

The Statistical Interpretation of Quantum Mechanics

Hans Rudolf Tschudi
Innere Altachen 4
CH-4800 Zofingen
Switzerland

hansrudolf.tschudi@alumni.ethz.ch

August 28, 2020

Abstract

The topic of the present inquiry is the foundation of the statistical interpretation of quantum mechanics, the interpretation used, explicitly or implicitly, in all practical applications of quantum theory. Although it was the first coherent interpretation of the complete quantum formalism and had eminent supporters as Born, Einstein, or Schrödinger, there was, from the outset, a deeply rooted dissatisfaction with the probabilistic character of its predictions inciting the search for alternatives. The most prominent among these is the so-called Copenhagen interpretation developed by Bohr, Heisenberg, Pauli, and others asserting that a pure quantum state does not describe an ensemble of physical systems as does the statistical interpretation, but provides a complete description of an individual system and that a quantum system can be only grasped in complementary classical pictures. At present, the statistical interpretation is considered as one of several different, but more or less equally founded interpretations. This is clearly no satisfactory state of affairs all the more than Copenhagen type of interpretations are plagued by all kinds of 'fundamental problems' completely absent in the statistical interpretation. Moreover, the interpretation of the quantum mechanical formalism should be based on better foundations than just on mere convenience or on the matter of taste. Such a better foundation is provided by the fact that physical laws do not relate directly to individual physical systems but to experimental arrangements, whole classes of similar physical systems, and to an individual system only insofar as it is a member of such a class. The most general physical statement is, therefore, a probability prediction for the outcome of single experiments revealing the statistical interpretation of quantum mechanics as the correct and fundamental one.

The statistical interpretation of quantum mechanics considers the experimental fact that physical processes are causal but basically indeterministic in contrast to the causal and deterministic classical theories. The statistical prognoses of quantum theories are objective properties of the experimental arrangement considered, and observable values are definite, objective properties of concrete, individual systems. These properties of an individual physical system always have definite values which may be determined by suitable measurements, but it is not possible to prepare the system in such a way that all observable values can be sharply predicted because of the noncommutativity of the quantum mechanical observable algebra. This indicates a limitation on principle of human command on nature. The concepts of objectivity and realism accord with the status of physics as the base science of engineering and technology. It is shown that these concepts do not entail paradoxes or contradictions in the statistical interpretation of quantum mechanics. It is shown that classical hidden variables theories, supposed to underly quantum theories, do not exist if the quantum mechanical Hilbert space has dimension greater than two. Quantum indeterminism is intrinsic, not reducible to mere subjective ignorance. In particular, there is no hidden variables model for Bohm's setting of the Einstein-Podolsky-Rosen thought experiment irrespective of whether it reproduces quantum mechanical predictions or not. This undermines the logical foundation of Bell's theorem. As a consequence, 'nonlocality proofs' for quantum mechanics based on Bell's inequality are not conclusive. In the statistical interpretation, the quantum mechanical description of isolated, free systems is, fortunately, strictly local.

Contents

1	Introduction	3
2	The general scheme of quantitative physical theories	6
2.1	Characteristics of physical statements	6
2.2	Interpretation of the notion 'probability'	8
2.3	Epistemological status of observables and states, attributes of concrete material systems . . .	9
2.4	The general formal scheme of quantitative physical theories (GFS)	10
2.5	Quantum and classical theories as examples of the GFS	13
3	The problem of hidden variables in quantum mechanics	15
3.1	Von Neumann's theorem	15
3.2	The Kochen-Specker theorem	18
3.2.1	The case of two dimensions	20
3.3	Bell's 1966 paper	22
3.4	Mermin's observable sets with no classical model	25
3.5	Criticism on the Kochen-Specker theorem	27
3.6	The class of classical models is incomplete	28
4	Further comments on state preparation and measurement of observable values	30
4.1	Platt's quantum mechanical model for the Stern-Gerlach experiment	30
4.2	Solution in the Heisenberg picture	32
4.3	Discussion of the Stern-Gerlach experiment	34
4.4	'Fundamental problems' in the Copenhagen interpretation	35
5	Bohm's version of the Einstein-Podolsky-Rosen thought experiment	37
5.1	Consequences of Bohm's setting	38
6	Bell's theorem and the question of locality	41
6.1	Bell's inequality and Bell's theorem	41

6.2	Bell's theorem without inequalities and probabilities	43
6.3	Quantum Nonlocality?	45
7	Galilei invariance and the quantum mechanics of nonrelativistic particles	48
7.1	The Galilei group	48
7.1.1	The Lie algebra of a matrix Lie group	49
7.1.2	A basis for the Lie algebra of the Galilei group	50
7.2	Unitary representations	52
7.3	Unitary ray representations	53
7.4	Unitary ray representations of the Galilei group in $\mathbb{L}^2(\mathbb{R}^3)$	55
7.4.1	The generators \mathbf{P} and \mathbf{L}	56
7.4.2	Time translations and time derivatives	57
7.4.3	There is no unitary representation of the Galilei group in $\mathbb{L}^2(\mathbb{R}^3)$	57
7.4.4	The generators \mathbf{Y} and H for unitary ray representations	58
7.4.5	Actions of 1-parameter unitary subgroups generated by $H, \mathbf{P}, \mathbf{L}, \mathbf{Y}$ on polynomials $T(\mathbf{X}, \mathbf{P})$	60
7.5	The quantum mechanics of a particle without inner degrees of freedom	61
7.6	An isolated system consisting of two interacting particles	62
A	Elements of probability theory	66
B	The Chebyshev inequality	70
C	Simultaneous eigenstates for Mermin's observables $A_1 - A_4$	71
D	The time evolution of S_1 and S_2	74
E	\mathbf{X} and \mathbf{P} are irreducible on $\mathbb{L}^2(\mathbb{R}^3)$	76

Chapter 1

One arrives at very implausible theoretical conceptions, if one attempts to maintain the thesis that the statistical quantum theory is in principle capable of producing a complete description of an individual system. On the other hand, those difficulties of theoretical interpretation disappear, if one views the quantum-mechanical description as the description of ensembles of systems.

A. Einstein [1, p. 671]

Introduction

About 35 years ago, the present author wrote in the introduction of a paper with the same subject: *More than 50 years after the formulation of quantum mechanics, the interpretation of a pure state (the wave function) still remains contested* [2]. This state of affairs does not seem to have changed much in the meantime. In recent paper entitled *Making better sense of quantum mechanics*, N. D. Mermin [3] writes: *More than ninety years after the invention of quantum mechanics, we find ourselves in a strange situation. Quantum mechanics works. Indeed, no theory of physics has ever had such spectacular success. Yet despite this unprecedented success there is notorious disagreement about ... The sentence fades away because it is not so easy to say what the disagreement actually is about. Everybody who has learned quantum mechanics agrees how to use it. 'Shut up and calculate!' There is no ambiguity, no confusion, and spectacular success. What we lack is any consensus about what one is actually talking about as one uses quantum mechanics. There is an unprecedented gap between the abstract terms in which the theory is couched and the phenomena the theory enables us so well to account for. We do not understand the meaning of this strange conceptual apparatus that each of us uses so effectively to deal with our world.*

A popular scientific account on this topic (*Interpretations of quantum mechanics*, English Wikipedia, April 2, 2020) lists more than 16 different interpretations: Copenhagen interpretation, quantum information theories, relational quantum mechanics, quantum Bayesianism (QBism), many worlds interpretation, consistent histories, ensemble or statistical interpretation, De Broglie-Bohm theory, quantum Darwinism, transactional interpretation, stochastic mechanics, objective collapse theories, consciousness causes collapse (von Neumann-Wigner interpretation), quantum logic, modal interpretations, many-minds interpretation, ... We omit a few approaches being not mere interpretations but extensions or alterations of the quantum mechanical formalism.

Although not all of these interpretations can be taken equally seriously, this colourful bouquet of coexisting opinions that partially contradict one another clearly illustrates the prevailing confusion. Its origin dates back to the early days of quantum mechanics when, after the formulation of wave mechanics by Schrödinger, an interpretation of the puzzling wave function was searched. Such an interpretation was given by Max Born already in 1926 [4]: if the wave function $\psi(\mathbf{x})$ of a particle is developed into a sum of eigenfunctions ϕ_n of a discrete observable, $\psi(\mathbf{x}) = \sum_n c_n \phi_n(\mathbf{x})$, then $|c_n|^2$ is the probability to measure the corresponding eigenvalue of the observable. Consequently, the quantity $|\psi(\mathbf{x})|^2$ is the probability density to find the particle at position \mathbf{x} . Suitable generalizations of these findings are the uncontested basis for the practical use of quantum mechanics. The question about the consequences for the epistemological status of the quantum mechanical formalism was posed by Born already in his 1926 paper. He understood these probabilities as objective and quantum mechanics as indeterministic. This point of view is the *Statistical Interpretation* [of quantum mechanics], *according to which a pure state (and hence also a general state) provides a description of certain statistical properties of an ensemble of similarly prepared systems, but need not provide a complete description of an individual system* [5, p. 360]. Advocates of this position are, besides Born, for

instance Einstein [1], Schrödinger [6, 7, 8], Popper [9], Blokhintsev [10], Ballentine [5].

This probabilistic character of physical predictions was hard to accept for many eminent physicists and a second view on the epistemological character of pure quantum states was developed and advocated by Bohr, Heisenberg, Pauli, and others. This view may be summarized under the heading *Copenhagen interpretation*¹, which asserts that a pure state provides a complete and exhaustive description of an individual system (e.g., an electron) [5, p. 360]. But it is shown in the present work that this is an epistemological misconception of the notion 'state' entailing numerous logical and interpretatory difficulties as 'Schrödinger's cat paradox', 'fundamental measurement problems', 'the collapse of the wave function if a conscious observer takes notice of the result of a performed measurement', ... to note the most notorious. The majority of the interpretations listed above accept the Copenhagen dogma and are attempts to cope with the resulting problems.

All these difficulties disappear at once if the statistical interpretation is adopted. Nevertheless and notwithstanding that it has prominent advocates and is everyday practice among physicists, the statistical interpretation is often considered as one amongst other, equally plausible, interpretations or even as an incomplete description of physical reality. It is the aim of the present work to show that the statistical interpretation is epistemologically well founded, free of internal inconsistencies, and needs no additional epistemological complements as e.g. Bohr's complementarity doctrine. There is much more to say and to understand than just "Shut up and calculate!".

The inquiry starts with a reflection on the nature of scientific statements (Secs. 2.1 – 2.3). Science is considered as a special and important field of human experience, its statements and assertions being characterized by their objectivity and universality. Science serves mankind to characterize, foresee, and master natural processes and is the basis of engineering and technology. As an immediate consequence of universality, physical statements do not primarily address single events (although single events may be analyzed with scientific methods), but infinite classes of similar events. The most general statement about such a class is a probability statement with respect to the single representatives of the class. Experiments are an important tool to test physical theories. To perform such an experiment, an individual system must be prepared in a *state* representing specific initial or boundary conditions, and *observables* representing physical properties of the system must be measured. The results of such measurements can then be confronted with theoretical predictions.

This structure may be cast into a general formal scheme (GFS) of quantitative physical theories (Sec. 2.4). The GFS contains states and observables as generic terms. For a state s and an observable A of a given physical system, the GFS provides the probability prediction $P_{A,s}(\sigma)$ to find the measured value of A in the interval σ . A deterministic description is included in this scheme by probability measures concentrated onto one point for all observables. The results of corresponding experiments are a random test for the asserted probability distribution. In the framework of the GFS, it remains undetermined how states and observables are mathematically represented, and how probability predictions are calculated. In Sec. 2.5, quantum and classical theories are described as formal schemes within the framework of the GFS. States and observables have, therefore, the same epistemological status in quantum as in classical theories. Quantum and classical theories differ in the mathematical representations of states and observables and, of course, in that they may provide different predictions for the same physical situation.

In the GFS, the real function $f(A)$ of an observable A can be defined by the relationship $P_{f(X)}(\sigma) = P_X(f^{-1}(\sigma))$ valid between the stochastic variables X and $f(X)$ for any interval σ . It is shown in Chap. 3, that this basic relationship rules out hidden variables theories that consider pure quantum states as superpositions of underlying dispersion free states providing sharp predictions for *all* observables of the system (an exception being quantum theories with 2-dimensional Hilbert space) [12]. The crucial point is that the noncommutative observable algebra of a quantum mechanical system cannot be mapped on the commuta-

¹This is clearly a rather summary description of the historical development. For a detailed account see [11].

tive observable algebra of such hidden variables theories (except the case of spin $1/2$) even if the mapping is required only for commutative subalgebras of the quantum system. Contributions to hidden variables theories due to von Neumann, Kochen and Specker, Bell, and Mermin are presented and discussed.

In Chap. 4, the characteristic features of state preparation and observable measurement are illustrated with a quantum mechanical model of the Stern-Gerlach experiment, this model being solved and discussed in the Heisenberg picture. In the statistical interpretation, the discussion of the measurement process is straight forward and distinguished by the complete absence of so-called 'fundamental measurement problems' which so severely bother Copenhagen type of interpretations. State preparation can consistently be described here as a separation and filtering process on ensembles. This discussion is especially illustrative with Bohm's version [13, p. 614 – 619], [14] of the Einstein-Podolsky-Rosen thought experiment [15] (see Chap. 5).

The positions of objectivity and realism are advocated here for the description of physical processes and it is shown that no paradoxes or contradictions result within the statistical interpretation. This position, inherent to the statistical interpretation, is nicely described by Mermin [16, p. 804]: *We wish to entertain the heretical view that the results of an experiment are not brought into being by the act of measurement itself. This heresy takes the state vector to describe an ensemble of systems and maintains that in each individual member of that ensemble every observable does indeed have a definite value, which the measurement merely reveals when carried out on that particular individual system. The quantum-mechanical rules, applied to a given state, give the statistics obeyed by those definite values in the ensemble described by that state. The uncertainty principle is not a restriction on the ability of observables to possess values in individual systems, but a limitation on the kinds of ensembles of individual systems it is possible to prepare If two observables fail to commute, then the uncertainty principle cannot prohibit both from having definite values in an individual system. It merely insists that it is impossible to prepare an ensemble of systems in which the values of neither observable fluctuate from one individual system to another.*

In Chap. 6, Bell's inequality and consequences that may (or should not) be drawn from it are discussed. With a view of Bohm's version of the Einstein-Podolsky-Rosen thought experiment, Bell develops, from seemingly very natural assumptions, a classical, local, hidden variables model for spin correlations [17] from which he derives an inequality which is violated by certain quantum mechanical correlations. He concludes that no local hidden variables model can reproduce all quantum correlations in Bohm's setting (Bell's theorem). Unfortunately, he expanded this statement later on to the erroneous assertion that *quantum mechanics* is nonlocal. But the theorem of Kochen and Specker [12] indicates that *no* such hidden variables model exists for Bohm's setting, the underlying Hilbert space being 4-dimensional. This was later shown explicitly by Mermin ([16, p. 809] and Sec. 3.4 of this work). However, Mermin did not draw the evident conclusion that Bell's model cannot describe the system of two spin- $\frac{1}{2}$ particles, contrary to Bell's assumption, because there is no such model. Bell's inequality does, therefore, not concern the spin correlations of Bohm's setting, and its application to the quantum mechanical spin correlations is futile. There is no logical foundation for Bell's theorem, and nonlocality is not the real issue. Two papers upholding that there is a problem with quantum mechanics and locality are critically examined (Secs. 6.2 and 6.3).

The noncommutativity of the quantum mechanical operator algebra is the decisive feature that renders classical hidden variables models impossible and lies at the root of the statistical character of quantum mechanical predictions. It is important to understand the origine of this feature. A contribution to this issue is presented in Chap. 7 where the physics of a free, nonrelativistic particle is developed from Galilei invariance within the quantum mechanical formalism. It is shown that there is a close connection between the noncommutative Lie algebras of unitary ray representations of the Galilei group and the observables 'position', 'momentum', 'kinetic energy', and 'angular momentum' of a particle, which explains the noncommutativity of this observable set. Clearly, this is no new insight, but an important issue in the context of the present work. The presentation given here is comparatively elementary and reveals the origin of noncommutativity of the observable algebra of an important physical concept.

Chapter 2

The general scheme of quantitative physical theories

In this chapter, some formal structures characteristic for the physical description of nature are reviewed and some basic notions and distinctions are established. It is appropriate to sketch here the epistemological position on which this work is based. Physical statements and assertions are considered as the symbolic representation of a particular segment of human experience, namely the part considered to be universally valid, by means of the language faculty proper to mankind. This peculiar part of human experience and its description are characterized more precisely in the following sections. We take the view of realism and objectivity stating that the material world exists and that physical processes take place irrespective whether a conscious being notices them. The role of the scientist is to perceive physical events, to produce such events in experiments, and to analyze and structure them by inventing an appropriate terminology and physical theories that are discussed and validated within the scientific community. We consider physical theories and models together with the notions they embody as free human creations being restricted, first of all, by the requirement of logical consistency and, in addition, by the general characteristics of physical statements, to be specified below, and, of course, by the aim that they provide statements about the physical world that can be falsified *i.e.* which are not mere tautologies, and turn out to be true. We think that the position of realism and objectivity is indispensable if physics is understood as the basis of engineering and technology. It is the aim of the present work to show that this position is logically consistent but entails that physical predictions are basically statistical.

2.1 Characteristics of physical statements

A characteristic trait of theoretical physical thinking consists of interpreting an individual observation as a special instance of a general law. As an example, take an observation like "*There was a thunderstorm in Zürich on August 15, 2018, at 5 p.m.*" For a physicist, neither place nor time are considered essential for its occurrence but rather meteorological conditions that, if realized, generate thunderstorms. Physicists expect nature to follow causal laws that can be formulated without explicit reference to geographical or historical circumstances. The causes can be brought about in principle by everybody, everywhere, at every time and always entail the same consequences. As the validity of causal laws is asserted for all situations where they apply, these laws have the logical form of *universal implications* [18]: "*At every place, for every time, and for every observer $s(S) \rightarrow p$* ". s denotes the causes, initial and boundary conditions, imposed on the system S , or the *state* into which the system is prepared, p the asserted consequences. A physical system is characterized by its set of observables or, vice versa, an observable denotes a physical property of the system. For instance: 'mass', 'position', 'momentum', 'energy', 'angular momentum', 'spin' are

the observables of a point mass. Physical properties are typically quantitative, their values being given as multiples of certain standard units.

Processes that follow universal causal laws are obviously important because their result can be predicted to a certain extent, and these processes offer so the possibility of human command on nature. A consequence of universality is the possibility to realize a physical system and specific initial and boundary conditions in an *experiment*. If the result of an experiment is once accepted by the scientific community, its validity is asserted also in future up to possible revisions in the course of scientific progress. Natural science owes much of its dynamics and effectiveness to the possibility to perform experiments. Clearly, important physical insights can be, and were, gained in contexts where a planned realization of causes exceeds human faculty, and where it must be relied on mere observations as *e.g.* in astronomy. However, experimental settings and the prediction of their results by physical theories are the generic case.

An experiment that can be confronted with a theoretical prediction is divided into two parts in a natural way. First of all, the state s of the system S must be *prepared*. In a second step, the value of an observable of the system is *measured*. This value can then be compared with the prediction p ; usually many repetitions of the experiment are necessary to arrive at a decisive conclusion. The preparation of the state and the subsequent measurement of an observable value are logically distinct parts of an experiment devised to test theoretical predictions. State preparation points towards the future whereas a measurement reveals the value of an observable at the very moment of its measurement. The value of the observable may be changed by the measurement; its value after the measurement is of no importance for the test of theoretical predictions. The system may even cease to exist after the measurement *e.g.* when a particle is absorbed by a detector. State preparation and the measurement of an observable value are both necessary and distinct parts of an experiment and should not be merged into one diffuse notion 'measurement'¹. The insight that state preparations and observable measurements serve different purposes is due to K. R. Popper [9, p. 25].

The universality of physical statements has a direct and important consequence: such a statement is not only asserted for the actual physical situation under consideration but is considered to be equally valid for every situation of the same kind. Physical statements refer, therefore, primarily to whole classes of such situations and to an actual situation only as far as it is a member of such a class. All members of a class are identical with respect to physical predictions. They have no individual properties in this regard. Preparations and measurements can be carried out following universal and generally valid instructions. Physical systems can be realized by using universal instructions such as *e.g.* prescriptions for the construction of an apparatus, for purification of copper, for the generation of a jet of electrons. Proper names are not really necessary in these instructions; occurring proper names can and should be removed.

The history of metrology [19] provides a striking illustration for this process. Until the middle of the 20th century, the units of time, length, and mass were based on the rotation of the earth and on artefacts (primary metre, primary kilogram). In 1960, the metre was redefined as a multiple of the wavelength of a spectral line of krypton-86. In 1967, the second was redefined as $9'192'631'770$ times the period of the groundstate hyperfine splitting frequency of the caesium-133 atom. In 1983, the definition of the metre was finally traced back to the definition of the second by giving the speed of light in vacuum the exact value of $299'792'458$ m/s. This process was brought to a close in May 2019, when the values of the seven base units metre (m), second (s), kilogram (kg), ampere (A), kelvin (K), mole (mol), and candela (cd) are determined by giving seven physical constants ($\Delta\nu(^{133}\text{Cs})_{\text{hfs}}$, speed of light c , Planck constant h , elementary electrical charge e , Boltzmann constant k , Avogadro constant N_A , luminous efficacy K_{cd}) exact numerical values [20], and by the identification of experiments to produce the base units from these given values of the physical constants.

