foster and J. W. Swanson, 79-101. Amherst: University of Massachusetts Press.

de Finetti, B. 1937. *La prévision: Ses lois logiques, ses sources objectives*. Paris: Institut Henri Poincaré.

D'Espagnat, B. 1990. Toward a separable 'empirical reality'?' *Foundations of Physics* 20: 1147-72.

Danieri, A. Loinger, G. M. Prosperi. 1962. Quantum theory of measurement and ergodicity conditions. *Nuclear Physics* 33: 297-319. Reprinted in (Wheeler, Zurek 1983): 657-79.

DeWitt, B. S. and N. Graham. 1973. *The Many–Worlds Interpretation of quantum mechanics*. Princeton: Princeton University Press.

Dieks, D. 1989. Quantum mechanics without the projection postulate and its realistic interpretation. *Foundations of Physics* 19: 1397-1423.

________. 1991. On some alleged difficulties in the interpretation of quantum mechanics. *Synthese* 86: 77-86.

Dowker, H. F., and J. J. Halliwell. 1992. Quantum mechanics of history: The decoherence functional in quantum mechanics. *Physical Review D* 46: 1580-1609.

Earman, J. 1986. *A primer on determinism*. Dordrecht: Reidel.

Einstein, A., B. Podolsky, and W. Rosen. 1935. Can quantum mechanical description of physical reality be considered complete? *Physical Review* 47: 777-80.

Everett, H. 1957.'Relative state' formulation of quantum mechanics. *Reviews of Modern Physics* 29: 454-62.

Everett, Hugh. 1973. The theory of the universal wave function. In: (DeWitt and Graham 1973).

Feynman, R. P. 1948. Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics* 20: 367-87.

________. 1965. *The character of physical law*. Cambridge, Mass.: M.I.T. Press.

Gell-Mann, M., and J. B. Hartle. 1990. Quantum mechanics in the light of quantum cosmology. In *Complexity, Entropy, and the Physics of Information*, ed. W. H. Zurek. Reading: Addison-Wesley.

Gibbs, Josiah Willard. 1902. *Elementary principles in statistical mechanics*. New York: C. Scribner's Sons.

Gillies, D. A. 1973. *An objective theory of probability*. London: Methuen.

Gleason, A. M. 1957. Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics* 6: 885-93.

Griffiths, R. J. 1984. Consistent histories and the interpretation of quantum mechanics. *Journal of Statistical Physics* 36: 219.

Guttmann, Y. Forthcoming. *Probability without Indeterminacy: The Concept of Probability in Statistical Physics.*

Harrison, E. R. 1981. *Cosmology. The Science of th Universe*. Cambridge: Cambridge University Press.

Hartle, James B. 1991. The quantum mechanics of cosmology. In *Jerusalem Winter School for Theoretical Physics: Quantum cosmology and baby universes*; Singapore: World Scientific.

Healey, R. 1984. How many worlds? *Nous* 18:591-616.

________. 1989 *The philosophy of quantum mechanics*. Cambridge, UK: Cambridge University Press.

________. 1995. 'Modal' interpretations, decoherence and the quantum measurement problem. Paper given at the workshop on *Quantum Measurement: Decoherence and Modal Interpretations*, University of Minnesota.

Heilbron, J. 1977. Lectures on the History of atomic physics 1900-1922. In *Storia della fisica del XX secolo: Proceedings of the International School of Physics "Enrico Fermi",* course LVII, 40-72. New York: Academic Press.

Heisenberg, Werner. 1925. Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik* 33: 879-93

Hughes, R. I. G. 1989. *The structure and interpretation of quantum mechanics*. Cambridge, Mass.: Harvard University Press.

James, W. 1981. *The Principles of Psychology*. Cambridge, Mass.: Harvard University Press.

Jarrett, J. P. 1984. On the physical significance of the locality conditions in the Bell arguments. *Nous* 18: 569-89.

Jauch, J. M. 1968. *Foundations of quantum mechanics*. Reading, Mass.: Addison-Wesley

Jones, M. R. 1991. Locality and holism: The metaphysics of quantum theory. Ph.D. diss., Stanford University.

Kant, I. 1929. *Critique of pure reason*. Translated by N. Kemp Smith. New York: St. Martin's Press.

________. 1952. *Critique of Judgement*. Translated by J. C. Meredith. Oxford: Clarendon Press.

Kim, J. 1984. Epiphenomenal and supervenient causation. In *Midwest Studies in Philosophy* IX, 257–70. University of Minnesota Press, Minneapolis.

Kochen, S., and E P. Specker. 1965. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics* 17: 59-87

Kolmogorov, A. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Julius Springer.