Physical statements are universal and apply to classes. Clearly, physical statements about individuals are

¹The disregard of this distinction has much obscured the discussion of the measurement process and is one of the reasons why 'fundamental measurement problems' emerged. The other, more fundamental reason is the assertion that a pure state ('the wave function') provides an exhaustive description of an *individual* physical system.

possible but their individuality does not matter. In physical cosmology, the universe is treated as an example of a certain class of physical systems. It has no physical consequences that our universe is the only accessible specimen. In scientific explorations of animate beings, their individual properties, their individual history, is not in the focus but typically figure as disturbances impairing the reproducibility of the results, disturbances which must be reduced as much as possible by suitable standardizations.

Universality and the abstraction from individuality make physical descriptions indirect and mediated in a specific manner: physical statements and the occurring expressions and terms refer directly to the abstract class of 'physical systems of the same kind' or 'experimental arrangements', to 'preparation instructions', and to 'measurement instructions'. They refer to a concretely realized physical system with its concretely realized preparation and measurement only in so far as it is a member of the occurring class. This opens the possibility for probabilistic statements (*i.e.* for statements about the probability of different possible outcomes), for causal statements that are not deterministic. The most general prediction for an experiment is the prediction how probable the different outcomes of measurements performed under identical conditions will be. Causal laws can then be expected only for these probability predictions not for measurement values. Deterministic predictions of measurement values are, nevertheless, included in this scheme. The probability measure of every observable is then concentrated onto a single point. It is a subtle but important point that physical theories do not provide predictions of objective properties of individual systems, but provide objective predictions of probability distributions of such properties.

The above characterization of physical statements as universal laws may be specified by stating that physical space is homogeneous and isotropic and physical time is homogeneous and that it is possible to formulate physical laws in such a way that they are invariant with respect to these symmetries of physical space and time. Thereby it is assumed that faithful standards for space and time measurements are available, these standards being used to define reference frames. Two such reference frames are physically equivalent if they can be made congruent by time translations and by space translations and rotations. In addition, two reference frames moving uniformly towards one another are also physically equivalent (Galilei and Lorentz invariance). It is revealing that assuming Galilei invariance allows to completely derive the physics of a nonrelativistic, free, quantum mechanical mass point (see Chap. 7).

2.2 Interpretation of the notion 'probability'

The axiomatic basis of probability theory given by Kolmogorov (see Appendix A) is widely accepted. However, the interpretation of the notion 'probability' is much more controversial. H. Rott and W. Lübbe [21, p. 605] define probability as *Bezeichnung für den Grad der Möglichkeit des Eintritts eines Ereignisses oder den Grad der Gewissheit oder der Glaubwürdigkeit einer Aussage*². Therefore, probability can be understood in two fundamentally different ways: *objective interpretations* assign probabilities to events as an objective measure of the frequency of their occurring, whereas *subjective interpretations* assign probabilities to statements or assertions as a measure of the subjective knowledge or belief that a statement is true. In the realms of natural sciences and technology, one mostly deals with objective probabilities. R. von Mises [22] stressed the fundamental difference between objective and subjective probabilities: from premises about degrees of belief we can never get a conclusion about the frequency of events on a mathematical or logical way [9, p. 29].

There are mainly three different kinds of objective interpretations of probability [9, p. 30]:

- i. The *classical interpretation* (J. Bernoulli, A. de Moivre, P. S. Laplace) takes probability as the proportion of the number of favourable to the number of equally possible cases, the latter being conjectured from symmetry considerations, and *e.g.* realized by throws of a regular die.

²term characterizing the degree of the possibility that an event will occur or the degree of certainty or credibility of a statement

ii. The *frequency interpretation* (J. Venn, R. von Mises, H. Reichenbach) according to which probabilities are limits of the relative frequencies with which certain attributes occur in an infinite sequence of random experiments.

The classical and the frequency interpretation have been severely criticized. The classical definition of probability is circular: 'equally possible' is much the same as 'equally probable'. In addition, it can hardly handle *anything like unequally possible cases such as playing with a loaded die; and it succumbs, like the subjective interpretation, to von Mises's criticism: there is no logical or mathematical bridge ... which leads from premises about possibilities to statistical conclusions about relative frequencies* [9, p. 31]. To decide whether a sequence of events is random is a second, central problem of the frequency interpretation. An other difficulty is that limiting values of relative frequencies apply uniquely to infinite sequences. In practice, however, there are only finite sequences available. Linking probabilities to infinite sequences prevents, furthermore, that the notion of probability can be applied to single events, the situation typical for causal but indeterministic behaviour.

In view of these shortcomings, K. Popper proposed

iii. the *propensity interpretation* of probability, where probability is understood as the propensity or proneness of an experimental arrangement, *i.e.* of a physical system and its state preparation, to produce a certain result in a single experiment. The propensity interpretation *differs from the ... frequency interpretation only in this — that it considers the probability as a characteristic property of the experimental arrangement rather than as a property of a sequence. The main point of this change is that we now take as fundamental the probability of the result of a single experiment with respect to its conditions, rather than the frequency of results in a sequence of experiments. Admittedly, if we wish to test a probability statement, we have to test an experimental sequence. But now the probability statement is not a statement about this sequence: it is a statement about certain properties of the experimental conditions, of the experimental set-up* [23, p. 67].

Our use of probability in Sec. 2.1 corresponds to the propensity interpretation with an emphasis on it that the notion 'experimental arrangement' refers to the mentioned abstract classes of physical systems and to preparation and measurement instructions. We conclude this section with the remark that there is no logical or mathematical bridge from objective probabilities to knowledge or belief either. A maxim, or axiom of applicability, is needed for the practical use of probability theory *e.g.* in order to decide if a given statistical sample is compatible with a theoretical probability distribution. This maxim may be formulated as follows: act as if events with sufficiently small probabilities would never take place. What is accepted as 'sufficiently small probability' is conventional and depends on the actual circumstances and on the preferences of the person that decides³. It is in this sense that a subject capable to take decisions enters the scene. We maintain that the probability distributions in physics should be considered as properties of the experimental arrangements and not of a conscious subject taking notice of them. This rightly accounts for their asserted universality which is the basis of their applications in engineering and technology.

2.3 Epistemological status of observables and states, attributes of concrete material systems

It was stressed in Sec. 2.1 that physical predictions refer to infinite abstract classes of equally prepared, equal systems and to a concrete, material system only as a member of such a class. Consequently, theoretical notions and terms used to formulate these predictions, namely observables, states, and probability measures⁴, are attributes of the description of such classes, not of their concrete material realizations. A state is a theor-

³Markus Fierz, private communication.

⁴In discussions concerning the interpretation of the quantummechanical formalism, emphasis is usually put only upon the notion state ('the wave function'). But the complete formalism should be considered.

tical notion, a symbol for initial and boundary conditions or for a preparation instruction in an experimental arrangement. The statement "*The material system S is in the state s* " is short for "*The material system S has undergone a preparation according to the instruction s* ". The supposition of the Copenhagen interpretation that a pure state is a property of this concrete system is an epistemological misconception leading to the numerous 'fundamental problems' of the Copenhagen interpretation completely absent in the statistical interpretation (see Chap. 4 and 5).

But what are the attributes of a concrete system? Attributes of concrete, material systems are the possible results of measurements on this system *i.e.* values of its observables. 'Velocity of 10 m/s' is an attribute that can be assigned to a concrete particle; "*This particle has a velocity of 10 m/s.*" is a meaningful sentence. 'Velocity' is not an attribute of a concrete particle. What could be the meaning of a sentence like "*This particle has velocity.*"? That it is not at rest? The sentence is then short for "*This particle has a velocity different from zero.*" — again a statement about measurement values of an observable. Or the sentence "*Velocity represents an observable that can be measured on a particle.*"? But this is a statement about the *theoretical notion* 'particle' not on 'this particle here', a real object. Likewise: 'colour' is not an attribute of a concretely present body, but 'red', 'green', 'blue' are.

We clearly advocate the position of objectivity and realism. Observables denote physical properties characterizing a theoretical model that is applied to a concrete, material system. By definition, all these properties can be measured at the material system to which the model applies, revealing so its properties namely the values of its observables. The result of an appropriate, error free measurement of an observable on an individual system is the observable value of this system at the instant of measurement, an objective property of the system alone. With this definition, all observable values are always definite. It is, however, in general not possible to simultaneously predict all these sharp values, as shown, for instance, in quantum mechanics.

Remember, in contrast, the position of Bohr that the result of an individual measurement must be considered as the joint product of the (microscopic) system under observation and the (according to this view necessarily macroscopic) measuring device (see N. Bohr in [1, p. 210]). A measurement value is then no expression of an objective property of the individual system alone. For Bohr and his followers, the behaviour of quantum objects can only be described in terms of complementary classical pictures, allegedly the only ones that can be grasped by human experience⁵. This position may be understood as the reaction of a generation of physicists, educated in the spirit of classical physics and faced with the puzzling quantum phenomena. But this is clearly not the position taken in this work: we consider the quantum mechanical description as completely understandable in itself and as the fundamentally correct one, and classical mechanics only as an important limiting case.

2.4 The general formal scheme of quantitative physical theories (GFS)

In the remaining part of this chapter, the notion 'physical system' refers to the mentioned class of 'equally prepared, equal systems' if not noted otherwise. The considerations in the previous sections may be formalized in the following manner: basic objects of causal, quantitative physical theories are *observables* A, B, \dots corresponding to the measurable quantities of the physical system in question, and *states* s_1, s_2, \dots corresponding to the preparation of initial or boundary conditions of the system. The most general quantitative prediction for a physical system is a statistical prediction for the observable values measured on individual, material realisations of the system subjected to the preparation which corresponds to the state s . A theory assigns, therefore, a *real probability measure* $P_{A,s}$ to an observable A and a state s . $P_{A,s}(\sigma)$ indicates the probability with which the *measured value* $m(A)$ of the observable A falls into the real Borel

⁵"...an essential element of ambiguity is involved in ascribing conventional physical attributes to atomic objects, as is at once evident in the dilemma regarding the corpuscular and wave properties of electrons and photons, where we have to do with contrasting pictures, each referring to an essential aspect of empirical evidence" (N. Bohr in [1, p. 210]).

set σ if an individual material representative of the system is prepared according to the state s . Or, the values $m(A)_1, m(A)_2, \dots, m(A)_n$ of the observable A measured on individual physical systems prepared according to the preparation instruction s form a random test of the probability measure $P_{A,s}$.

In terms of probability theory, outlined in Appendix A, this *general formal scheme of quantitative physical theories* (GFS) assigns a probability space $(\Omega_A, \Sigma_A, P_{A,s})$ to an observable A and a state s . The sample space Ω_A is the *spectrum* of the observable A , the set of all possible measurement values of A . The spectrum is assumed to be a Borel set. Ω_A is an intrinsic property of the observable, which is, in particular, independent of the state into which a system is prepared. The observable A is represented by the stochastic variable mapping a point ω of the sample space Ω_A onto itself.

It is advantageous to embed $(\Omega_A, \Sigma_A, P_{A,s})$ into the probability space $(\mathbb{R}, \mathcal{B}, P_{A,s})$ where \mathcal{B} is the Borel σ -field (see Appendix A) and $P_{A,s}(\sigma)$, $\sigma \in \mathcal{B}$, is now equal to the former $P_{A,s}(\sigma \cap \Omega_A)$, $\sigma \cap \Omega_A \in \Sigma_A$. The spectrum Ω_A is then characterized by the condition that $\overline{\Omega_A}$ is the union of all Borel sets σ with $P_{A,s}(\sigma) = 0$ for all states s . In the sequel, only this embedding is considered. The following separation postulates serve the economy of description:

- (SI) Observables separate states. Two states s_1 and s_2 may be identified, $s_1 = s_2$, if $P_{A,s_1}(\sigma) = P_{A,s_2}(\sigma)$ for all observables A and for all Borel sets σ .
- (SII) States separate observables. Two observables A and B may be identified, $A = B$, if $P_{A,s}(\sigma) = P_{B,s}(\sigma)$ for all states s and for all Borel sets σ .

This general formal scheme together with the separation postulates corresponds to the axioms I and II of Mackey's axiomatics of quantum mechanics [24, p. 63].

In addition, an operational meaning of the identity $A = B$ must be provided for individual measurements: if A is measured on an individual system yielding the measured value $m(A) = a$, then this is also a measurement of $B = A$ with the same measured value $m(B) = a$. This is the only reasonable definition if the value of an observable is considered as an objective property of the individual system. If, on the contrary, $m(B) = b \neq a$ is assumed for individual measurements, then it must be explained why many independent realisations of B -measurements yield a random sample of the probability distribution $P_{A,s}$.

This formal description of a physical theory may appear rather general. It entails, nevertheless, some additional structure due to the properties of Borel sets, probability measures, and stochastic variables.

First: the function $f(A)$ of an observable A is already defined on the level of the probability spaces (Mackey's axiom III). Let f be a Borel function *i.e.* the full inverse $f^{-1}(\sigma)$ of a Borel σ is again a Borel set. Then it is

$$P_{f(A),s}(\sigma) = P_{A,s}(f^{-1}(\sigma)) \quad (2.1)$$

for every state s and for every Borel set σ . On the operational level of single measurements, this means that a measurement of the observable A yielding the value $m(A) = a$ is also a measurement of the observable $f(A)$ with the measurement value $f(a)$ or

$$m(f(A)) = f(m(A)) \quad (2.2)$$

In accordance with this relationship, the *expectation value of an observable $f(A)$, given the state s* , equals

$$E_s(f(A)) \equiv \langle f(A) \rangle_s = \int_{-\infty}^{+\infty} f(x) dF_{A,s}(x) \quad (2.3)$$

where $F_{A,s}$ is the *distribution function*, $F_{A,s}(x) \equiv P_{A,s}((-\infty, x])$ (see Eqns. (A.4), (A.5), and (A.8)).

For the unit observable I , it is always $m(I) = 1$. Therefore, $\Omega_I = \{1\}$, $F_{I,s}(x) = \Theta(x - 1)$ for all states s and, thus, $dF_{I,s}(x) = \delta(x - 1) dx$, and

$$E_s(I) = 1 \quad (2.4)$$

This result may also be derived directly from Eq. (2.3) with arbitrary observables by setting $f(x) \equiv 1$.

The probability measure $P_{A,s}(\sigma)$ can be recovered as the expectation value of the characteristic function $\chi_\sigma(A)$ in the state s

$$P_{A,s}(\sigma) = E_s(\chi_\sigma(A)) \quad (2.5)$$

(see Eq. (A.19)). In a fully developed theory, it thus suffices to assign mathematical objects to states, observables, and to the expectation value of an observable in a given state.

The fact that real functions of observables are, in the GFS, defined already on the level of the probability measures is important for the appropriate formulation and discussion of the problem of hidden variables in quantum mechanics (see Sec. 3.2).

Secondly: on the level of probability measures, convex linear combinations of states can be defined (axiom IV of Mackey). Let s_1, s_2, \dots be a countable sequence of states of the system and p_1, p_2, \dots a countable sequence of real numbers such that $p_n > 0$ and $\sum_n p_n = 1$. The convex linear combination

$$P_{A,s} \equiv \sum_n p_n P_{A,s_n} \quad (2.6)$$

is then again a probability measure for every observable A and defines a state s as a linear combination of states s_n which may formally be denoted by $s = \sum_n p_n s_n$. On the operational level, this means that every random sample $m(A)_1, m(A)_2, \dots, m(A)_n$ measured on systems prepared in the state s can be subdivided into samples belonging to the same observable measured on systems prepared in the states s_n . The numbers p_n indicate the relative frequencies the measured values fall into the subsamples belonging to s_n . A *pure* state s is defined by the condition that every decomposition $s = p_1 s_1 + p_2 s_2$, $p_1 + p_2 = 1$, implies that either $p_1 = 0$ or $p_2 = 0$. It may not always be meaningful to interpret the convex linear combination s in the sense that the system is prepared into the state s_n with probability p_n . The preparation instruction corresponding to s may have no relation to the preparation instructions for the states s_n (consider for instance Gibbs states!).

The GFS provides only few additional structures except the relationships (2.1) – (2.6). In particular, the sum and the product of two observables A and B can be defined within the GFS only if A and B are Borel functions $f(C)$ and $g(C)$ of a third observable C . It is then $A + B = (f + g)(C)$ and $A \cdot B = (f \cdot g)(C)$ in accordance with the relationships $m(A + B) = m(A) + m(B)$ and $m(A \cdot B) = m(A) m(B)$ for measured values. These operational definitions make sense only if the values of A and B can be simultaneously measured irrespective of the state into which the system was prepared. It is well known from quantum mechanics that this cannot generally be assumed. However, the addition of a real constant r to an observable A is a special instance of Eq. (2.1) with $f(x) = r + x$. This yields $m(rI + A) = r + m(A)$ for the measurement values, and $F_{rI+A,s}(x) = F_{A,s}(x - r)$ for the distribution function (A.4). The expectation value of $aA + bB$ is linear if A and B are functions of C :

$$E_s(aA + bB) = aE_s(A) + bE_s(B) \quad (2.7)$$

as follows from Eq. (2.3).

The given, purely formal description of physical theories is by no means exhaustive. An understanding of physical theories as black boxes spitting out probability measures would be far too narrow. A theory interpreted this way could only and strictly by applied to the experiment for which it was designed. There would be no foundation for an extension to other, even very similar, situations. It is a decisive strength of physical theories that, besides the prognoses for the group of experiments they are designed for, they also give hints how to treat different physical problems or even indicate principles generally valid for all physical theories. The GFS is, nevertheless, a suitable scheme to discuss some of the so-called interpretation problems of quantum mechanics.

2.5 Quantum and classical theories as examples of the GFS

The GFS provides observables and states only on a symbolic level and contains no relationships between observables except that an observable may be a real function of an other one. In order to establish additional relationships between observables and to be able to compute probability measures, a developed theory is needed. In such a theory, states and observables correspond to mathematical objects and there exist instructions how to calculate the probability measure $P_{A,s}$ or the expectation values $E_s(A)$. Quantum and classical theories are examples [24].

Quantum mechanics *in its statistical interpretation* clearly fits into the GFS. An observable A is represented by a *self-adjoint operator* $\hat{A} = \hat{A}^\dagger$ acting on a Hilbertspace \mathcal{H} , the state s by a *state operator* ρ_s , the expectation value of an observable A in the state s is given by $\langle A \rangle_s = \text{Tr}(\rho_s \hat{A})$.⁶ A pure state is represented by a projection operator Q on a 1-dimensional subspace of \mathcal{H} , and the state operator ρ_s by a convex linear combination of such projections: $\rho_s = \sum_n p_n Q_n$, $p_n > 0$, $\sum_n p_n = 1$. This linear combination should not be confounded with the spectral decomposition of ρ_s ; the projections Q_n do not necessarily all commute. State operators are nonnegative, self-adjoint operators with a discrete spectrum and trace, $\text{Tr}(\rho)$, equal to 1.

The distribution function $F_{A,s}$ is given by

$$F_{A,s}(x) = \text{Tr}(\rho_s E_{\hat{A}}(x)) \quad (2.8)$$

where $E_{\hat{A}}(x)$ denotes the spectral decomposition of the operator \hat{A} . As a consequence of Eq. (2.3) and of the linearity of the trace, it is

$$\begin{aligned} \langle f(A) \rangle_s &\equiv E_s(f(A)) = \int_{-\infty}^{\infty} f(x) dF_{A,s}(x) = \int_{-\infty}^{\infty} f(x) d \text{Tr}(\rho_s E_{\hat{A}}(x)) \\ &= \text{Tr}\left(\rho_s \int_{-\infty}^{\infty} f(x) dE_{\hat{A}}(x)\right) = \text{Tr}(\rho_s f(\hat{A})) \end{aligned} \quad (2.9)$$

for all functions $f(A)$ for which the integrals exist.

Eq. (2.9) entails that the observable $f(A)$ is represented by the operator $f(\hat{A})$ if A is represented by \hat{A} . Choose now $f(x) = \chi_{\bar{\Omega}_A}(x)$, the characteristic function of the complement of the spectrum of observable A . It is $\langle \chi_{\bar{\Omega}_A}(A) \rangle_s = 0$ for all states s and thus $\chi_{\bar{\Omega}_A}(\hat{A}) = \hat{0}$, the null operator, by the separation postulate (SII). The spectrum of the operator \hat{A} is, therefore, equal to the spectrum of the observable A [26, Appendix 1].

Two observables A and B whose quantum mechanical representatives \hat{A} and \hat{B} do not commute cannot be simultaneously sharply predicted for all pure states. A *classical theory* is, on the other hand, characterized by the condition that all states can be represented as convex linear combinations of *atomic states* in which all observables attain sharp values. The set of all atomic states is the *phase space* Λ . It is assumed that Λ is a measure space. An atomic state is then represented by a probability measure concentrated on one point $\lambda_0 \in \Lambda$, which may symbolically be denoted by $\delta(\lambda - \lambda_0)$. A general state s is represented by a general probability measure μ_s on the phase space. The decomposition of a classical general state into atomic states is always unique in contrast to the corresponding situation in quantum mechanics. An observable A is represented by a real, measurable function \tilde{A} of atomic states, $\tilde{A} : \lambda \in \Lambda \rightarrow \tilde{A}(\lambda) \in \mathbb{R}$. A typical example of a classical theory is a system of N identical point masses, where the phase space Λ equals \mathbb{R}^{6N} with the volume element $1/(N! h^{3N}) d^{3N}x d^{3N}p$, and the observables are real functions which are symmetric in the indices on this phase space.

For a system prepared in an atomic state, it is $E_{\delta_{\lambda-\lambda_0}}(A) = \tilde{A}(\lambda_0)$ and $E_{\delta_{\lambda-\lambda_0}}(A^2) = \tilde{A}(\lambda_0)^2$ for all observables A i.e. the result of the measurement of any observable is exactly predicted. The classical

⁶For a concise introduction into the mathematical formalism of quantum mechanics see e.g. [25, Chap. 2].

expectation value of the observable $f(A)$ in a general state s is given by

$$E_s(f(A)) = \int_{\Lambda} f(\tilde{A}) d\mu_s(\lambda) \quad (2.10)$$

and the real probability measure $P_{A,s}(\cdot)$ by

$$P_{A,s}(\sigma) = \int_{\Lambda} \chi_{\sigma}(\tilde{A}) d\mu_s(\lambda) = \mu_s(\tilde{A}^{-1}(\sigma)) \quad (2.11)$$

$\tilde{A}^{-1}(\sigma)$ being the full inverse image of σ under \tilde{A} . The set $\{\tilde{A}(\lambda) \mid \lambda \in \Lambda\}$ corresponds to the spectrum of the observable A . $f(A)$ is represented by $f(\tilde{A}(\lambda))$ because of the identity $\tilde{A}^{-1}(f^{-1}(\sigma)) = (f(\tilde{A}))^{-1}(\sigma)$.

Although quantum theories and classical theories both fit into the GFS, they possess quite different structures. In classical theories, the product $\tilde{A}\tilde{B}$ of the two representatives \tilde{A} and \tilde{B} of the observables A and B is, in general, a real measurable function on the phase space and represents the product observable. In quantum mechanics, the product of two self-adjoint operators \hat{A} , \hat{B} is self-adjoint only if they commute, and can, therefore, only in this case be interpreted as a representative of an observable $AB = BA$. In classical theories, simultaneous sharp predictions for all observables are possible, the system then being in an atomic state, and statistical predictions are considered as consequences of an inaccurate state preparation. In a quantum theory, there are no atomic states and the statistical nature of predictions is intrinsic, not reducible to mere inaccuracies of preparation. In a classical theory, there is *one* probability space (Λ, Σ, μ_s) for each state s , and all observables A are represented by random variables \tilde{A} over this one probability space. This is in contrast to the situation in quantum theories, where two observables A and B whose operators \hat{A} and \hat{B} do not commute are, in general, random variables in different probability spaces for a given state s .

At the level of the GFS, a real linear combination $aA + bB$, $a, b \in \mathbb{R}$, of two observables can only be defined if A and B are functions of a third observable C . As a consequence, there is no general additivity of expectation values in the framework of the GFS. This is in contrast to quantum and to classical theories where linear combinations of arbitrary observables can be considered as observables and the expectation values, given by Eq. (2.9) and Eq. (2.10), respectively, are linear. In classical theories, this appears quite natural because the linearity of the expectation value for all states is equivalent to $(a\tilde{A} + b\tilde{B})(\lambda) = a\tilde{A}(\lambda) + b\tilde{B}(\lambda)$ *i.e.* the measurement value of the linear combination of observables is equal to the same linear combination of the individual measurement values. In quantum theories, the linearity of the expectation values, an immediate consequence of its definition (2.9), appears less obvious because the linearity of the corresponding measurement values is not given. A well known example is the spin variable (\mathbf{e}, \mathbf{S}) of a spin- $\frac{1}{2}$ particle which has the eigenvalues $\pm\hbar/2$ for all directions $\mathbf{e} \in \mathbb{R}^3$. The measurement value of the spin variable $(\mathbf{e}_1 + \mathbf{e}_2, \mathbf{S}) / \|\mathbf{e}_1 + \mathbf{e}_2\|$ is thus different from the corresponding linear combination of the measurement values of $(\mathbf{e}_1, \mathbf{S})$ and $(\mathbf{e}_2, \mathbf{S})$ unless the directions \mathbf{e}_1 and \mathbf{e}_2 are parallel or antiparallel.

The meanings of the notions 'state', 'observable', 'observable value' and of the actions 'preparation of a state' and 'measurement of an observable' are identical for quantum mechanics and for classical theories. States, observables, and expectation values are, however, represented by different mathematical objects. The decisive formal difference is that the quantum mechanical representation of observables form a noncommutative algebra whereas their classical representatives all commute. From an epistemological point of view, the essential difference between classical and quantum theories is that the former provides the possibility of simultaneous, sharp predictions for *the values of all observables* whereas the latter only provides predictions for *probability distributions of observable values*. This may appear as a shortcoming of quantum theories since the position of realism upholds that *all* observables of a material system *always* have definite, sharp values. Quantum theories thus seem to give merely an incomplete description. This conviction, most prominently held by Einstein, Podolsky, and Rosen [15] and later on by Bell [27], led to a deepened inquiry into so-called hidden variables theories, a topic discussed in Chap. 3. The main result of this discussion is that a more complete description than the one provided by quantum mechanics does, in general, not exist.

Chapter 3

The problem of hidden variables in quantum mechanics

In the present context, the interest in classical theories lies in their epistemological simplicity: classical systems can always be considered to be in an atomic state where all observables simultaneously are sharply predictable, but it is not always clear in which atomic state. Classical theories directly deal with observable values, the properties of material systems; they appear, in this sense, concrete. In contrast, quantum theories only provide statistical predictions for observable values, a seemingly weaker and more abstract description. It is an interesting question if it is possible, nevertheless, to reproduce the statistical predictions of the quantum mechanical description of a physical system with a classical theory. As a paradigm serves the relationship between thermodynamics and statistical mechanics where the atomic states of the statistical description of the system may be considered as hidden variables, a suitable average over them yielding the measurable thermodynamic quantities.

The notion 'hidden variables' is not uniquely defined. Consequently, different kinds of hidden variables theories have been proposed (for a survey see [28] and [11, Chapt. 7]). The hidden variables theories discussed in this chapter address the concerns of those who admit that the quantum mechanical predictions are correct but feel uncomfortable with their statistical nature and aim at a theoretical framework that allows, together with a pure quantum mechanical state ψ , to predict the measurement results of *all observables* of an *individual* system. Einstein, being a prominent advocate of this position, was not satisfied by quantum mechanics as a statistical theory, although he admitted its consistency and success, and looked for a 'complete description' of a physical system [1, p. 671 – 673], [15].

3.1 Von Neumann's theorem

An early contribution to the question if hidden variables are possible was given in 1932 by J. von Neumann in his book '*Mathematische Grundlagen der Quantenmechanik*' being translated into English in 1955 [29]. Von Neumann concluded that hidden variables are not possible. His authority settled the discussion about hidden variables for a considerable time until J. Bell [30] pointed out that some of the seemingly plausible assumptions of von Neumann are, in fact, not so obvious (an early criticism on similar grounds by Grete Hermann [31, p. 99 – 104] seems to have been entirely ignored).

Von Neumann formulated his no-go theorem for hidden variables within the scope of an axiomatic formulation of quantum mechanics. His essential postulates are (in curly brackets von Neumann's identifications):

- (i) {I} There is a one-to-one correspondance between physical observables A and self-adjoint operators \hat{A} acting on some Hilbertspace [29, p. 313].

- (ii) {I} Let $f(A)$ be a real function of an observable A *i.e.* the relationship (2.2) for the corresponding measurement values holds. Then $f(A)$ is represented by the operator $f(\hat{A})$ [29, p. 313].
- (iii) {II} The sum of several observables $A, B, C \dots$ is represented by the operator $\hat{A} + \hat{B} + \hat{C} + \dots$, whether or not these operators are mutually commutative [29, p. 314].
- (iv) {A'} There exists an expectation function $E(A)$ of observables A , their (real) expectation values with respect to some statistical ensemble. $E(A) \geq 0$ for nonnegative observables (*i.e.* if all measurement values are nonnegative) [29, p. 311].
- (v) {B'} The expectation function is linear

$$E(aA + bB) = aE(A) + bE(B)$$

for arbitrary observables A and B , arbitrary real numbers a and b , and all statistical ensembles with which the expectation values are calculated. [29, p. 311].

From these postulates, von Neumann derives the trace formula $E(A) = \text{Tr}(\rho A)$ where ρ is a nonnegative, self-adjoint operator characteristic of the underlying statistical ensemble but independent of the observable A . In order to identify ρ with the quantum mechanical state operator, the additional condition $E(I) = 1$ is needed, self-evident in the GFS (see Eq. (2.4), but not required by von Neumann.

In a second step, von Neumann shows that the trace formula for expectation functions precludes the existence of dispersion free expectation functions, characterized by the condition $E(A^2) = E(A)^2$ for all observables A , and, hence, the idea of hidden variables stating that a pure quantum state can be considered as a convex linear combination of dispersion free substates. Since von Neumann takes the postulates (i) – (v) as a genuine and defining part of quantum theory, he concludes: *It is therefore not, as is often assumed, a question of a reinterpretation of quantum mechanics, — the present system of quantum mechanics would have to be objectively false, in order that another description of the elementary processes than the statistical one be possible* [29, p. 325].

A summary and critical assessment of von Neumann's proof are given in [5, p. 374 – 375] and in [26] where, in addition, Bell's and, in [26], also Hermann's criticisms of von Neumann's position are provided. We present here the economic and elegant proof given by S. Kochen and E. Specker [12, p. 81]. Their scope, being focussed on the question of hidden variables, is narrower than von Neumann's. Kochen and Specker summarize the conditions that, according to von Neumann, the expectation value $E(\cdot)$ of any self-adjoint operator A in a dispersion free state should fulfill by the following postulates:

$$\begin{aligned}
\text{(I)} \quad & E(I) = 1 \\
\text{(II)} \quad & E(aA) = aE(A) \\
\text{(III)} \quad & E(A^2) = E(A)^2 \\
\text{(IV)} \quad & E(A + B) = E(A) + E(B)
\end{aligned} \tag{3.1}$$

Conditions (I) and (II) follow, in the framework of the GFS, from Eq. (2.3), condition (III) characterizes a dispersion free state, and condition (IV) is assumed to hold for all self-adjoint operators.

Theorem of von Neumann *There is no function $E(\cdot)$ satisfying the conditions (I) — (IV).*

Proof First, Kochen and Specker show the following corollary:

Corollary *In a dispersion free state, any expectation value of a self-adjoint operator lies in its spectrum.*

Proof of the corollary The following simple lemma is needed:

Lemma Assume conditions (I), (II), (III) and

$$(IV') \quad E(A + B) = E(A) + E(B) \quad \text{if} \quad AB = BA \quad (3.2)$$

Then it is $E(AB) = E(A) E(B)$.

Proof of the lemma It is

$$\begin{aligned} E(A)^2 + 2E(A) E(B) + E(B)^2 &= (E(A) + E(B))^2 \\ &= (E(A + B))^2 \\ &= E((A + B)^2) \\ &= E(A^2 + 2AB + B^2) \\ &= E(A^2) + E(2AB) + E(B^2) \\ &= E(A)^2 + 2E(AB) + E(B)^2 \end{aligned} \quad (3.3)$$

and, hence, $E(AB) = E(A) E(B)$.

QED

Assume now that $(A - E(A) I)$ has an inverse B . The two operators commute and it follows from the assumptions of the lemma that

$$\begin{aligned} 1 = E(I) &= E((A - E(A) I) B) = E(A - E(A) I) E(B) \\ &= (E(A) - E(A)) E(B) \\ &= 0 \end{aligned} \quad (3.4)$$

a contradiction! Therefore, $(A - E(A) I)$ has no inverse, and $E(A)$ lies in the spectrum of A or, in the case of a discrete spectrum, is equal to an eigenvalue of A .

QED

But condition (IV), Eq. (3.1), is then violated, because the eigenvalue of a linear combination of two non-commuting operators is, in general, different from the corresponding linear combination of the eigenvalues of the single operators as shown by the example of the spin variables of a spin- $\frac{1}{2}$ particle given in Sec. 2.5. This concludes Kochen and Specker's proof of von Neumann's theorem.

QED

Note that it is not assumed in this proof that every self-adjoint operator on a Hilbert space corresponds to an observable in contrast to von Neumann's postulate (i) above, this assumption being rather artificial. In general, only a few operators stand for physically significant observables (for an example see Chap. 7).

In view of the corollary, it is almost trivial that the unrestricted validity of von Neumann's postulate (v) rules out dispersion free states in quantum mechanics. Bell criticizes the unlimited assumption of postulate (v) for dispersion free states as unfounded on physical grounds (see Sec. 3.3) and qualifies it lateron even as 'silly' [16, p. 805]. However, in order to do justice to von Neumann's work, it has to be reminded that the topic of his pioneering book is *the mathematical foundation of quantum mechanics*. Within this scope and in consideration of von Neumann's conviction at that time that quantum mechanics is a statistical theory, his postulate (v) being true for quantum mechanical expectation values is appropriate and appears as a necessary completion of the formal definition of the sum of observables given in postulate (iii). This may explain von Neumann's statement cited above, and why he stated his postulate (v) for dispersion free states, too.

However that may be, more important than this controversy is the fact that hidden variables theories are ruled out even if the additivity of the expectation value is postulated only for commuting subalgebras. This was shown independently by Kochen and Specker and, ironically, also by Bell about 35 years after von Neumann's contribution.

3.2 The Kochen-Specker theorem

The analysis in the previous section indicates that the construction of hidden variables theories based on dispersion free states also requires a new definition of the mathematical representation of observables. In a dispersion free state, each observable is represented by its sharply predicted expectation value. If a hidden variables theory is complete in the sense that it can reproduce all quantum mechanical expectation values, one inevitably arrives at a classical theory as described in Sec. 2.5 where the carriers of dispersion free, *i.e.* atomic, states build the phase space and where observables are represented by real functions on this phase space necessarily forming a commutative algebra. It is the achievement of Kochen and Specker [12] to have reformulated the problem of hidden variables in quantum mechanics in this manner. Their condition on the map of the quantum theory to a classical theory is that it respects the different statuses of observables and states and the fundamental relationship (2.1).

A classical model of a quantum theory consists, therefore, on a map v of the quantum mechanical operators \hat{A} to the real measurable functions \tilde{A} over the phase space Λ of the classical theory and on a map w of the state operator ρ_s to a probability measure μ_s on the phase space

$$v : \quad \hat{A} \rightarrow v_{\hat{A}} \equiv \tilde{A} : \lambda \in \Lambda \rightarrow \tilde{A}(\lambda) \in \mathbb{R} \quad (3.5)$$

$$w : \quad \rho_s \rightarrow w(\rho_s) \equiv \mu_s : \sigma \subset \Lambda \rightarrow [0, 1] \quad (3.6)$$

such that the quantum mechanical expectation values are reproduced by their classical equivalents

$$P_{A,s}(\sigma) = \text{Tr}(\rho_s \chi_\sigma(\hat{A})) = w(\rho_s)((v_{\hat{A}})^{-1}(\sigma)) = \mu_s(\tilde{A}^{-1}(\sigma)) \quad (3.7)$$

The atomic states of the underlying classical theory have the status of hidden variables since dispersion free states cannot be prepared according to quantum mechanics and to experience.

To respect the different statuses of observables and states means that the map (3.5) may not be generalized to $\hat{A} \rightarrow v_{\hat{A},\rho}(\lambda)$. This is one of the reasons why we do not consider Bohm's pilot wave interpretation of quantum mechanics [32, 33] a decent classical hidden variables model. An other reason is its grossly nonlocal character.

The relationship (2.1), valid already on the level of the probability measures of observables in the GFS, imposes the necessary condition

$$v_{f(\hat{A})} = f(v_{\hat{A}}) \quad (3.8)$$

on the map v in order to ensure a consistent definition of the real function $f(A)$ of an observable A . In particular, if $\hat{A} = f(\hat{C})$ and $\hat{B} = g(\hat{C})$, then $\hat{A} + \hat{B} = f(C) + g(C) = (f+g)(\hat{C})$ and $\hat{A}\hat{B} = f(C)g(C) = (fg)(\hat{C})$ and from condition (3.8)

$$v_{\hat{A}+\hat{B}} = v_{\hat{A}} + v_{\hat{B}} \quad (3.9)$$

$$v_{\hat{A}\hat{B}} = v_{\hat{A}} v_{\hat{B}} \quad (3.10)$$

A constant aI can be considered as a function of any observable. It holds, therefore, for all observables A that $v_{a\hat{I}+\hat{A}} = a\tilde{I} + v_{\hat{A}}$ with $\tilde{I} : \lambda \in \Lambda \rightarrow 1$. Remember that relationships analogous to (3.9) and (3.10) are valid also for individual measurement values because of Eq. (2.2) which entails the same structure for measurement values as does Eq. (3.8) for functions of an observable.

Stated in this manner, the question whether hidden variables exist is now a purely mathematical problem concerning the possibility to map the noncommuting observable algebra of a quantum theory on the commuting observable algebra of a classical theory. Kochen and Specker [12] show that this is, in general, not possible even if this map is required only for commuting subalgebras of the quantum theory.

Theorem of Kochen and Specker *There is no classical model in the sense of Equations (3.5), (3.6), and (3.8) of a quantum theory whose Hilbert space has a dimension greater than 2 if a sufficient number of operators are observables.*

Proof In order to simplify the notation we no longer distinguish between the symbols for observables and for the related quantum mechanical operators and denote both by A , Q and so on. We show now that no classical model of the spin operators of a spin-1 particle exists which respects the conditions (3.5), (3.6), and (3.8). The relevant Hilbert space is isomorphic to \mathbb{C}^3 and the observables are represented by Hermitian 3×3 matrices if a fixed orthonormal basis of \mathcal{H} is chosen. Let the triplet $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of unit vectors in \mathbb{R}^3 indicate the axes of a right-handed Cartesian coordinate frame. Set the spin $\mathbf{J} = \hbar \mathbf{M}$. In a simultaneous eigenbasis of \mathbf{M}^2 and of the component $(\mathbf{e}_3, \mathbf{M})$, the operators $(\mathbf{e}_j, \mathbf{M})$ have the following matrix representation [34, Sec. 24]:

$$(\mathbf{e}_1, \mathbf{M}) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (\mathbf{e}_2, \mathbf{M}) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad (\mathbf{e}_3, \mathbf{M}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (3.11)$$

Using this matrix representation one easily verifies that the operators

$$Q(\mathbf{e}_j) = Q(-\mathbf{e}_j) \equiv 1 - (\mathbf{e}_j, \mathbf{M})^2, \quad j = 1, 2, 3 \quad (3.12)$$

satisfy the relationships

$$Q(\mathbf{e}_j) Q(\mathbf{e}_k) = Q(\mathbf{e}_k) Q(\mathbf{e}_j) = \delta_{jk} Q(\mathbf{e}_j), \quad j, k = 1, 2, 3 \quad (3.13)$$

$$Q(\mathbf{e}_1) + Q(\mathbf{e}_2) + Q(\mathbf{e}_3) = 1 \quad (3.14)$$

for any orthonormal basis $\{\mathbf{e}_j\}$ of \mathbb{R}^3 . The projection $Q(\mathbf{e})$ projects on the eigenspace to the eigenvalue 0 of the spin component (\mathbf{e}, \mathbf{M}) and corresponds to the proposition " (\mathbf{e}, \mathbf{M}) has the value 0".