Kripke, S. 1963. Semantical considerations in modal logic. *Acta Philosophica Fennica* 16: 83-94.

Laplace, Pierre S., marquis de. 1995. *Philosophical Essay on Probabilities*. Translated by Andrew I. Dale. New York: Springer.

Lehner, Christoph. 1986. Das Fulgurationsproblem und Kants Dialektik der teleologischen Urteilskraft. Seminar paper, University of Munich.

________. 1997. What it feels like to be in a superposition. And why. *Synthese* 110: 191-216.

Lewis, D. 1986. *On the plurality of worlds*. Oxford: Blackwell

Lockwood, M. 1989. *Mind, brain, and the quantum*. Oxford: Blackwell.

London, F. W. and E. Bauer. 1982. The theory of observation in quantum mechanics. Translated by A. Shimony, J. A. Wheeler, W. H. Zurek, J. McGrath, and S. McLean. In (Wheeler and Zurek, 1983, 217-59).

Lüders, G. 1951. Über die Zustandsänderung durch den Meßprozeß. *Annalen der Physik* 8: 323–328.

Mises, Richard von. 1957. *Probability, Statistics and Truth*. Translated by Hilda Geiringer. London: Allen and Unwin.

Mott, N. 1929. The Wave Mechanics of α-Ray Tracks. *Proceedings of the Royal Society* A126: 79–84.

Neumann, Johann von. 1932. *Mathematische Grundlagen der Quantenmechanik*. Berlin: Julius Springer.

Prosser, Reese T. 1966. A new formulation of particle mechanics. *Memoirs of the American Mathematical Society,* vol. 61.

Putnam, H. 1962. What theories are not. In *Logic, Methodology and the Philosophy of Science: Proceedings of the 1960 International Congress,* eds. E. Nagel, P. Suppes, and A. Tarski, 240-51. Stanford: Stanford University Press.

Ramsey, F. P. 1950. Truth and Probability. In *The Foundations of Mathematics,* ed.R. B. Braithwaite. London: Routledge and Kegan Paul.

Saunders, S. 1996. Relativism. In *Perspectives on Quantum Reality,* ed. R. Clifton. Dordrecht: Kluwer.

Shimony, A. 1980. The point we have reached. *Epistemological Letters*, June 1980.

Suppe, F. 1977. The search for philosophic understanding of scientific theories. In *The Structure of Scientific Theories*, ed. F. Suppe. 2d ed., 1-241. Urbana: University of Illinois Press

________. 1989. *The semantic conception of theories and scientific realism.* Urbana: University of Illinois Press

Suppes, Patrick. 1966. The Probabilistic argument for a nonclassical logic in quantum mechanics. *Philosophy of Science* 33: 14-21

________. 1967. What is a scientific theory? In*Philosophy of Science Today*, ed. S. Morgenbesser, 55-67. New York: Basic Books.

Suppes, Patrick, and Mario Zanotti. 1976. On the determinism of hidden variable theories with strict correlation and conditional statistical independence of onservables. In *Logic and Probability in Quantum Mechanics,* ed. P. Suppes, 445-55. Dordrecht: Reidel.

Unruh, W. G., and W. H. Zurek. 1989. Reduction of a wave packet in quantum Brownian motion. *Physical Review D* 40: 1071

van Fraassen, B. C. 1991 *Quantum mechanics: An empiricist view*. Oxford: Clarendon Press.

Wheeler, J. A. 1957. Assessment of Everett's "relative state" formulation of quantum theory. *Review of Modern Physics* 29: 463-65

________. 1977. Include the observer in the wave function? In *Quantum Mechanics, a Half Century Later*, J. Leite Loper and M. Paty, eds. Dordrecht: Reidel.

________. 1979-81. Law without law. In (Wheeler, Zurek 1983, 182-213)

Wheeler, J. A. and W. H. Zurek eds. 1983. *Quantum theory and measurement.* Princeton: Princeton University Press

Wigner, E. 1961. Remarks on the mind-body question. Reprinted in (Wheeler, Zurek 1983, 168-81).

________. 1970. On hidden variables and quantum mechanical probabilities. *American Journal of Physics* 38: 1005-9

________. 1976. Interpretation of quantum mechanics. In (Wheeler and Zurek 1983): 269-314

Zurek, W. H. 1991 Decoherence and the transition from quantum to classical. *Physics Today* 44: 36-44.
Replies by L. Ballentine, P. Pearle, et al.: *Phys. Tod.* 46 (1993): 13-15, 81-90).

________. 1993. Preferred states, predictability, classicality, and the environment-induced decoherence. *Progress of Theoretical Physics* 89: 281–312.