For any orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, the projections $Q(\mathbf{e}_j)$ can be considered as functions of an observable A , because all $Q(\mathbf{e}_j)$ commute. Choose

$$A = a_1 Q(\mathbf{e}_1) + a_2 Q(\mathbf{e}_2) + a_3 Q(\mathbf{e}_3) \quad \text{with} \quad a_1 < a_2 < a_3 \quad (3.15)$$

Then it is

$$Q(\mathbf{e}_j) = f_j(A), \quad j = 1, 2, 3 \quad (3.16)$$

with

$$Q(\mathbf{e}_1) = f_1(A) = \frac{(A - a_2)(A - a_3)}{(a_1 - a_2)(a_1 - a_3)} \quad (3.17)$$

and so on. We therefore relate a projection $Q(\mathbf{e})$ to each direction $\|\mathbf{e}\| = 1$ in \mathbb{R}^3 in such a way that the properties (3.13), (3.14), and (3.16) are valid for every orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Because of Eqns. (3.16) and (3.9), the classical representations of the operators $Q(\mathbf{e})$ generate, for each value $\lambda \in \Lambda$, a function

$$p(\mathbf{e}) \equiv v_{Q(\mathbf{e})}(\lambda) \quad (3.18)$$

with the following properties

$$p(\mathbf{e}_1) + p(\mathbf{e}_2) + p(\mathbf{e}_3) = 1 \quad (3.19)$$

if $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is an orthonormal basis of \mathbb{R}^3 , and, for all directions \mathbf{e} ,

$$p : \mathbf{e} \rightarrow \{0, 1\} \quad (3.20)$$

because $v_{Q(\mathbf{e})}(\lambda)$ is a characteristic function: $Q(\mathbf{e})^2 = Q(\mathbf{e})$ implies $v_{Q(\mathbf{e})}^2 = v_{Q(\mathbf{e})}$. But Eqns. (3.19) and (3.20) cannot be fulfilled according to a theorem due to A. M. Gleason [35, 36]:

Theorem of Gleason *Let \mathcal{S} denote the unit sphere of a complex or real Hilbert space \mathcal{H} , $\mathcal{S} = \{\psi \in \mathcal{H} / \|\psi\| = 1\}$, and $p(\phi)$ a frame function i.e. a nonnegative real function on \mathcal{S} such that $\sum_n p(\phi_n) = 1$ for any orthonormal basis $\{\phi_n\}$ of \mathcal{H} . Then for a complex (or real) Hilbert space of dimension at least three, there exists, for any frame function, a self-adjoint (or symmetric) operator T defined on \mathcal{H} such that $p(\phi) = (\phi, T\phi)$. The eigenvalues of T are nonnegative and T belongs to the trace class with $\text{Tr}(T) = 1$.*

T is a state operator. Frame functions are thus continuous. The function $p(\mathbf{e})$ fulfilling condition (3.19) is a frame function on the unit sphere of \mathbb{R}^3 and cannot have uniquely the values 0 and 1. This concludes the proof of the theorem of Kochen and Specker. **QED**

Actually, Kochen and Specker argue on the basis of the operators $(\mathbf{e}_j, \mathbf{M})^2$ instead of $Q(\mathbf{e})$ and show that Eqns. (3.19) and (3.20) are contradictory without using Gleason's theorem. They construct a finite subset of orthonormal triplets with altogether 117 different directions \mathbf{e}_n , for which it is not possible to assign the values 0 and 1 in such a way that the value 0 occurs exactly once in each orthonormal triplet.

3.2.1 The case of two dimensions

The above proof breaks down in two dimensions; a function $p(\mathbf{e})$ on the unit circle, for which $p(\mathbf{e}_1) + p(\mathbf{e}_2) = 1$ if $(\mathbf{e}_1, \mathbf{e}_2) = 0$, can be arbitrarily given on a quadrant. In fact, Kochen and Specker provide a classical model in the sense of Eqns. (3.5), (3.6), and (3.8) for the self-adjoint operators on a 2-dimensional Hilbertspace and for pure states.

In order to introduce this classical model let us first analyse the structure of Hermitian matrices on \mathbb{C}^2 . The defining condition $A = A^\dagger$ implies that A is determined by 4 real parameters and can be given as a real linear combination of the identity operator I and the Pauli spin matrices $\sigma_1, \sigma_2, \sigma_3$. In an eigenbasis of σ_3 , the spin matrices are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3.21)$$

From this representation, the principal properties of spin matrices are readily verified. It is

$$\sigma_j^\dagger = \sigma_j, \quad \text{Tr}(\sigma_j) = 0, \quad \sigma_j^2 = I, \quad \sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3 + \text{cycl.} \quad (3.22)$$

and all σ_j have eigenvalues ± 1 . Any Hermitian 2×2 -matrix A can be written as

$$A = a_0 I + \sum_{j=1}^3 a_j \sigma_j \equiv a_0 I + \|\mathbf{a}\| (\mathbf{e}_A, \boldsymbol{\sigma}) \quad (3.23)$$

with real parametres a_0 and \mathbf{a} . A has 2 different eigenvalues $r_2 > r_1$ if $\|\mathbf{a}\| > 0$. Set in this case

$$\sigma(A) = \frac{2}{r_2 - r_1} A - \frac{r_2 + r_1}{r_2 - r_1} I \quad (3.24)$$

r_2 being mapped to $+1$, r_1 to -1 . Comparison wit Eq. (3.23) reveals that

$$a_0 = \frac{r_2 + r_1}{2}, \quad \|\mathbf{a}\| = \frac{r_2 - r_1}{2}, \quad (\mathbf{e}_A, \boldsymbol{\sigma}) = \sigma(A) \quad (3.25)$$

The matrices $f(A)$ and $\sigma(A)$ are diagonal in an eigenbasis $\{\psi_1, \psi_2\}$ of A such that $A\psi_k = r_k\psi_k$, $k = 1, 2$,

$$f(A) = \begin{pmatrix} f(r_2) & 0 \\ 0 & f(r_1) \end{pmatrix}, \quad \sigma(A) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3.26)$$

and it becomes evident that $f(A)$ is the linear combination

$$f(A) = \frac{f(r_2) + f(r_1)}{2} I + \frac{f(r_2) - f(r_1)}{2} (\mathbf{e}_A, \boldsymbol{\sigma}) \quad (3.27)$$

of I and $(\mathbf{e}_A, \boldsymbol{\sigma})$.

According to a remark after Eq. (3.10), it suffices to define the map v for all operators of the form $(\mathbf{e}, \boldsymbol{\sigma})$ in order to obtain a consistent classical model, condition (3.8) imposing no further restrictions. The limited bearing of this condition in the case of two dimensions is the reason why classical models are possible here.

Alternatively, we may choose the projections

$$Q(\mathbf{e}) = \frac{1}{2} (I + (\mathbf{e}, \boldsymbol{\sigma})) \quad (3.28)$$

instead of the operators $(\mathbf{e}, \boldsymbol{\sigma})$ in order to describe the classical model. $Q(\mathbf{e})$ projects on the eigenvector to the eigenvalue 1 of $(\mathbf{e}, \boldsymbol{\sigma})$. It is $Q(\mathbf{e}) + Q(-\mathbf{e}) = 1$, $Q(\mathbf{e}) Q(-\mathbf{e}) = 0$, and $f(A) = f(r_2) Q(\mathbf{e}_A) + f(r_1) Q(-\mathbf{e}_A) = f(r_1) I + (f(r_2) - f(r_1)) Q(\mathbf{e}_A)$.

Let $Q_\psi = \frac{1}{2} (I + (\mathbf{e}_\psi, \boldsymbol{\sigma}))$ denote the projection operator on the pure state ψ , the eigenstate of $(\mathbf{e}_\psi, \boldsymbol{\sigma})$ with eigenvalue +1. The expectation value $\langle Q(\mathbf{e}) \rangle_\psi$ is given by

$$\langle Q(\mathbf{e}) \rangle_\psi = \text{Tr} (Q_\psi Q(\mathbf{e})) = \frac{1}{2} (1 + (\mathbf{e}_\psi, \mathbf{e})) \quad (3.29)$$

We are now ready to present the classical model of Kochen and Specker for the operators of a 2-dimensional Hilbertspace [12, Sec. 6]. As phase space, they choose $\mathcal{S}^2 = \{\mathbf{e} \in \mathbb{R}^3 / \|\mathbf{e}\| = 1\}$ with the usual surface element. The classical counterpart of $Q(\mathbf{e}_A)$ is given by

$$v_{Q(\mathbf{e}_A)}(\mathbf{e}) = \begin{cases} 1 & \text{if } (\mathbf{e}, \mathbf{e}_A) > 0 \\ \frac{1}{2} & \text{if } (\mathbf{e}, \mathbf{e}_A) = 0 \\ 0 & \text{if } (\mathbf{e}, \mathbf{e}_A) < 0 \end{cases} \quad (3.30)$$

The measure μ_ψ representing the pure quantum state ψ shall follow from a continuous probability distribution ρ_ψ . Kochen and Specker show that, with the ansatz (3.30), the only possible choice for ρ_ψ is

$$\rho_\psi(\mathbf{e}) = \frac{1}{\pi} \times \begin{cases} (\mathbf{e}, \mathbf{e}_\psi) & \text{if } (\mathbf{e}, \mathbf{e}_\psi) \geq 0 \\ 0 & \text{if } (\mathbf{e}, \mathbf{e}_\psi) < 0 \end{cases} \quad (3.31)$$

With (3.30) and (3.31), the quantum mechanical expectation values are reproduced implying

$$\langle Q(\mathbf{e}_A) \rangle_\psi = \frac{1}{2} (1 + (\mathbf{e}_\psi, \mathbf{e}_A)) = \int_{\mathcal{S}^2} d\omega \rho_\psi v_{Q(\mathbf{e}_A)} \quad (3.32)$$

for arbitrary directions \mathbf{e}_ψ and \mathbf{e}_A . Geometric visualization readily reveals this formula to be correct for $(\mathbf{e}_\psi, \mathbf{e}_A) = 0$ and $(\mathbf{e}_\psi, \mathbf{e}_A) = 1$. The general case may be checked by direct evaluation [12, p. 79].

It should, however, be mentioned that this classical model is by no means equivalent to the quantum mechanical description, because important structural properties of the spin- $\frac{1}{2}$ algebra as *e.g.* the commutation relations

$$\frac{1}{i} [(\mathbf{e}, \boldsymbol{\sigma}), (\mathbf{e}', \boldsymbol{\sigma})] = 2 (\mathbf{e} \times \mathbf{e}', \boldsymbol{\sigma}) \text{ corresponding to } \frac{1}{i} [Q(\mathbf{e}), Q(\mathbf{e}')] = Q(\mathbf{e} \times \mathbf{e}') - \frac{1}{2} I$$

cannot be represented. The classical model does not provide a physical description of spin- $\frac{1}{2}$ observables even though it reproduces all quantum mechanical predictions.

Let us face the positions of von Neumann and of Kochen and Specker. In contrast to the postulate (v) of von Neumann, Kochen und Specker use the linearity of the expectation values only on commuting subsets of the quantum mechanical operators. Von Neumann introduces dispersion free states in quantum theories, which is inadequate. Kochen and Specker situate the dispersion free states into the realm of classical models where they naturally occur. Kochen and Specker construct a classical hidden variable model conforming with their conditions (3.5), (3.6), and (3.8) for quantum theories with a 2-dimensional Hilbert space. The dispersion free states of von Neumann are impossible even in this case.

3.3 Bell's 1966 paper

In a seminal paper [30], J. S. Bell reconsidered von Neumann's arguments against hidden variables theories. Stimulated by Gleason's paper [35], he derived a mathematical result equivalent to the mathematical content of the Kochen-Specker theorem, but drew quite different conclusions out of it. Both results were discovered independently of each other. Bell's mathematical result is

Bell's Corollary *Let \mathcal{S} denote the unit sphere of a complex or real Hilbertspace \mathcal{H} of dimension greater than two, $\mathcal{S} = \{\psi \in \mathcal{H} / \|\psi\| = 1\}$, and p a nonnegative function on \mathcal{S} such that $\sum_n p(\psi_n) = 1$ for every orthonormal basis $\{\psi_n\}$ of \mathcal{H} . Let ψ and ϕ be in \mathcal{S} with $p(\psi) = 1$ and $p(\phi) = 0$.*

Then $\|\phi - (\phi, \psi) \psi\| > \frac{1}{2} |(\phi, \psi)|$. In the case of a real Hilbert space, this amounts to $1 - (\phi, \psi)^2 > \frac{1}{4} (\phi, \psi)^2$ or to $\cos \theta \equiv |(\phi, \psi)| < 2/\sqrt{5}$ implying $\theta > 26.56^\circ$. In the case of a complex Hilbert space, it follows that $1 - 2 \Re((\phi, \psi)^2) + |(\phi, \psi)|^2 > \frac{1}{4} |(\phi, \psi)|^2$. This inequality must be valid for arbitrary phase factors of ϕ and ψ implying again $|(\phi, \psi)| < 2/\sqrt{5}$.

The function p is a frame function as defined by Gleason. The following lemmas A and B are valid for frame functions:

Lemma A *If $p(\psi) = 1$ and $(\psi, \phi) = 0$ then $p(\phi) = 0$*

Proof ψ and ϕ may be considered as basis vectors of an orthonormal basis $\psi, \phi, \phi_3, \phi_4, \dots$ of \mathcal{H} . **QED**

Lemma B *If $(\phi_1, \phi_2) = 0$ and $p(\phi_1) = p(\phi_2) = 0$ then $p(c_1 \phi_1 + c_2 \phi_2) = 0$ for all $c_1, c_2 \in \mathbb{C}$, or $c_1, c_2 \in \mathbb{R}$, respectively, with $|c_1|^2 + |c_2|^2 = 1$.*

Proof Let $\phi_1, \phi_2, \phi_3, \phi_4, \dots$ and $\psi_1, \psi_2, \phi_3, \phi_4, \dots$ denote two orthonormal bases of \mathcal{H} . As a consequence $p(\psi_1) + p(\psi_1) = p(\phi_2) + p(\phi_2) = 0$ and $p(\psi_1) = p(\psi_2) = 0$ because $p(\psi_1) \geq 0$ and $p(\psi_2) \geq 0$. Set $\psi_1 = c_1 \phi_1 + c_2 \phi_2$. **QED**

Proof of Bell's corollary ϕ and ψ span a 2-dimensional subspace $\mathcal{V} \subset \mathcal{H}$. Choose a vector $\psi' \in \mathcal{V} \cap \mathcal{S}$ such that $(\psi, \psi') = 0$. ψ and ψ' are an orthonormal basis of \mathcal{V} and it is, with $\epsilon \equiv (\phi, \psi')$,

$$\phi = (\phi, \psi) \psi + \epsilon \psi'$$

implying $|\epsilon| = \|\phi - (\phi, \psi) \psi\|$. We can assume that $(\phi, \psi) \neq 0$.

Let ψ'' be a vector in \mathcal{S} with $(\psi'', \psi) = (\psi'', \psi') = 0$, and as a consequence, $(\psi'', \phi) = 0$ (here the condition that the dimension of \mathcal{H} is greater than 2 is needed). It is $p(\psi') = p(\psi'') = 0$ because of the supposition

$p(\psi) = 1$ and of Lemma A. Let γ be a real number different from 0 and

$$\begin{aligned}\psi_a &= \frac{1}{d_a} (\phi + \gamma^{-1} \epsilon \psi''), & d_a &= \|\phi + \gamma^{-1} \epsilon \psi''\| \\ \psi_b &= \frac{\epsilon}{d_b} (-\psi' + \gamma \psi''), & d_b &= \|\epsilon(-\psi' + \gamma \psi'')\|\end{aligned}$$

ψ_a and ψ_b are orthogonal, $(\psi_a, \psi_b) = 0$. Since $p(\phi) = p(\psi)' = p(\psi'') = 0$, it follows $p(\psi_a) = p(\psi_b) = 0$ with Lemma B. Again with Lemma B, it is

$$\begin{aligned}0 &= p\left(\frac{1}{\sqrt{d_a^2 + d_b^2}} (d_a \psi_a + d_b \psi_b)\right) \\ &= p\left(\frac{1}{\sqrt{d_a^2 + d_b^2}} ((\phi, \psi) \psi + \epsilon (\gamma + \gamma^{-1}) \psi'')\right) \\ &= p\left(\frac{|(\phi, \psi)|}{\sqrt{d_a^2 + d_b^2}} \left(\psi + \frac{\epsilon}{(\phi, \psi)} (\gamma + \gamma^{-1}) \psi''\right)\right) \\ &= p\left(\frac{|(\phi, \psi)|}{\sqrt{d_a^2 + d_b^2}} \left(\psi + \eta (\gamma + \gamma^{-1}) e^{i\beta} \psi''\right)\right)\end{aligned}$$

with $\epsilon/(\phi, \psi) = \eta e^{i\beta}$, $\eta = |\epsilon/(\phi, \psi)|$. If $\eta \leq \frac{1}{2}$, there exist values of γ with $\eta(\gamma + \gamma^{-1}) = \pm 1$ implying $p((\psi + e^{i\beta} \psi'')/\sqrt{2}) = p((\psi - e^{i\beta} \psi'')/\sqrt{2}) = 0$. It follows with Lemma B that $p(\psi) = 0$ contradicting the assumption $p(\psi) = 1$ we began with. Therefore $|\epsilon| = \|\phi - (\phi, \psi) \psi\| > \frac{1}{2} |(\phi, \psi)|$. **QED**

Bell stated his corollary *not* with the intention of proving the impossibility of hidden variables but rather to indicate a loophole in order to render hidden variables possible. The topic of his paper [30] is a critical discussion of von Neumann's argument that the proof of the nonexistence of dispersion free states in quantum mechanics discards hidden variables theories. Von Neumann's essential assumption in his proof is: *Any real linear combination of any two Hermitian operators represents an observable, and the same linear combination of expectation values is the expectation of the combination* [30, p. 448]. This assumption is true for quantum mechanical states but von Neumann requires it also for the hypothetical dispersion free states. It is, however, rather trivial that this requirement can then not be upheld as we have seen in Sec. 3.1.

Bell criticises von Neumann's assumption on physical grounds: *A measurement of a sum of noncommuting observables cannot be made by combining trivially the results of separate observations on the two terms — it requires a quite distinct experiment. For example the measurement of σ_x for a magnetic particle might be made with a suitably oriented Stern Gerlach magnet. The measurement of σ_y would require a different orientation, and of $(\sigma_x + \sigma_y)$ a third and different orientation. But this explanation of the nonadditivity of allowed (measurement) values also establishes the nontriviality of the additivity of expectation values. The latter is a quite peculiar property of quantum mechanical states, not to be expected a priori. There is no reason to demand it individually of the hypothetical dispersion free states, whose function is to reproduce the measurable peculiarities of quantum mechanics when averaged over* [30, p. 449]. This criticism clearly reveals a central point.

Bell was aware that the work of Gleason [35] *apparently enables von Neumann's result to be obtained without objectionable assumptions about noncommuting observables. ... The relevant corollary of Gleason's work is that, if the dimensionality of the state space is greater than two, the additivity requirement for expectation values of commuting operators cannot be met by dispersion free states. This will now be proved, and then its significance discussed* [30, p. 450].

In the proof, Bell considers the set of projection operators $P(\phi)$ projecting onto vectors ϕ of a Hilbert space \mathcal{H} and their expectation values $\langle P(\phi) \rangle$. For a complete and orthogonal set $\{\phi_n\}$ of \mathcal{H} , it is $P(\phi_n)P(\phi_m) = \delta_{nm}P(\phi_n)$ and $\sum_n P(\phi_n) = I$. Bell does not specify if the expectation values $\langle P(\phi) \rangle$ denote quantum mechanical or dispersion free expectations, but assumes that they generate a frame function. This is true for the quantum mechanical case: the projection formula

$$P(\phi)\psi = \frac{(\phi, \psi)}{(\phi, \phi)} \phi$$

shows that $P(\cdot)$ is actually a function of the unit vectors $\phi/\sqrt{(\phi, \phi)}$ and the quantum mechanical expectation value for the state ρ , $\langle P(\phi) \rangle_\rho = \text{Tr}(\rho P(\phi))$, is a frame function, because $\langle P(\phi) \rangle_\rho$ is nonnegative as the expectation value of a projector and $\sum_n \langle P(\phi_n) \rangle_\rho = \langle \sum_n P(\phi_n) \rangle_\rho = 1$.

Dispersion free states whose expectation values $\langle P(\phi) \rangle$ are supposed to generate a frame function are ruled out by Bell's corollary. *For such states each projector has expectation value either 0 or 1. It is clear ... that both values must occur; and since there are no other values possible, there must be arbitrarily close pairs ψ, ϕ with different expectation values 0 and 1, respectively. But we saw above such pairs could not be arbitrarily close. Therefore, there are no dispersion free states.* [30, p. 451].

Bell's findings invalidate, for his purposes, his conclusions drawn above from the criticism of von Neumann's theorem to a great extent, and so is he forced to invalidate, in turn, the consequences of his own proof: *Are the requirements imposed, which are satisfied by quantum mechanical states, reasonable requirements on the dispersion free states? Indeed they are not. Consider the statement (B), [lemma B with $p(\phi) = \langle P(\phi) \rangle$]. The operator $P(\alpha\phi_1 + \beta\phi_2)$ commutes with $P(\phi_1)$ and $P(\phi_2)$ only if either α or β is zero. Thus in general measurement of $P(\alpha\phi_1 + \beta\phi_2)$ requires a quite distinct experimental arrangement. We can therefore reject (B) on the grounds already used: it relates in a nontrivial way the results of experiments which cannot be performed simultaneously; the dispersion free states need not have this property, it will suffice if the quantum mechanical averages over them do. How did it come about that (B) was a consequence of assumptions in which only commuting operators were explicitly mentioned? The danger in fact was not in the explicit but in the implicit assumptions. It was tacitly assumed that measurement of an observable must yield the same value independently of what other measurements may be made simultaneously. Thus as well as $P(\phi_3)$ say, one might measure either $P(\phi_2)$ or $P(\psi_2)$, where ϕ_2 and ψ_2 are both orthogonal to ϕ_3 but not to one another. These different possibilities require different experimental arrangements; there is no a priori reason to believe that the results for $P(\phi_3)$ should be the same* [30, p. 451].

This argumentation is, however, much less convincing here as it is with the setting of von Neumann. First of all, lemma B is a *mathematical* consequence for frame functions and it seems odd to criticise it on *physical* grounds. Instead, one may argue that the classical, dispersion free hidden states do not necessarily generate frame functions. But even then the expectation value of $P(c_1\phi_1 + c_2\phi_2)$ for dispersion free states cannot be freely chosen. Remind that the issue here is the possibility of classical dispersion free states under the condition that the quantum mechanical expectation values $\langle P(\phi) \rangle_\chi^{\text{qm}}$ for a pure state χ are convex linear combinations of classical dispersion free expectation values $\langle P(\phi) \rangle_\lambda^{\text{cl}}$. In view of conditions (2.10), (3.5) and (3.6), these linear combinations have the form

$$\langle P(\phi) \rangle_\chi^{\text{qm}} = \int_\Lambda \widetilde{P(\phi)}(\lambda) d\mu_\chi(\lambda) \equiv \int_\Lambda \langle P(\phi) \rangle_\lambda^{\text{cl}} d\mu_\chi(\lambda)$$

for all pure states χ . The classical expectations $\langle P(\phi) \rangle_\lambda^{\text{cl}}$ can attain only the values 0 or 1 and none of the probability measures μ_χ can be concentrated onto one point. As a consequence, if $\langle P(\phi) \rangle_\chi^{\text{qm}} = 0$ then $\langle P(\phi) \rangle_\lambda^{\text{cl}} = 0$ almost everywhere in Λ i.e. but for sets of measure zero, and if $\langle P(\phi) \rangle_\chi^{\text{qm}} = 1$ then $\langle P(\phi) \rangle_\lambda^{\text{cl}} = 1$ almost everywhere in Λ . Now, lemma A, lemma B, and Bell's corollary are valid for the quantum mechanical expectation values, $\langle P(\phi) \rangle_\chi^{\text{qm}}$ being frame functions. Therefore, if, in lemma B, $(\phi_1, \phi_2) = 0$, $\langle P(\phi_1) \rangle_\chi^{\text{qm}} = 0$, and $\langle P(\phi_2) \rangle_\chi^{\text{qm}} = 0$ imply $\langle P(c_1\phi_1 + c_2\phi_2) \rangle_\chi^{\text{qm}} = 0$ then that

all corresponding classical expectations, $\langle P(\phi_1) \rangle_\chi^{\text{cl}}, \langle P(\phi_2) \rangle_\chi^{\text{cl}}, \langle P(c_1\phi_1 + c_2\phi_2) \rangle_\chi^{\text{cl}}$ are equal to 0 almost everywhere in Λ . Further, in lemma A, lemma B, and in the proof of Bell's corollary, only Hilbert space vectors ϕ are essential, for which $\langle P(\phi) \rangle_\chi^{\text{qm}}$ is either 0 or 1. The corresponding dispersion free expectation values $\langle P(\phi) \rangle_\chi^{\text{cl}}$ must have the same values almost everywhere in Λ as the quantum mechanical expectation values do. They entail, therefore, the same restrictions for the vectors ϕ and ψ in Bell's corollary as the quantum mechanical expectation values. The consequences of Bell's corollary apply to the dispersion free states, too. It is not necessary to assume *a priori* that the dispersion free expectation values generate a frame function.

We, finally, mention that the presuppositions as well as the conclusions of the corollary, as stated by Bell in Ref. [30], differ in some details from the presuppositions and conclusions given above. Bell allows vectors of arbitrary positive length as arguments for P . He takes, as above, two vectors $\phi, \psi \in \mathcal{H}$ with $\langle P(\phi) \rangle = 0$ and $\langle P(\psi) \rangle = 1$ but comes to the conclusion that then $\|\phi - \psi\| > \frac{1}{2} \|\psi\|$. To show this he normalizes ψ and writes ϕ in the form

$$\phi = \psi + \epsilon\psi'$$

with $\|\psi'\| = 1$ and $(\psi, \psi') = 0$. But this ansatz introduces undesirable interdependences between ϕ and ψ . Taking the scalar product with ψ on both sides yields $(\phi, \psi) = (\psi, \psi) = 1$. This implies that $1 = (\phi, \psi) < \|\phi\| \|\psi\|$ or $\|\phi\| > 1$. Bell's ansatz is, therefore, *not* possible for vectors ϕ and ψ both of length 1, the only case where an inequality of the form $\|\phi - \psi\| > \frac{1}{2} \|\psi\|$ really makes sense.

Bell's corollary may substitute Gleason's theorem [35] in the above proof of the theorem of Kochen and Specker. The functions $p(\mathbf{e})$ of Eq. (3.19) fulfill the presuppositions of Bell's corollary for a real Hilbert space. But Eq. (3.20) implies that directions \mathbf{e} and \mathbf{e}' with $p(\mathbf{e}) = 1$ and $p(\mathbf{e}') = 0$ are arbitrarily close, which is excluded by Bell's corollary. One easily finds four right-trihedrons with altogether nine different directions, to which the values 0 and 1 cannot be consistently assigned [16, p. 808]: choose the x , y , and z axis of a Cartesian frame as the first trihedron. Set the value of p on the z axis equal to 1, implying that $p = 0$ on the x and the y axis, and consider three additional trihedrons generated by rotations of the first one around the x axis by angles of 25° , 50° , and 75° .

Bell's corollary excludes quantum mechanical expectation values that are dispersion free for all observables even if linearity is assumed merely for commuting observables. This result is related to but not identical with the result of Kochen and Specker. To obtain the Kochen-Specker theorem, consider a complete and orthogonal set $\{\phi_n\}$ of vectors in \mathcal{H} as the eigenbasis of a purely discrete observable A with nondegenerate eigenvalues a_n , $A\phi_n = a_n\phi_n$. Then it is $P(\phi_n) = \chi_{[a_n]}(A)$, and $\sum_n P(\phi_n) = I$, $P(\phi_n)P(\phi_m) = \delta_{nm}P(\phi_m)$ as before. There is no classical hidden variables model for the set of quantum mechanical projections $P(\phi)$ because their classical representatives $v_{P(\phi)}(\lambda)$ fulfill, for all λ , the relationships $\sum_n v_{P(\phi_n)}(\lambda) = 1$ for any orthogonal and complete set $\{\phi_n\}$, and it is $v_{P(\phi)}(\lambda) \in \{0, 1\}$. For every fixed value of λ , the $v_{P(\phi)}(\lambda)$ define a frame function $p(\phi)$ over \mathcal{H} whose only values are 0 or 1. But such frame functions do not exist, according to Gleason's theorem and to Bell's corollary, if the dimension of \mathcal{H} is greater than two. Considered in this way, it becomes clear that Bell's argumentation based on measurement values misses the point. The nonexistence of classical hidden variables theories is a mathematical consequence of the crucial relationship (3.8) and of the structural incompatibility of the quantum mechanical and the classical observable algebras.

3.4 Mermin's observable sets with no classical model

The proof of the theorem of Kochen and Specker seemingly only deals with commuting subsets of quantum mechanical operators. This obscures that, infact, it is tried to map a set of quantum mechanical operators, noncommutative in its totality, into the totally commutative set of phase space functions of the classical model, which turns out to be impossible except for the case of two dimensions. N. D. Mermin [16, p. 809 – 811] constructs a set of nine operators acting on a 4-dimensional Hilbert space and a set of ten operators,

acting on an 8-dimensional Hilbert space, both with no classical models, which reveal this point especially clearly. We first discuss the second case.

Mermin's ten operators A_j are expressed in terms of the spin matrices σ_1 and σ_3 of three independent spin- $\frac{1}{2}$ particles denoted with superscripts a, b , and c (Mermin actually uses the components σ_2 and σ_3). They form a subset of the self-adjoint operators on the 8-dimensional Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\begin{aligned} A_1 &= \sigma_3^a \otimes \sigma_3^b \otimes \sigma_3^c & A_2 &= \sigma_1^a \otimes \sigma_1^b \otimes \sigma_3^c & A_3 &= \sigma_1^a \otimes \sigma_3^b \otimes \sigma_1^c & A_4 &= \sigma_3^a \otimes \sigma_1^b \otimes \sigma_1^c \\ A_5 &= \sigma_1^a \otimes I \otimes I & A_6 &= I \otimes \sigma_1^b \otimes I & A_7 &= I \otimes I \otimes \sigma_1^c \\ A_8 &= \sigma_3^a \otimes I \otimes I & A_9 &= I \otimes \sigma_3^b \otimes I & A_{10} &= I \otimes I \otimes \sigma_3^c \end{aligned} \quad (3.33)$$

It is $A_j^2 = I, j = 1, \dots, 10$ and

$$\begin{aligned} \text{(I)} \quad & A_1, A_2, A_3, A_4 \quad \text{all mutually commute} \quad A_1 A_2 A_3 A_4 = -I \\ \text{(II)} \quad & A_1, A_8, A_9, A_{10} \quad \text{all mutually commute} \quad A_1 A_8 A_9 A_{10} = I \\ \text{(III)} \quad & A_2, A_5, A_6, A_{10} \quad \text{all mutually commute} \quad A_2 A_5 A_6 A_{10} = I \\ \text{(IV)} \quad & A_3, A_5, A_9, A_7 \quad \text{all mutually commute} \quad A_3 A_5 A_9 A_7 = I \\ \text{(V)} \quad & A_4, A_8, A_6, A_7 \quad \text{all mutually commute} \quad A_4 A_8 A_6 A_7 = I \end{aligned} \quad (3.34)$$

In addition, A_1, A_2, A_3, A_4 all are products of commuting operators. But some of the pairs $A_j A_k$ anticommute: $A_5 A_8 = -A_8 A_5, A_6 A_9 = -A_9 A_6, A_7 A_{10} = -A_{10} A_7, A_5 A_1 = -A_1 A_5, \dots$. It is because of these anticommutations that the product $A_1 A_2 A_3 A_4 = -I$, even though these four observables all commute and every one of these A_j is a product of three commuting observables.

That A_1, A_2, A_3, A_4 all commute implies that there exists a system of eight common eigenvectors ψ_1, \dots, ψ_8 of the $A_k, k = 1, \dots, 4$. The projection operators Q_j with $Q_j \psi_l = \delta_{jl} \psi_l, j, l = 1, 2, \dots, 8$, can be considered as functions $f_j(B)$ of the operator $B = \sum_{j=1}^8 j Q_j$ and, in turn, the $A_k, k = 1, \dots, 4$, as functions $g_k(Q_1, \dots, Q_8) = g_k(f_1(B), \dots, f_8(B)) = h_k(B)$. Similar considerations apply to the products $A_1 = A_8 A_9 A_{10}$ and so on. In terms of the classical model and in consideration of condition (3.10) which applies here, it is, therefore,

$$(v_{A_j}(\lambda))^2 = v_{A_j^2}(\lambda) = v_I(\lambda) = 1 \quad (3.35)$$

and

$$-1 = v_{A_1 A_2 A_3 A_4}(\lambda) = v_{A_1}(\lambda) v_{A_2}(\lambda) v_{A_3}(\lambda) v_{A_4}(\lambda) \quad (3.36)$$

and

$$v_{A_1}(\lambda) = v_{A_6}(\lambda) v_{A_{10}}(\lambda) v_{A_8}(\lambda) \quad (3.37)$$

and so on. Now each function $v_{A_j}, j = 5, 6, \dots, 10$, appears exactly twice in the product $v_{A_1} v_{A_2} v_{A_3} v_{A_4}$ which thus equals $+1$ not -1 — a plain contradiction! The reason for this contradiction is obvious: although only commuting observables are mapped in every of the consecutive steps, one tries, in fact, to map a partially noncommuting set of quantum mechanical observables into the totally commuting set of phase space functions. This is impossible as a consequence of the consistency requirement (3.8). It becomes evident that the noncommutative structure of quantum mechanical operators is intrinsic and cannot be circumvented.

Mermin's first example is based on the spin matrices of two independent spin- $\frac{1}{2}$ particles, $\frac{\hbar}{2} \sigma^a$ and $\frac{\hbar}{2} \sigma^b$, acting on the Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2$. The nine operators, arranged in a square array, are

$$\begin{array}{ccc} \sigma_1^a \otimes I & I \otimes \sigma_1^b & \sigma_1^a \otimes \sigma_1^b \\ I \otimes \sigma_2^b & \sigma_2^a \otimes I & \sigma_2^a \otimes \sigma_2^b \\ \sigma_1^a \otimes \sigma_2^b & \sigma_2^a \otimes \sigma_1^b & \sigma_3^a \otimes \sigma_3^b \end{array} \quad (3.38)$$

The three operators in each row and in each column all commute. As a consequence of the properties (3.22) of the Pauli spin matrices, the products of the operators of each row equals I . The product of the operators of the column on the left and on the middle equal I , but the product of the operators of the right column is $-I$. A reasoning analogous to the one above leads to a contradiction for the classical representatives of the nine operators. There is no classical model for Bohm's setting discussed in Chap. 5, a fact that significantly lowers the impact of Bell's theorem (Sec. 6.1). The crucial point is again that the noncommutative structure of the quantum mechanical operators is not represented by the classical model.

It is a remarkable feature of the theorems of Kochen and Specker and of Mermin that only the map v of the observables, Eq. (3.5), together with Eq. (3.8) is needed in the proofs. The map w , Eq. (3.6), of the states is not used and, as a consequence, the requirement that a classical hidden variables model should reproduce the quantum mechanical expectation values is irrelevant. The incompatibility between the noncommutative quantum mechanical observable algebra and the commutative classical observable algebra precludes hidden variables.

3.5 Criticism on the Kochen-Specker theorem

Let us first state that a trivial classical model of the GFS can always be constructed by a slight modification of an argument given by Kochen and Specker [12, p. 63]: let \mathbb{I} be some index set which may be finite, countable, $\mathbb{I} = \mathbb{N}$, or noncountable (e.g. $\mathbb{I} = \mathbb{R}$). Choose a subset $\{A_i / i \in \mathbb{I}'\}$, $\mathbb{I}' \subset \mathbb{I}$, of the set of all observables in question in such a way that every observable can be obtained as a Borel function $f(A_i)$ for some index $i \in \mathbb{I}'$. Choose $\mathbb{R}^{\mathbb{I}'}$ as the phase space Λ . The observable A_i is represented by the real function $\tilde{A}_i : \mathbb{R}^{\mathbb{I}'} \rightarrow \mathbb{R}$ depending linearly on the coordinate x_i and being independent of all other coordinates. The observable $f(A_i)$ is represented by $f(x_i)$ as usual. The state s is represented by the product measure $\prod_{i \in \mathbb{I}'} P_{A_i, s}(\Delta_i)$, $\Delta_i \subseteq \mathbb{R}_i$. As a consequence, no relationships between observables exist in this classical model except that an observable may be a real function of another one. But this is a mere list of probability measures cast into the structure of a classical model. Such a construct can hardly be called a theory, a physical theory being much more than a mere list of observations.

Bell (see e.g. [30], [27]) and Belinfante [28] argue that the value $m(A)$ of an observable A measured on an individual system should depend on which other observables are simultaneously measured together with A (*contextuality of measurement values*). Consider for instance the situation given by the equations (3.13) – (3.17). The observable $Q(\mathbf{e}_1)$ can be represented in the form $Q(\mathbf{e}_1) = f_1(A)$ in many different ways depending on the orthonormal triplet $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ chosen. A measurement of A is a simultaneous measurement of $Q(\mathbf{e}_1)$, $Q(\mathbf{e}_2)$, $Q(\mathbf{e}_3)$, but the measurement acts for different observables A require different measurement set-ups. Therefore, it seems plausible that the measured value $m(Q(\mathbf{e}_1))$ depends on which observable A the measurement was performed¹.

Such an assumption clearly undermines the proof given by Kochen and Specker. But is it really plausible? We think not. First of all, this assumption implies that the value of an observable ceases to be an objective property of the probed individual system alone, but that it depends on the measurement device chosen, a consequence particularly lacking attractivity in connection with classical hidden variables models. It is, moreover, undesirable. Every splitting up of quantum mechanical observables weakens the theoretical description. It invalidates structural relationships between observables and is a step towards the trivial classical model described above.

¹For a support of this view, Bell [30, p. 447] takes recourse to Bohr's remark on *the impossibility of any sharp separation between the behaviour of atomic objects and the interaction with the measuring instruments which serve to define the conditions under which the phenomena appear* (N. Bohr in [1, p. 210]). However, to cite Bohr in favour of hidden variables theories is misleading since Bohr abhorred the idea of hidden variables [16, p. 811]. The background of Bohr's remark is his concept of complementarity and his conviction that the wave function provides an exhaustive description of an individual physical system.

Secondly, this kind of contextuality does not happen in the quantum mechanical framework, as pointed out by Mermin [16, p. 811 – 812]: *Most strikingly, although it is indisputable that measuring A with mutually commuting B, C, \dots requires a different experimental arrangement from measuring it with mutually commuting L, M, \dots whenever some of L, M, \dots fail to commute with some of B, C, \dots , it is nevertheless an elementary theorem of quantum mechanics that the joint distribution $p(a, b, c, \dots)$ for the first experiment yields precisely the same marginal distribution $p(a)$ as does the joint distribution $p(a, l, m, \dots)$ for the second, in spite of the different experimental arrangements. If we do the experiment to measure A with B, C, \dots on an ensemble of systems prepared in the state Ψ and ignore the results of other observables, we get exactly the same statistics for A as we would have obtained had we instead done the quite different experiment to measure A with L, M, \dots on that same ensemble. The obvious way to account for this, particularly when entertaining the possibility of a hidden-variables theory, is to propose that both experiments reveal a set of values for A in the individual systems that is the same, regardless of which experiment we choose to extract them from. If one is attempting a hidden-variables model at all, it seems not unreasonable to expect the model to provide the obvious explanation for this striking insensitivity of the distribution to changes in the experimental arrangement — namely, that the hidden variables are noncontextual.*

Thirdly, in a classical model, the atomic state $\delta(\lambda - \lambda_0)$, into which the system is supposed to be, is concentrated on the point λ_0 of the phase space, the atomic states or their carriers being the hidden variables. The value $\tilde{A}(\lambda_0)$ of the classical representative of the observable A is the expectation value of this observable in the atomic state $\delta(\lambda - \lambda_0)$ and equal to the asserted measurement value $m(A)$. In the context of the classical hidden variables model, Bell's and Belinfante's objections mean that the one quantum mechanical observable $Q \equiv Q(\mathbf{e}_1)$ could be represented by several functions $\tilde{Q}, \tilde{Q}', \tilde{Q}'', \dots$ of the classical model. All these functions have the property to reproduce the quantum mechanical expectation values of Q :

$$\langle Q \rangle_s = \text{Tr}(\rho_s Q) = \int_{\Lambda} \tilde{Q}(\lambda) d\mu_s(\lambda) = \int_{\Lambda} \tilde{Q}'(\lambda) d\mu_s(\lambda) = \dots$$

for all states s . The differences $\tilde{Q}'(\lambda) - \tilde{Q}(\lambda)$, $\tilde{Q}''(\lambda) - \tilde{Q}(\lambda)$, \dots thus all represent the constant 0 but these representatives are not identically equal to 0 almost everywhere for all states s . This contradicts the requirement that every functional value of the classical representative of an observable should be equal to a measurement value of this observable. Conversely, the classical representative of an observable is unique if the constant 0 is solely represented by the function $0 \cdot \tilde{I} : \lambda \in \Lambda \rightarrow 0$.

Finally, every observable $Q(\mathbf{e})$ defined by Eq. (3.12) has a significance independent of being a member of a triplet $\{Q(\mathbf{e}_1), Q(\mathbf{e}_2), Q(\mathbf{e}_3)\}$: $Q(\mathbf{e}) = 1 - (\mathbf{e}, \mathbf{M})^2$ corresponds to the proposition "*the spin component (\mathbf{e}, \mathbf{M}) has the value 0*". Equations (3.13) – (3.17) solely follow from the properties of the spin-1 operator algebra, equations (3.15) and (3.16) make sure that, by applying the relationships (3.9) and (3.10), the expectation values of the set $\{\tilde{Q}(\mathbf{e})\}$ in an atomic state generate a frame function. These considerations have nothing to do with the question how $Q(\mathbf{e})$ is measured on an actual system. The relationships (3.15) and (3.16) may be exploited for a simultaneous measurement of the $Q(\mathbf{e}_j)$ of an orthonormal triplet $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. But this way is by no means necessary. The most obvious method to determine the value of $Q(\mathbf{e})$ consists in measuring the value of the spin component (\mathbf{e}, \mathbf{M}) , and contextuality plays no role.

3.6 The class of classical models is incomplete

The theorem of Kochen and Specker shows that there are no logically consistent classical models of quantum theories respecting the structure of the general formal scheme, except for the case of a 2-dimensional Hilbert space. Quantum mechanics is remarkably unimportant for the proof: the equations (3.5) – (3.7), the notions 'real function of an observable', 'sum or product of real functions of the same observable', 'proposition', can already be expressed in the GFS. Quantum mechanics is needed only in order to ensure that the set of observables $Q(\mathbf{e})$ with the properties (3.13), (3.14), and (3.16) really exists. The theorem of Kochen and

Specker reveals, infact, that there is a set of observables within the GFS, which can be described by quantum mechanics but not in classical physics. The crucial point is that the commutativity of the classical observable algebra is a too restrictive requirement.

This is a remarkable answer to the question *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?*, the title of the famous paper by Einstein, Podolsky, and Rosen [15]. Here, a theory is considered complete if it contains *an element corresponding to each element of reality. A sufficient condition for the reality of a physical quantity is the possibility of predicting it with certainty, without disturbing the system.* The authors devise a physical system in a pure quantum state, where two physical quantities described by two noncommuting operators simultaneously have physical reality in the above sense. They conclude that a pure quantum state hardly can be considered as the complete description of the physical reality of an *individual* system, an opinion still upheld by Bohr in his reply [37].

From our point of view, the question of completeness must be posed differently: a state is *never* a property of an individual, concrete system even not in classical theories, all the less its 'complete description'. It is a property of the experimental arrangement, of the infinite class of equally prepared equal systems. We propose to call a class of physical theories complete if it allows for the formulation of any system of prognoses relevant in the covered field. We think that nonrelativistic quantum mechanics of physical systems with a finite number of degrees of freedom is complete in this sense; the theorem of Kochen and Specker shows that classical theories are incomplete.

Classical theories express the idea of determinism, most clearly pictured by Laplace's demon: a demon knowing the forces acting between all atoms of the universe and, at a given moment, all their positions and velocities would be able to calculate the total history of the universe from its earliest past to its most remote future. Statistical scatter is then solely the result of the subjective ignorance about the atomic state of the universe. The theorem of Kochen and Specker shows that this conception of the physical world is too narrow, that already — or just — small physical systems behave differently. Physical processes are intrinsically indeterministic; the statistical predictions of pure quantum states are objective, not reducible to mere subjective ignorance.

Chapter 4

Further comments on state preparation and measurement of observable values

In the framework of the GFS, the statement " $P_{A,s}(\sigma)$ is the probability to find the measured value of the observable A in the interval σ if the system is prepared in the state s " presupposes, firstly, that there exist one or several devices suitable to obtain the desired measurement values and that a measurement reveals the value of an observable as an objective property of the inquired material system, secondly, that a series of measurement values forms a sample of the probability distribution $P_{A,s}(\cdot)$. The first point is almost tautological: physical observables are measurable quantities *by definition*. What sense could be attributed to an observable, for which no measurement procedure exists? The second point does not concern the possibility of measurements but the validity of theoretical predictions. Similar arguments apply to state preparations. From our point of view, the epistemological status of state preparation and measurements of observable values are both given on the level of the GFS and are, therefore, the same in classical as in quantum theories. The two concepts differ only by the mathematical description of states, observables, and expectation values. There are no 'fundamental' preparation and measurement problems in the GFS and, therefore, neither in classical physics nor in quantum mechanics in its statistical interpretation.

A different topic is the question why a specific state preparation or a measurement of an observable value can be performed with a given apparatus. Such a discussion can be held only on the level of a developed theory. The quantum mechanical model of the Stern-Gerlach experiment, described below, provides an example. It might seem unsatisfactory in this discussion that no detailed theoretical description of diaphragms and impervious screens forming the particle jets or of the particle detector are included in the model. But it is sufficient to confirm *empirically* that a given device provides a suitable state preparation or observable measurement and it may be desirable, but is not necessary, to have a detailed insight how the apparatus works. It suffices to ascertain *empirically* that a diaphragm made from metal sheets sorts out a light ray, that this light ray can be split into its different colours by a glass prism, and that a photographic plate (or the eye!) is a suitable light detector. Such empirical knowledge is, moreover, indispensable for methodological reasons: theories like Maxwell's theory of light or theories concerning the structure of matter and the interaction between matter and light are necessarily based upon empirical knowledge. But only with such theories it is possible, in turn, to have an understanding why glass prisms or photographic plates function as they do.

4.1 Platt's quantum mechanical model for the Stern-Gerlach experiment

Published shortly before the formulation of modern quantum mechanics by Heisenberg and Schrödinger, the Stern-Gerlach experiment [38] was an important impetus to the development of quantum mechanics by

establishing the quantization of intrinsic spin. In the pedagogical literature, this experiment was discussed a long time only by treating the translational motion of the particle classically (see *e.g.* [5, p. 363]) until D. E. Platt [39] provided a genuinely quantum mechanical analysis of an idealized model of the Stern-Gerlach experiment in 1990.

Platt starts with the Schrödinger equation for the spinor wave function of a spin-1/2 particle. The Hamilton operator H contains the familiar coupling term between the magnetic field \mathbf{B} , treated classically, and spin \mathbf{S}

$$H = \frac{\mathbf{P}^2}{2m} + \frac{g\mu_B}{\hbar} \mathbf{B} \cdot \mathbf{S}$$

g is the gyromagnetic ratio, μ_B the Bohr magneton, and $\mathbf{S} = \frac{\hbar}{2} \boldsymbol{\sigma}$ where $\boldsymbol{\sigma}$ denotes the Pauli spin matrices. The incoming particle jet is directed along the x_1 -axis, and the inhomogeneous magnetic field is chosen as

$$\mathbf{B}(\mathbf{x}) = (0, \beta x_2, B_0 - \beta x_3) \quad (4.1)$$

satisfying the conditions $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{B} = \mathbf{0}$ characteristic of a static magnetic field in vacuum. With this Hamiltonian, the Schrödinger equations for the components of the spinor wave function

$$\Psi(\mathbf{x}, t) = \begin{pmatrix} \psi_+(\mathbf{x}, t) \\ \psi_-(\mathbf{x}, t) \end{pmatrix} \quad (4.2)$$

become

$$i\hbar \partial_t \psi_+ = -\frac{\hbar^2}{2m} \nabla^2 \psi_+ - i\frac{\hbar\gamma}{2} x_2 \psi_- + \frac{\hbar}{2} (\omega - \gamma x_3) \psi_+ \quad (4.3)$$

$$i\hbar \partial_t \psi_- = -\frac{\hbar^2}{2m} \nabla^2 \psi_- + i\frac{\hbar\gamma}{2} x_2 \psi_+ - \frac{\hbar}{2} (\omega - \gamma x_3) \psi_-$$

with $\partial_t \equiv \frac{\partial}{\partial t}$ and the substitutions

$$\omega = \frac{g\mu_B B_0}{\hbar} \quad (4.4)$$

$$\gamma = \frac{g\mu_B \beta}{\hbar} \quad (4.5)$$

Platt sets

$$\Psi(t) = U_P(t) \check{\Psi}(t) \quad (4.6)$$

with

$$U_P(t) = \begin{pmatrix} e^{-i\omega t/2} & 0 \\ 0 & e^{+i\omega t/2} \end{pmatrix} \quad (4.7)$$

which yields

$$i\hbar \partial_t \check{\Psi}_+ = -\frac{\hbar^2}{2m} \nabla^2 \check{\Psi}_+ - i e^{i\omega t} \frac{\hbar\gamma}{2} x_2 \check{\Psi}_- - \frac{\hbar}{2} \gamma x_3 \check{\Psi}_+ \quad (4.8)$$

$$i\hbar \partial_t \check{\Psi}_- = -\frac{\hbar^2}{2m} \nabla^2 \check{\Psi}_- + i e^{-i\omega t} \frac{\hbar\gamma}{2} x_2 \check{\Psi}_+ + \frac{\hbar}{2} \gamma x_3 \check{\Psi}_-$$

Consider now time averaged spinor functions

$$\bar{\psi}_{\pm}^{(\tau)}(\mathbf{x}, t) = \int_{-\infty}^{+\infty} dt' w_{\tau}(t - t') \check{\Psi}_{\pm}(\mathbf{x}, t') \quad (4.9)$$

with, *e.g.*, the weight function

$$w_\tau(t) = \frac{1}{\sqrt{2\pi}\tau} e^{-t^2/\tau^2}, \quad \tau > 0 \quad (4.10)$$

By partial integration one obtains

$$\partial_t \bar{\psi}_\pm^{(\tau)}(\mathbf{x}, t) = \int_{-\infty}^{+\infty} dt' w_\tau(t - t') \partial_t \check{\Psi}_\pm(\mathbf{x}, t')$$

and from that the Schrödinger equations

$$i\hbar \partial_t \bar{\psi}_+^{(\tau)} = -\frac{\hbar^2}{2m} \nabla^2 \bar{\psi}_+^{(\tau)} - i\frac{\hbar\gamma}{2} x_2 \int_{-\infty}^{+\infty} dt' w_\tau(t - t') \check{\Psi}_-(t') e^{i\omega t'} - \frac{\hbar}{2} \gamma x_3 \bar{\psi}_+^{(\tau)} \quad (4.11)$$

$$i\hbar \partial_t \bar{\psi}_-^{(\tau)} = -\frac{\hbar^2}{2m} \nabla^2 \bar{\psi}_-^{(\tau)} + i\frac{\hbar\gamma}{2} x_2 \int_{-\infty}^{+\infty} dt' w_\tau(t - t') \check{\Psi}_+(t') e^{-i\omega t'} + \frac{\hbar}{2} \gamma x_3 \bar{\psi}_-^{(\tau)}$$

The integrals over t' in the coupling terms are Fourier integrals where the integrands $w_\tau(t - t') \check{\Psi}_\mp(t')$ feature a continuous first derivative with respect to t' for fixed values of t , τ , and \mathbf{x} , or are smoother, and their modulus is integrable over any finite interval $[t_1, t_2]$. Under these conditions, the Fourier integrals are of the order $O(\omega^{-n})$ with $n \geq 3$ [40, Chapt. 4]. For sufficiently high values of ω , it is possible to find a value of τ such that $\bar{\psi}_\pm^{(\tau)} \approx \check{\Psi}_\pm$ and that the coupling terms become insignificant. The Schrödinger equations Eq. (4.11) then decouple and are approximately valid also for $\check{\Psi}_\pm$.

4.2 Solution in the Heisenberg picture

The solution and proper discussion of the decoupled Schrödinger equations is, however, still involved and Platt takes recourse to Ehrenfest's theorem in order to proceed. This is not really satisfactory. In addition, Platt fails, we believe, to bring the task of discussing the Stern-Gerlach experiment to an end. A suitable basis for such a thorough discussion is provided if Platt's model is solved in the Heisenberg picture. We retain from Platt's analysis that the Hamiltonian

$$\check{H}(t) = \begin{cases} \frac{\mathbf{P}^2}{2m} - \gamma X_3 S_3 & \text{if } 0 \leq t < T \\ \frac{\mathbf{P}^2}{2m} & \text{if } t \geq T \end{cases} \quad (4.12)$$

describes the time evolution of the spinor $\check{\Psi}$ in a good approximation. T may be considered as the time of flight of the particle across the magnetic field (4.1) confined to some x_1 -interval. With this Hamiltonian, the analysis shall now be carried out in the Heisenberg picture where the time evolution is carried by the observables and the states are constant in time.

The time evolution of the spinor $\check{\Psi}$ is given by $\check{\Psi}(t) = \check{U}(t) \Psi(0)$ with

$$\check{U}(t) = \begin{cases} e^{-\frac{i}{\hbar} \check{H}(0)t} & \text{if } 0 \leq t \leq T \\ e^{-\frac{i}{\hbar} \check{H}(T)(t-T)} e^{-\frac{i}{\hbar} \check{H}(0)T} & \text{if } t \geq T \end{cases} \quad (4.13)$$

The transition to the Heisenberg picture follows from the identity

$$\begin{aligned} \langle A \rangle_\rho(t) &= (\Psi(t), A(0) \Psi(t)) \\ &= (U_P(t) \check{\Psi}(t), A(0) U_P(t) \check{\Psi}(t)) \\ &= (U_P(t) \check{U}(t) \Psi(0), A(0) U_P(t) \check{U}(t) \Psi(0)) \\ &= (\Psi(0), \check{U}^{-1}(t) U_P^{-1}(t) A(0) U_P(t) \check{U}(t) \Psi(0)) \\ &\equiv (\Psi(0), A(t) \Psi(0)) \end{aligned} \quad (4.14)$$

valid for all observables A and all spinors $\psi(0)$, yielding

$$A(t) = \check{U}^{-1}(t) U_P^{-1}(t) A(0) U_P(t) \check{U}(t) \quad (4.15)$$

The Hamiltonian \check{H} governs, therefore, the time evolution of all observables but with the time dependent initial condition $U_P^{-1}(t) A(0) U_P(t)$. It is $U_P^{-1}(t) A(0) U_P(t) = A(0)$ for the observables $\mathbf{X}(0)$, $\mathbf{P}(0)$, and $S_3(0)$, diagonal in the spinor indices. The time evolution of these observables is then given by the Heisenberg equation

$$d_t A = \frac{i}{\hbar} [\check{H}, A] \quad (4.16)$$

($d_t \equiv \frac{d}{dt}$), which must be solved separately for the time intervals $0 \leq t < T$ and $t \geq T$ where the Hamiltonian (4.12) is constant. The time evolution conserves the canonical commutation relations $[X_j, X_k] = 0$, $[P_j, P_k] = 0$, $\frac{i}{\hbar} [P_j, X_k] = \delta_{jk} I$, $[X_j, S_k] = 0$, $[P_j, S_k] = 0$, and $\frac{i}{\hbar} [S_j, S_k] = \sum_{l=1}^3 \epsilon_{jkl} S_l$ for $j, k \in \{1, 2, 3\}$, because $[A(t), B(t)] = \check{U}^{-1}(t) U_P^{-1}(t) [A(0), B(0)] U_P(t) \check{U}(t)$.

In the Heisenberg equation, the Hamilton operator \check{H} is considered as a function of the time dependent operators $\mathbf{P}(t)$, $X_3(t)$, and $S_3(t)$. It follows then from the Heisenberg equation and the canonical commutation relations that

$$d_t \mathbf{X}(t) = \frac{1}{m} \mathbf{P}(t), \quad \mathbf{X}(0) = \mathbf{x} \quad (4.17)$$

and

$$S_3(t) = S_3(0), \quad S_3(0) = \frac{\hbar}{2} \sigma_3 \quad (4.18)$$

$$P_j(t) = P_j(0), \quad \mathbf{P}(0) = \frac{\hbar}{i} \nabla \quad (4.19)$$

$$X_j(t) = X_j(0) + \frac{t}{m} P_j(0) \quad (4.20)$$

for $j = 1, 2$ and $t \geq 0$. But it is

$$d_t P_3(t) = \begin{cases} \gamma_3 S_3(0) & \text{if } 0 \leq t < T \\ 0 & \text{if } t \geq T \end{cases} \quad (4.21)$$

and it follows from Eq. (4.21) and (4.17) that

$$P_3(t) = \begin{cases} P_3(0) + \gamma t S_3(0) & \text{if } 0 \leq t \leq T \\ P_3(0) + \gamma T S_3(0) & \text{if } t \geq T \end{cases} \quad (4.22)$$

and

$$X_3(t) = \begin{cases} X_3(0) + \frac{t}{m} P_3(0) + \frac{\gamma}{m} \frac{t^2}{2} S_3(0) & \text{if } 0 \leq t \leq T \\ X_3(T) + \frac{(t-T)}{m} P_3(T) & \\ = X_3(0) + \frac{t}{m} P_3(0) + \frac{\gamma}{m} (Tt - \frac{T^2}{2}) S_3(0) & \text{if } t \geq T \end{cases} \quad (4.23)$$

To summarize the results obtained so far: the quantities P_1 , P_2 , S_3 are constants of motion, the time dependence of X_1 and X_2 is that of free motion, a term proportional to S_3 is added to the free motion for P_3 and X_3 . A further consequence of Eq. (4.16) is that the Hamilton operator is constant in each of the time intervals $0 \leq t < T$ and $t \geq T$ as it is easily verified by direct calculation using (4.12), (4.22), and (4.23).

The time evolution of S_1 and S_2 is evaluated in Appendix D with the result that a spin precession due to the gradient of the magnetic field is superimposed to the precession due to the constant part of the magnetic field.

4.3 Discussion of the Stern-Gerlach experiment

Eq. (4.22) and Eq. (4.23) suffice for a thorough discussion of the Stern-Gerlach experiment. The immediate idea for such a discussion is to consider the expectation values of $P_3(t)$ and $X_3(t)$ for a jet of particles prepared in a state with state operator ρ . In the time interval $0 \leq t \leq T$, they are then equal to

$$\langle P_3(t) \rangle_\rho = \langle P_3(0) \rangle_\rho + \gamma t \langle S_3(0) \rangle_\rho \quad (4.24)$$

$$\langle X_3(t) \rangle_\rho = \langle X_3(0) \rangle_\rho + \frac{t}{m} \langle P_3(0) \rangle_\rho + \frac{\gamma}{m} \frac{t^2}{2} \langle S_3(0) \rangle_\rho \quad (4.25)$$

If ρ is an eigenstate of the S_3 -operator, one of two distinct trajectories appears according to the eigenvalues of $S_3(0)$. Taking the time derivative of the equation resulting from Eq. (4.24) reproduces Eq. (9) of [39]. This equation is valid only for spin eigenstates. But the real feature of the Stern-Gerlach experiment is that the splitting of the incoming particle jet into two jets with given values of S_3 is achieved, *irrespective* of the spin state chosen, be it pure or mixed (for a successful experiment, there are, of course, conditions to be fulfilled by the initial positions and momenta of the particles).

Another shortcoming of Platt's analysis is that the dispersion of P_3 or X_3 is ignored. For a valuable discussion of the Stern-Gerlach experiment, dispersion must be accounted for and it is necessary to show the strict correlation between the result of a momentum or position measurement on a particle and the value of its spin *for each individual measurement*. In the Schrödinger picture, this task necessitates the solution of the decoupled Schrödinger equations (4.11) with appropriate initial values $\check{\psi}_\pm(\mathbf{x}, 0)$ of the spinor wave functions. Then it must be shown that there are disjoint intervals (a_+, b_+) and (a_-, b_-) such that the expectation values $\langle \chi_{(a_+, b_+)}(X_3) \rangle_{\check{\psi}_+(t)}$ and $\langle \chi_{(a_-, b_-)}(X_3) \rangle_{\check{\psi}_-(t)}$ are close to 1 at sufficiently high values of t .

The operator equations Eq. (4.22) and Eq. (4.23) provide a much more direct access. They signify that a measurement of $P_3(t)$ is also a measurement of the observable represented by the right-hand side of Eq. (4.22) with identical result and analogously for $X_3(t)$. The operators $P_3(0)$ and $X_3(0)$ both commute with $S_3(0)$, which means that each of the pairs $P_3(0)$, $S_3(0)$, and $X_3(0) + (t/m) P_3(0)$, $S_3(0)$, possesses a common eigenbasis. A measurement value of the observable $P_3(t)$ is, in the case of $t \geq T$, the sum of the measurement values of the observables $P_3(0)$ and $\gamma T S_3(0)$. Similar considerations apply for the measurement values of $X_3(t)$. Remember that it is *not* generally true that the operator equation $C = A + B$ implies the additivity of the corresponding measurement values (see Sec. 2.5).

Let now

$$\Delta_\rho(A) \equiv \sqrt{\text{Var}(A)_\rho} = \sqrt{\langle A^2 \rangle_\rho - \langle A \rangle_\rho^2} \quad (4.26)$$

denote the dispersion of the observable A in the state ρ and let n be a real number, *e.g.* greater than 10. It follows from Chebyshev's inequality (B.1) that the measured values of $P_3(0)$ lay in the interval

$$\left(\langle P_3(0) \rangle_\rho - \frac{n}{2} \Delta_\rho(P_3(0)), \langle P_3(0) \rangle_\rho + \frac{n}{2} \Delta_\rho(P_3(0)) \right) \quad (4.27)$$

with probability greater or equal than $1 - (2/n)^2$. As a consequence of the considerations above, measured values of $P_3(t)$ lay, for $t \geq T$, with the same probability in the intervals

$$\left(\langle P_3(0) \rangle_\rho - \frac{n}{2} \Delta_\rho(P_3(0)) \pm \frac{\gamma \hbar T}{2}, \langle P_3(0) \rangle_\rho + \frac{n}{2} \Delta_\rho(P_3(0)) \pm \frac{\gamma \hbar T}{2} \right) \quad (4.28)$$

according to the measured values $\pm \hbar/2$ of S_3 . The two intervals are disjoint if

$$\hbar \gamma T = g \mu_B \beta T > n \Delta_\rho(P_3(0)) \quad (4.29)$$

In this case, the value of the spin component $S_3(0)$ can be concluded from the measured momentum of the particle with probability greater or equal than $1 - (2/n)^2$.

As for a measurement of $X_3(t)$, a similar derivation yields the inequality

$$\begin{aligned}
\left(\frac{\hbar\gamma}{m} \left(Tt - \frac{T^2}{2}\right)\right)^2 &> n^2 \Delta_\rho^2 \left(X_3(0) + \frac{t}{m} P_3(0)\right) \\
&= n^2 \left\{ \Delta_\rho^2(X_3(0)) + \left(\frac{t}{m}\right)^2 \Delta_\rho^2(P_3(0)) \right. \\
&\quad \left. + \frac{t}{m} \left(\langle X_3(0)P_3(0) + P_3(0)X_3(0) \rangle_\rho - 2 \langle X_3(0) \rangle_\rho \langle P_3(0) \rangle_\rho \right) \right\}
\end{aligned} \tag{4.30}$$

as a condition for path separation in the case of $t \geq T$. For a sufficiently high value of t , the terms quadratic in t dominate and the intervals corresponding to different values of $S_3(0)$ are again disjoint if inequality Eq. (4.29) holds.

The particles in each of the disjoint particle jets have the same value of the spin component S_3 *i.e.* they form a sample for a pure spin state *irrespective of the initial spin state* — it may even be mixed. The Stern-Gerlach apparatus can be used to prepare particle jets in a pure spin state. The correlation between momentum or position and spin value is a consequence of dynamics and does not depend on the spin state. The state, denoting initial or boundary conditions, plays a role only in so far as the chosen initial conditions for position and momentum of the particle must ensure a later path separation according to the value $+\hbar/2$ or $-\hbar/2$ of the spin component S_3 . Inequality Eq. (4.29) provides a necessary condition for this and design criteria for the experiment. It is, not surprisingly, favourable to keep the product βT big and the dispersion $\Delta_\rho(P_3(0))$ small. The occurrence of dispersions (4.26) and their use in the ensuing argumentation make evident that the statistical interpretation of quantum mechanics underlies. Such a discussion is clearly not conceivable if a pure state is taken as a property of an individual system.

On the other hand, if the incoming particle jet is the output of some experiment and two particle detectors are placed into the paths of the disjoint output jets then the Stern-Gerlach apparatus is a measurement device for the S_3 -observable because the value of S_3 is not changed by the interaction with the magnetic field at least within the approximations used to get the effective Hamiltonian (4.12). The position of the particle is here the discriminating pointer observable. It is obvious that no quantum mechanical model of the interactions between particle and detector must be integrated into the description of the measurement apparatus to render the measurement process conclusive.

The relationship $S_3(t) = S_3(0)$ implies, on the level of individual systems, that the value of this observable, an objective property of the individual system, is constant in time and not changed by the measurement process. It is out of question *“that measured values do not preexist the measurement[, and that] a measurement does not merely ‘read off’ the values, but enacts or creates them by the process itself”* [41, p. 1694].

4.4 ‘Fundamental problems’ in the Copenhagen interpretation

Measurement process and measurement theory are usually discussed in the Schrödinger picture and in a rather generic and abstract manner. A clear presentation and critical assessment of this approach is provided by Ballentine [25, Sec. 9.2, 9.3]. He concludes that the discussion of the measurement process in this manner entails serious difficulties for Copenhagen type interpretations of quantum mechanics.

It is, in fact, characteristic for the Copenhagen interpretation that ‘fundamental problems’ occur. Because a pure quantum state is there ‘attached’ to the single, concrete system, preparation cannot be understood as a separation and filtering process on ensembles as described in Sec. 5.1, but must be treated on the level of the single particle (‘preparation problem’). As an illustration, consider, now in the Schrödinger picture, Platt’s model of the Stern-Gerlach device with initial pure state $\Psi(\mathbf{x}, 0) = \psi(\mathbf{x}, 0) \otimes \chi_1^\pm = \psi(\mathbf{x}, 0) \otimes (\chi_3^+ \pm \chi_3^-)/\sqrt{2}$, where χ_j^\pm is the eigenstate of S_j with eigenvalue $\pm\hbar/2$. According to the decoupled equations (4.11), the initial state of the particle is transformed to $\Psi(\mathbf{x}, t) = (\psi^+(\mathbf{x}, t) \otimes \chi_3^+ \pm \psi^-(\mathbf{x}, t) \otimes$

$\chi_3^-)/\sqrt{2}$. The wave function $\psi(\mathbf{x}, 0)$, the initial value of $\psi^\pm(\mathbf{x}, t)$, shall be chosen in such a way that $\int \psi^+(\mathbf{x}, t)^* \psi^-(\mathbf{x}, t) d^3x \ll 1$ for sufficiently high values of t . A pure S_3 spin state is attained if the 'wrong' half of the wave function is discarded and the 'right' half renormalized. On the operational level, this means that 'wrong' particles are eliminated. The processes of discarding and renormalization cannot be properly described by a unitary transform of the wave function on the level of the individual particle. A recourse to ensembles cannot be avoided; this process is usually called 'transition from the coherent to an incoherent superposition'.

A 'measurement problem' occurs if the observable A with eigenstates $\{\phi_n\}$ is measured on a system in a pure state ψ not being such an eigenstate, which is the generic case. More than one value of the measured observable are then somehow present in the single, concrete system, which becomes so a carrier of contradictory properties. These undesirable consequences do not only hold for microscopic systems but also for the macroscopic measurement devices due to the unitarity of the quantum mechanical time evolution (Schrödinger's cat paradox [6, p. 812]). If one of these values, a_1 say, results from a measurement of this observable, the wave function, a property of the individual system according to Copenhagen doctrine, must consequently equal ϕ_1 after measurement, the transition $\psi \rightarrow \phi_1$ being termed 'the collapse of the wave function'. Such a consideration clearly can be carried out only in the Schrödinger picture. What is 'the collapse of the wave function' in the unitarily equivalent Heisenberg picture with its constant states?

We must forgo a critical acclaim of the vast discussion on this topics. The interested reader may consult [5, Sec. 4], or, for a historical presentation, the treatise of M. Jammer [11, Chap. 11]. We only retain that no such fundamental measurement or preparation problems arise in the statistical interpretation of quantum mechanics. These problems are fictitious, a consequence of the mistaken assumptions of the Copenhagen doctrine, which are, in the first place, the assumption that a pure state is a complete description of an individual system. A second point causing confusion is that no clear distinction is made between state preparation and observable measurement, which entails the unfounded assertion that, after a measurement, the individual system must be in an eigenstate of the measured observable.

Chapter 5

Bohm's version of the Einstein-Podolsky-Rosen thought experiment

In the context of the thought experiment of Einstein, Podolsky, and Rosen [15], David Bohm [13, p. 614 – 619], [14] devised a setting that considerably facilitates discussion and is also the setting used by Bell to formulate his famous theorem (see Chap. 6). Two spin- $\frac{1}{2}$ particles a and b are bound into an unstable state of total spin 0 and disintegrate conserving spin. No physical interactions between the two particles remain after disintegration. We confine ourselves to the description of the spin space of the particles represented by the tensor product $\mathbb{C}^2 \otimes \mathbb{C}^2$. The spin component in direction \mathbf{e} is given by $(\mathbf{e}, \boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ denote the Pauli spin matrices (3.21). The spin variables of the particles a and b have the form $(\mathbf{e}^a, \boldsymbol{\sigma}) \otimes I$ and $I \otimes (\mathbf{e}^b, \boldsymbol{\sigma})$, the total spin of the system is $\mathbf{S} = \boldsymbol{\sigma} \otimes I + I \otimes \boldsymbol{\sigma}$. The state ψ_0 of total spin 0 is given by

$$\psi_0 = \sqrt{\frac{1}{2}} (\psi_+ \otimes \psi_- - \psi_- \otimes \psi_+) \quad (5.1)$$

$\{\psi_+, \psi_-\}$ being some orthonormal basis of \mathbb{C}^2 . A unitary transform of this basis conserves the form of ψ_0 , expression (5.1) is valid for every such orthonormal basis.

The expectation value of the spin variables of particle a in the state ψ_0 equals

$$\langle \boldsymbol{\sigma} \otimes I \rangle_{\psi_0} = (\psi_0, \boldsymbol{\sigma} \otimes I \psi_0) = \frac{1}{2} ((\psi_+, \boldsymbol{\sigma} \psi_+) + (\psi_-, \boldsymbol{\sigma} \psi_-)) = \frac{1}{2} \text{Tr}(\boldsymbol{\sigma}) = \mathbf{0} \quad (5.2)$$

The spin state of each a -particle after disintegration, taken as a separate, free system, is thus described by the state operator

$$\rho_0 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (5.3)$$

and likewise for the spin state of particle b . For an arbitrary direction \mathbf{e} , ρ_0 is the sum $\frac{1}{2} (Q(\mathbf{e}) + Q(-\mathbf{e}))$, where $Q(\mathbf{e})$ projects on the eigenvector of the spin component $(\mathbf{e}, \boldsymbol{\sigma})$ with eigenvalue +1.

The operator describing the correlation of the spin components $(\mathbf{e}^a, \boldsymbol{\sigma}) \otimes I$ and $I \otimes (\mathbf{e}^b, \boldsymbol{\sigma})$ is given by the product

$$((\mathbf{e}^a, \boldsymbol{\sigma}) \otimes I) (I \otimes (\mathbf{e}^b, \boldsymbol{\sigma})) = (\mathbf{e}^a, \boldsymbol{\sigma}) \otimes (\mathbf{e}^b, \boldsymbol{\sigma}) \quad (5.4)$$

In terms of measurement values, it is, for every instance n of a measurement series,

$$m((\mathbf{e}^a, \boldsymbol{\sigma}) \otimes (\mathbf{e}^b, \boldsymbol{\sigma}))_n = m((\mathbf{e}^a, \boldsymbol{\sigma}))_n m((\mathbf{e}^b, \boldsymbol{\sigma}))_n \quad (5.5)$$

because the operators $(\mathbf{e}^a, \boldsymbol{\sigma}) \otimes I$ and $I \otimes (\mathbf{e}^b, \boldsymbol{\sigma})$ commute. Both operators can, therefore, be considered as real functions of a third operator and Eq. (2.2) applies. The product formula (5.5) indicates how to measure the correlation operator (5.4) and corresponds to the operational definition of the correlation of two observables A and B as the limit $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N m(A)_n m(B)_n$.

The expectation value of the correlation operator (5.4) in the state ψ_0 equals

$$\langle (\mathbf{e}^a, \boldsymbol{\sigma}) \otimes (\mathbf{e}^b, \boldsymbol{\sigma}) \rangle_{\psi_0} = (\psi_0, (\mathbf{e}^a, \boldsymbol{\sigma}) \otimes (\mathbf{e}^b, \boldsymbol{\sigma}) \psi_0) = -(\mathbf{e}^a, \mathbf{e}^b) \equiv -\cos(\theta) \quad (5.6)$$

and its variance

$$\text{Var}((\mathbf{e}^a, \boldsymbol{\sigma}) \otimes (\mathbf{e}^b, \boldsymbol{\sigma}))_{\psi_0} = 1 - \cos^2(\theta) = \sin^2(\theta) \quad (5.7)$$

because $((\mathbf{e}^a, \boldsymbol{\sigma}) \otimes (\mathbf{e}^b, \boldsymbol{\sigma}))^2 = I \otimes I$.

The variance vanishes for $\theta = 0$ and $\theta = \pi$ *i.e.* ψ_0 is an eigenstate of $(\mathbf{e}, \boldsymbol{\sigma}) \otimes (\mathbf{e}, \boldsymbol{\sigma})$ with eigenvalue -1 for any direction \mathbf{e} . Or in terms of the Chebyshev inequality (B.1): it is $P_{\psi_0}(|(\mathbf{e}, \boldsymbol{\sigma}) \otimes (\mathbf{e}, \boldsymbol{\sigma}) + 1| \geq \epsilon) = 0$ for all $\epsilon > 0$: the probability measure $P_{(\mathbf{e}, \boldsymbol{\sigma}) \otimes (\mathbf{e}, \boldsymbol{\sigma}), \psi_0}(\cdot)$ is concentrated onto the point -1 . Eq. (5.5) yields then $m((\mathbf{e}, \boldsymbol{\sigma}) \otimes (\mathbf{e}, \boldsymbol{\sigma}))_n \equiv -1$ for any direction \mathbf{e} . There is a strict anticorrelation between the measured values of the same spin components of particles a and b . Quantum mechanics provides, therefore, the following prediction: *If the spin component $(\mathbf{e}, \boldsymbol{\sigma})$ is measured on particle a of a system of two particles, prepared in the state ψ_0 , and the measurement value equals $+1$, then a measurement of the same spin component on particle b always yields the value -1 , and analogously if the value -1 is measured on particle a .* This is a statement about a connection between *measured values*, and not plainly between observable values. The statement is void if no spin measurement is performed on particle a . This contrasts with the situation where particle b is prepared in the eigenstate of $(\mathbf{e}, \boldsymbol{\sigma})$ with eigenvalue -1 . There, the definite quantum mechanical prediction needs no recourse to measurements on another particle.

5.1 Consequences of Bohm's setting

Let us now discuss in some detail the transition from the total system after disintegration to its separated, free parts. That the two particles are considered as separated means that the system of the two particles is described solely by observables of the form $A \otimes I$ and $I \otimes B$. Observables corresponding to correlations between the particles are not contained in this set. Correlations may nevertheless exist but they are ignored and they can be ignored exactly if there are no physical interactions between the particles *i.e.* if the Hamiltonian of the combined system has the form $H^a \otimes I + I \otimes H^b$, where H^a determines the time evolution of particle a after disintegration. The set of observables $\{A \otimes I, I \otimes B\}$ does not suffice to separate all states of the tensor product space $\mathcal{H}^a \otimes \mathcal{H}^b$. These states are divided, according to the separation postulate (SI) of Sec. 2.4, into classes. Every state in such a class yields the same expectation value for an observable in the considered reduced set of observables. Each class contains an element of the form $\rho^a \otimes \rho^b$ which is uniquely determined if the observable sets $\{A \otimes I\}$ and $\{I \otimes B\}$ separate all pure states of the Hilbert spaces \mathcal{H}^a and \mathcal{H}^b . The state operators ρ^a and ρ^b denote the states of the particles a and b if they are taken as separated. With respect to Bohm's setting, this means that the spin states of the separated particles a and b are both given by the state operator ρ_0 , Eq. (5.3). A measurement of any spin component $(\mathbf{e}, \boldsymbol{\sigma})$ yields the values $+1$ and -1 with equal probability $\frac{1}{2}$.

On the other hand, the two disintegrated particles may still be considered as parts of a total system, in which spin correlations persist. A sequence of such correlated particles a and b moving in the $-x_1$ - and $+x_1$ -direction, respectively, shall be prepared in an experiment. The σ_3^a component is measured on every a -particle yielding a sequence of measurement values $m(\sigma_3^a)_1, m(\sigma_3^a)_2, \dots, m(\sigma_3^a)_n, \dots, m(\sigma_3^a)_N$. Measurements of σ_3 on the corresponding b -particles give the sequence $m(\sigma_3^b)_1, m(\sigma_3^b)_2, \dots, m(\sigma_3^b)_n, \dots, m(\sigma_3^b)_N$. Quantum theory provides two predictions. First, each of the two sequences is a sample of the

probability distribution generated by the state operator ρ_0 *i.e.* the values $+1$ and -1 must occur in a random fashion with frequencies $\frac{1}{2}$ in the long run. Second, the product of corresponding measurement values $m(\sigma_3^a)_n m(\sigma_3^b)_n = -1$ for each value n due to the strict anticorrelation. The two sequences may serve to test these two theoretical predictions.

If these theoretical predictions are accepted, the measurement sequence $\{m(\sigma_3^a)_n\}$ on the a -particles and anticorrelation can be used to prepare subsequences of b -particles being samples for the eigenstates of the σ_3^b observable. The preparation of these samples is done by *separation* and *filtering*. The total sequence of b -particles forms a sample belonging to the state operator ρ_0 with the mentioned distribution of the measurement values $m(\sigma_3^b)_n$. Owing to the strict anticorrelation, the sequence of measured values $\{m(\sigma_3^a)_n\}$ on the a -particles allows to sort out the b -particles into two subsamples which form samples for the eigenstates of σ_3^b to the eigenvalues $+1$ and -1 (*separation*). The desired subsample is then obtained by eliminating the particles with the undesired measurement values (*filtering*). This filtering may be achieved *e.g.* by putting an absorbing screen into the path of the b -particles at suitable times or, without machinery and without physically interfering with the b -particles, by mere bookkeeping using the total sequence of b -particles in an experiment but subdividing the results of the measurements on the b -particles into two groups according to the corresponding measurement values of the a -particles. With the experimental possibility to flip spin direction, the information provided by the spin measurements on particles a may be used to turn the spins of the b -particles if necessary and to double the intensity of the particle jet belonging to a pure eigenstate of σ_3^b . In order to enable these procedures, a signal has to be sent from each experiment on a particle a to the corresponding experiment on particle b transmitting the result of the σ_3^a -measurement. Causality and locality are respected thereby.

If, instead of the σ_3 -component, the σ_2 -component is measured on the a -particles, nothing has changed for the the series of b -particles. It still forms a sample whose spin state is given by the state operator ρ_0 , and the measurement series of the σ_3 -components of the b -particles do not depend on which spin component is measured on the a -particles, or if a measurement on the particles a is performed at all. However, it is now possible to decompose this sample into two subsamples belonging to eigenstates of the σ_2^b -observable and to prepare a sequence of b -particles to a pure σ_2 -state by separation and filtering. The experimenter obviously has to decide according to which property of the b -particles, σ_2 or σ_3 , he wants to sort out. The two separations result in different subsamples belonging to different pure states which cannot be decomposed further.

Equal considerations apply to any spin component $(\mathbf{e}, \boldsymbol{\sigma})$, \mathbf{e} being perpendicular to the x_1 -direction. This provides the remarkable possibility to prepare, by transmitting the results of spin measurements on the a -particles, a jet of b -particles to any desired spin eigenstate without performing spin measurements on the b -particles. This is a consequence, first, of the symmetry properties of the spin state ψ_0 expressed by Eq. (5.1), second, of the strict anticorrelation between the spins of the two particles prepared into the singlet state ψ_0 , the anticorrelation being established before the separation of the two particles has taken place, third, of the conservation of all individual spin values for free particles, and, fourth, of the property of the state ρ_0 of the b -particles, considered as isolated, to be, for an arbitrary direction \mathbf{e} , equal to the sum $\frac{1}{2}(Q(\mathbf{e}) + Q(-\mathbf{e}))$, $Q(\mathbf{e})$ projecting on the eigenvector to the eigenvalue $+1$ of the spin component $(\mathbf{e}, \boldsymbol{\sigma})$. There is no necessity to assume any interaction at a distance between the free particles a and b .

Note that the possibilities of state preparation are restricted here not of observable measurements. In the special setting of Bohm, it is possible to simultaneously measure σ_2 and σ_3 : due to the strict anticorrelations, we may consider a measurement of σ_3 on particle a also as a measurement of the value of σ_3 on particle b and can, in addition, measure the value of σ_2 on this particle.¹ Likewise, an apparatus can be devised to simultaneously measure the position and the momentum of a particle with any desired precision [5, p. 365].

¹The values $m(\sigma_2^b)$ and $m(\sigma_3^b)$ thus obtained refer to the time just before the measurement of σ_2^b . It is obviously not possible to prepare a state with sharp values of σ_2^b and σ_3^b with this procedure because the value of σ_3^b is changed by the interaction with the magnetic field generated for the measurement of σ_2^b .

The Heisenberg uncertainty relations are restrictions for the preparation of states: it is not possible to prepare a physical system in such a way that the values of two noncommuting observables can be simultaneously sharply predicted. But this does not mean that "*the electron, at any instance of time, does not have a definite position and velocity. [Physicists] . . . convert what is not measurable by them into the unreal and the nonexistent*" [42, p. 300]. By the way, simultaneous measurements of the two noncommuting observables in question are not required for a test of the uncertainty relations. Repeating the experiment many times and measuring the values of one observable allows to determine the dispersion of this observable provided that it can be measured with a precision clearly exceeding its dispersion. In a second series of experiments, the dispersion of the other observable is determined.

The discussion of Bohm's experimental setting is effected here within the scheme of the statistical interpretation of quantum mechanics, where state preparation can be done by separation and filtering on an ensemble, and this kind of discussion is feasible only within the scheme of the statistical interpretation. The 'preparation problem' of the Copenhagen interpretation appears here in an especially glaring light. If a pure state is understood as a property 'pinned' to a concrete, material system, the processes of separation and filtering are then obviously not possible and it must be admitted that the *b*-particle is in a pure spin state after the measurement of a value of a spin component on particle *a*. Which pure state depends on which spin component was measured on particle *a* and on the result of this measurement although, admittedly, no physical interactions between the two particles exist at that time. This state of affairs was denoted by Einstein as '*Spooky action at a distance*' [43, p. 158], and is responsible for much of the mystery woven around the so-called entangled states.

This difficulty seems to originate from the fact that the two particles, free at the moment of the spin measurement on particle *a* were once in interaction and correlations persist. To assume quantum mechanics being nonlocal or contextual provides a way out if one wants to maintain the Copenhagen interpretation. Fostered by the authority of Bell, the assertion that quantum mechanics is a nonlocal theory gets some support. However, as shown in Chap. 6, quantum nonlocality cannot be deduced from Bell's theorem, nor from an analysis of measured values of some cleverly designed thought experiment. In addition, the idea of nonlocality, if properly reasoned out, is hardly attractive. The possibility to isolate a system independently of its history, and to prepare it according to well defined initial and boundary conditions independently of its past is a fundamental methodological principle and prerequisite of physical theory building, and constitutes the epistemic character of physical statements in an essential manner. For methodological reasons, these presuppositions cannot be cancelled in a later stage of theory formation. One cannot dispense with the possibility of realizing isolated systems in physics neither practically nor theoretically.

Chapter 6

Bell's theorem and the question of locality

In a seminal paper, entitled '*On the Einstein Podolsky Rosen Paradox*', J. S. Bell [17] shows that certain quantum mechanical predictions cannot be met by local hidden variables theories. In the introduction, he states that *it is the requirement of locality, or more precisely that the result of a measurement on one system be unaffected by operations on a distant system with which it has interacted in the past, that creates the essential difficulty*. This statement exerted a considerable influence on the subsequent discussions up to now. It is often ignored in these discussions that Bell's conclusion concerns hidden variables theories that ought to reproduce statistical predictions of quantum mechanics *not* quantum mechanics itself.

6.1 Bell's inequality and Bell's theorem

Bell considers Bohm's setting of the Einstein - Podolsky - Rosen experiment, but his interest is focussed on spin correlations. The elegance and clarity of Bell's exposition of the topic can hardly be surpassed. We, therefore, use Bell's formulation [17, Sec. II] to this end (for consistency with our mathematical notations).

With the example advocated by Bohm and Aharonov [14], the EPR argument is the following. Consider a pair of spin one-half particles formed somehow in the singlet spin state and moving freely in opposite directions. Measurements can be made, say by Stern-Gerlach magnets, on selected components of the spins σ_1 and σ_2 . If measurement of the component (\mathbf{a}, σ_1) , where \mathbf{a} is some unit vector, yields the value $+1$ then, according to quantum mechanics, measurement of (\mathbf{a}, σ_2) must yield the value -1 and vice versa. Now we make the hypothesis [of locality], and it seems one at least worth considering, that if the two measurements are made at places remote from one another the orientation of one magnet does not influence the result obtained with the other. Since we can predict in advance the result of measuring any chosen component of σ_2 , by previously measuring the same component of σ_1 , it follows that the result of any such measurement must actually be predetermined. Since the initial quantum mechanical wave function does not determine the result of an individual measurement, this predetermination implies the possibility of a more complete specification of the state.

Let this more complete specification be effected by means of parameters λ . It is a matter of indifference in the following whether λ denotes a single variable or a set, or even a set of functions, and whether the variables are discrete or continuous. However, we write as if λ were a single continuous parameter. The result A of measuring (\mathbf{a}, σ_1) is then determined by \mathbf{a} and λ , and the result B of measuring (\mathbf{b}, σ_2) in the same instance is determined by \mathbf{b} and λ , and

$$A(\mathbf{a}, \lambda) = \pm 1, \quad B(\mathbf{b}, \lambda) = \pm 1 \quad (6.1)$$

The vital assumption ... is that the result B for particle 2 does not depend on the setting \mathbf{a} of the magnet for particle 1, nor A on \mathbf{b} .

If $\rho(\lambda)$ is the probability distribution of λ then the expectation value of the product of the two components (\mathbf{a}, σ_1) and (\mathbf{b}, σ_2) is

$$C(\mathbf{a}, \mathbf{b}) = \int A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda) \rho(\lambda) d\lambda \quad (6.2)$$

This should equal the quantum mechanical expectation value, which for the singlet state $[\psi_0, \text{Eq. (5.1)}]$ is

$$\langle (\mathbf{a}, \sigma_1) \otimes (\mathbf{b}, \sigma_2) \rangle_{\psi_0} = -(\mathbf{a}, \mathbf{b}) \quad (6.3)$$

But it will be shown that this is not possible.

For the sequel, we follow a later publication of Bell [44, p. 156 – 157] and the presentation by Ballentine [25, 20.3] where a more general inequality is derived under weaker assumptions. Here, two alternative settings \mathbf{a} and \mathbf{a}' are considered for the first Stern-Gerlach magnet and two settings \mathbf{b} and \mathbf{b}' for the second. Then

$$\begin{aligned} |C(\mathbf{a}, \mathbf{b}) - C(\mathbf{a}, \mathbf{b}')| &= \left| \int (A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda) - A(\mathbf{a}, \lambda) B(\mathbf{b}', \lambda)) \rho(\lambda) d\lambda \right| \\ &= \left| \int A(\mathbf{a}, \lambda) B(\mathbf{b}', \lambda) (1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}, \lambda)) \rho(\lambda) d\lambda \right. \\ &\quad \left. - \int A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda) (1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}', \lambda)) \rho(\lambda) d\lambda \right| \\ &\leq \int (1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}, \lambda)) \rho(\lambda) d\lambda + \int (1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}', \lambda)) \rho(\lambda) d\lambda \\ &= 2 \pm (C(\mathbf{a}', \mathbf{b}) + C(\mathbf{a}', \mathbf{b}')) \end{aligned}$$

or

$$|C(\mathbf{a}, \mathbf{b}) - C(\mathbf{a}, \mathbf{b}')| + |C(\mathbf{a}', \mathbf{b}) + C(\mathbf{a}', \mathbf{b}')| \leq 2 \quad (6.4)$$

Besides Eq. (6.2) only the conditions $|A(\mathbf{a}, \lambda)| \leq 1$ and $|B(\mathbf{b}, \lambda)| \leq 1$ have been used. Bell's original inequality, Eq. (15) of Ref. [17], follows from inequality (6.4) by setting $\mathbf{b}' = \mathbf{c}$, $\mathbf{a}' = \mathbf{b}$, and by assuming strict anticorrelation, $C(\mathbf{b}, \mathbf{b}) = -1$, an assumption not necessary to derive (6.4).

Contradictions with quantum mechanical predictions become manifest if the expectation values (6.3) are inserted into inequality (6.4) yielding

$$|(\mathbf{a}, \mathbf{b} - \mathbf{b}')| + |(\mathbf{a}', \mathbf{b} + \mathbf{b}')| \leq 2 \quad (6.5)$$

For given unit vectors \mathbf{b} and \mathbf{b}' , the lefthand side is maximal, if \mathbf{a} is chosen parallel to $\mathbf{b} - \mathbf{b}'$ and \mathbf{a}' parallel to $\mathbf{b} + \mathbf{b}'$, and is then equal to $\sqrt{2} (\sqrt{1 - (\mathbf{b}, \mathbf{b}')} + \sqrt{1 + (\mathbf{b}, \mathbf{b}')})$. This expression attains its maximum value, $2\sqrt{2}$, for $(\mathbf{b}, \mathbf{b}') = 0$ surpassing the limit 2 of inequality (6.4) by more than 40%. If \mathbf{a}' is chosen equal to \mathbf{b} , Bell's original assumption, the maximum value of the lefthand side of (6.5) is attained for \mathbf{a} being parallel to $\mathbf{b} - \mathbf{b}'$ and $(\mathbf{b}, \mathbf{b}') = \frac{1}{2}$ and equals 2.5.

In his Conclusion, Bell comments as follows: *In a theory in which parameters are added to quantum mechanics to determine the results of individual measurements, without changing the statistical predictions, there must be a mechanism whereby the setting of one measuring device can influence the reading of another instrument, however remote. Moreover, the signal involved must propagate instantaneously, so that such a theory could not be Lorentz invariant.* This statement is *Bell's theorem*. It states that the quantum mechanical correlations of Bohm's setting cannot be reproduced by a local hidden variables theory.

This result appears surprising in view of the simple and seemingly natural assumptions (6.1) and (6.2). Note, however, that Bell's model is a classical theory in the sense of Sec. 2.5, where the spin observables are represented by the real functions $A(\mathbf{a}, \lambda)$, $B(\mathbf{b}, \lambda)$ and the correlation operator by the product $A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda)$, where a state is represented by the probability density $\rho(\lambda)$, and where expectation values are given by

expressions of the form (6.2) or, more generally, of (2.10). It is shown in Sec. 3.4 that *no* such classical hidden variables model exists for the set of the quantum mechanical observables $(\mathbf{a}, \sigma) \otimes I, I \otimes (\mathbf{b}, \sigma), (\mathbf{a}, \sigma) \otimes (\mathbf{b}, \sigma)$ irrespective of whether it reproduces quantum mechanical expectation values or not. The reason is that basic structural relationships like $\sigma_1 \sigma_2 = -\sigma_2 \sigma_1$ cannot be represented by a classical model. Bell's model does not describe the system of two spin- $\frac{1}{2}$ particles, contrary to Bell's assumption, because there is no such model. Bell's inequality (6.4) does, therefore, not concern spin-spin correlations, and inequality (6.5) is futile.

Bell's correlation model is a model of its own, independent of a quantum mechanical description except of being motivated by a quantum mechanical system, and unspecific to a degree that should arouse suspicion. None of the symmetry properties characteristic of spin-spin correlations, $(-\mathbf{a}, \sigma) \otimes I = -((\mathbf{a}, \sigma) \otimes I)$ implying $A(-\mathbf{a}, \lambda) = -A(\mathbf{a}, \lambda)$, $C(\mathbf{b}, \mathbf{a}) = C(\mathbf{a}, \mathbf{b})$, $C(\mathbf{a}, \mathbf{a}) = -1$, is used for the derivation of Bell's inequality (6.4), and the latter is valid for a classical model of any families of observables $\{(A(a)/a \in \mathbb{I}_a\}$ and $\{B(b)/b \in \mathbb{I}_b\}$ where the spectra of all $A(a)$ and $B(b)$ are contained in the interval $[-1, +1]$ and $\mathbb{I}_a, \mathbb{I}_b$ are arbitrary index sets (if a classical model exists at all).

It is astonishing that, despite the overwhelming success of quantum mechanics from a theoretical point of view as well as from general experimental evidence, the different bearing of these two schemes was scarcely noticed and taken as an argument against Bell's model or at least as an incentive to question its foundations. Experimental tests were considered important and a series of experiments was devised to decide between Bell's correlation model and quantum mechanics starting with the work of Clauser *et al.* [45]. With considerable sagacity, numerous physical and metaphysical loopholes of the experimental setting were closed in the progress of these experiments (see *e.g.* [46, 47, 48]). But there is no confirmed experiment deciding against quantum mechanics and in favour of Bell's inequality. The logical consequence of this finding is that Bell's correlation model for Bohm's setting, based on assumptions (6.1) and (6.2), is falsified by experiment, and that no conclusions should be drawn out of it.

In view of all these facts, the importance of Bell's model as a useful physical theory is doubtful. The crucial point to state is not that any hidden variables theory reproducing quantum mechanical predictions would necessarily be nonlocal, unfortunately altered later on by Bell to the wrong assertion that *... the quantum correlations are locally inexplicable* [44, p.153], but that there are no classical hidden variables theories for Bohm's setting at all. 'Nonlocality' is, therefore, not the real issue, the less 'quantum nonlocality'. However, it is Bell's accomplishment to have shown that local hidden variables theories have testable consequences that may conflict with quantum mechanical predictions. He incited thereby a renewed interest in hidden variables theories and controversial discussions which helped to clarify the epistemical status of quantum theory.

6.2 Bell's theorem without inequalities and probabilities

In the course of these discussions, variants of Bell's theorem have been invented that dispense with inequalities and probabilities and concerne only perfect correlations rather than statistical ones, but maintain the aim to show that there is a problem with quantum mechanics and the principle of locality which was expressed by Einstein in following terms: *But on one supposition we should, in my opinion, absolutely hold fast: the real factual situation of the system S_2 is independent of what is done with the system S_1 , which is spatially separated¹ from the former* [1, p. 85]. The idea is due to Greenberger, Horne, and Zeilinger and presented in a paper written together with A. Shimony [49]. We discuss a version given by Mermin [16].

Mermin argues on the basis of his ten spin variables (3.33). They are now interpreted as spin components

¹'Spatially separated' understood in the sense of relativity theory: S_1 and S_2 lay in space-time regions that cannot be connected with light rays.

of three spin- $\frac{1}{2}$ particles forming a physical system initially confined into a small region and prepared in a simultaneous eigenstate of the four observables A_1, A_2, A_3, A_4 . The system then disintegrates into three isolated particles conserving the initial spin state. As soon as the three particles are sufficiently apart so that the mutual spin interactions have ceased the individual spin variables are constant.

Let ψ denote such a simultaneous eigenstate:

$$A_k = a_k \psi, \quad k = 1, 2, 3, 4 \quad (6.6)$$

Obviously, it is

$$a_k = \pm 1, \quad a_1 a_2 a_3 a_4 = -1 \quad (6.7)$$

because of the definitions of the four mutually commuting observables A_k and of the equation (3.34I). In terms of measurement values for a system prepared in the state ψ and with the definitions of A_1, \dots, A_4 as products of commuting operators, the following system of equations is obtained (see Sec. 2.4)

$$\begin{aligned} \text{(i)} \quad & m(\sigma_3^a) m(\sigma_3^b) m(\sigma_3^c) = m(A_1) = a_1 \\ \text{(ii)} \quad & m(\sigma_1^a) m(\sigma_1^b) m(\sigma_3^c) = m(A_2) = a_2 \\ \text{(iii)} \quad & m(\sigma_1^a) m(\sigma_3^b) m(\sigma_1^c) = m(A_3) = a_3 \\ \text{(iv)} \quad & m(\sigma_3^a) m(\sigma_1^b) m(\sigma_1^c) = m(A_4) = -a_1 a_2 a_3 \end{aligned} \quad (6.8)$$

All these measurement values are equal to ± 1 .

So far, only the rules of quantum mechanics have been applied. To conclude the argument, the principle of locality is involved stating, for instance, that the measurement value $m(\sigma_1^a)$ must be the same irrespective whether, in addition, σ_1^b and σ_3^c are measured, line (ii) of Eq. (6.8), or σ_3^b and σ_1^c , line (iii). Applying this principle to each of the six spin variables seduces to the conclusion that a unique value must be attributed to each of the six spins. The contradiction is then readily obtained by forming the the product of lines (i) to (iv) of Eq. (6.8). The righthand side of the product equals -1 , but the lefthand side is equal to $+1$ because each measurement value appears twice in the product.

It seems that the locality principle directly conflicts here with quantum mechanics, not only with hidden variables theories reproducing quantum mechanical predictions. However, the quantities $m(\sigma)$ in Eq. (6.8) are not definite spin values because ψ is *not* an eigenstate of the spins (see below), but Eq. (6.8) are quantum mechanical predictions concerning correlations between *measured* spin values. (6.8i) should, therefore, be read as "If measurements of the observables σ_3^a, σ_3^b , and σ_3^c were performed with results $m(\sigma_3^a), m(\sigma_3^b), m(\sigma_3^c)$, then the product of these measured values equals a_1 ". No contradiction can arise within quantum mechanics. For only three of the six spin variables can be measured at once with the above experimental setting. Only *one* of the four equations (6.8) is then involved. A second experiment is needed to measure the other three spins. But the values of the first triple in the second experiment may deviate from their values in the first experiment. One never gets a coherent set of six *really measured values* for which all identities of Eq. (6.8) simultaneously would apply.

How does Mermin's setting present itself in the statistical interpretation of quantum mechanics? The situation is obviously similar to the one in Bohm's setting. In Appendix C, a common eigenbasis of the observables A_1, \dots, A_4 is calculated, and it is shown that the spin state of the particles a, b, c , considered as isolated, is given by the state operator (5.3), $\rho_0 = \frac{1}{2} I$, for all vectors of this eigenbasis. The measured values of a spin component $m(\sigma_k^\xi)_n$ with $n \in \{1, \dots, N\}$, $k \in \{1, 3\}$, and $\xi \in \{a, b, c\}$ in a series of equally prepared particle triplets form, as in the setting of Bohm, a random sequence of $+1$ and -1 whose mean value goes to 0 as N goes to infinity. The equations (6.8) may be used either to test the asserted correlations or, in a series of experiments, to prepare subseries of particles belonging to the eigenstates of a spin variable σ_k . Assume, for instance, that the σ_1 -components of particles a and b are measured in a series of identically prepared particle triplets. The identity given by line (ii) then allows to split the series of particles c into two

subseries corresponding to the eigenstates of the observable σ_3^c by mere bookkeeping without touching the c -particles. The other equations of (6.8) are of no importance in this case.

Mermin assigns values to all six spin observables appearing in Eq. (6.8). He does not make clear whether he considers these values as given by some theory or as given by measurements. As for the first interpretation, remember that physical theories provide statements about probability distributions of observable values, and about observable values only if the system is in a eigenstate of the corresponding observable. To assign definite values to all six spin observables implicitly assumes that the particles are prepared in a simultaneous eigenstate of these six observables, which is possible for classical theories but not in quantum mechanics; in the present example, the spins are in the state where the predictions of all spin values are maximally uncertain. According to the Kochen-Specker theorem, there is no classical theory for this physical system.

It remains the second interpretation: to assume the existence of six simultaneously available measurement values is an act of *counterfactual definiteness*. This concept goes back to H. P. Stapp [50] and may be characterized as *term ... to describe the character of statements about "what would have occurred" in a measurement that we could have performed, but did not actually perform* [25, p. 608], and in which the same status is given to these hypothetical measurement values as to the really measured ones. It is counterfactual because, in the example given, only *three* spin component out of the set of mutually noncommuting spin components of particle a , b , c can be measured at the same instance in an actual series of experiments. Conclusions should be drawn only from the *one* spin correlation that can be calculated from the actually measured spins. All the other, not measured spin correlations are mere possibilities, not available in reality.

In the statistical interpretation of quantum mechanics, an observable is considered as a random variable in some probability space. It is characteristic for a random variable that a random experiment is needed to reveal its actual value out of the set of all possible values and that these other values have no relevance beyond the fact of being possible. Quantum mechanical predictions should be understood exactly in this sense. The Good Lord seems to play at dice, indeed, even if Einstein did not like the idea².

6.3 Quantum Nonlocality?

Elaborating an idea due to P. H. Eberhard [53], H. P. Stapp takes it upon himself to show, in a paper entitled '*Quantum Nonlocality*', that *the validity of the predictions of quantum theory in certain spin-correlation experiments entails a violation of Einstein's locality idea that no causal influence can act outside the forward light cone* [54, p. 427]. The physical situation considered is that of Bohm's experimental setting with measurement sequences of spin components $\sigma^{a'}$ or $\sigma^{a''}$ on particle a , and $\sigma^{b'}$ or $\sigma^{b''}$ on particle b , where

$$\begin{aligned}\sigma^{a'} &\equiv (\mathbf{e}^{a'}, \boldsymbol{\sigma}) & \text{with } \mathbf{e}^{a'} &= (1, 0, 0) \\ \sigma^{a''} &\equiv (\mathbf{e}^{a''}, \boldsymbol{\sigma}) & \text{with } \mathbf{e}^{a''} &= (0, 1, 0) \\ \sigma^{b'} &\equiv (\mathbf{e}^{b'}, \boldsymbol{\sigma}) & \text{with } \mathbf{e}^{b'} &= (1, 0, 0) \\ \sigma^{b''} &\equiv (\mathbf{e}^{b''}, \boldsymbol{\sigma}) & \text{with } \mathbf{e}^{b''} &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)\end{aligned}\tag{6.9}$$

The system of the two particles is prepared into the spin state ψ_0 given by Eq. (5.1). From the operational definition (5.4) of correlations in terms of measured values and from the quantum mechanical expectation value, Eq. (5.6), of the spin correlation operator in the state ψ_0 , the following equations for the four possible

²In a letter to Max Born, Einstein wrote [51, p. 97]: *Die Quantenmechanik ist sehr achtunggebietend. Aber eine innere Stimme sagt mir, daß das noch nicht der wahre Jakob ist. Die Theorie liefert viel, aber dem Geheimnis des Alten bringt sie uns kaum näher. Jedenfalls bin ich überzeugt, daß der nicht würfelt* (Quantum mechanics is certainly imposing. But an inner voice tells me that this is not yet the real thing. The theory says a lot, but does not really bring us nearer to the secret of the "old one". I, at any rate, am convinced that He does not throw dice [letter of December 4, 1926, translation by Irene Born [52]]).

spin-spin correlations result

$$\begin{aligned}
m(\sigma^{a'})_n m(\sigma^{b'})_n &= -1 \quad \text{for } n = 1, \dots, N & (a) \\
\left| \frac{1}{N} \sum_{n=1}^N m(\sigma^{a'})_n m(\sigma^{b''})_n + \frac{1}{\sqrt{2}} \right| &= \frac{1}{\sqrt{2}} \epsilon_{2,N} \geq 0 & (b) \\
\left| \frac{1}{N} \sum_{n=1}^N m(\sigma^{a''})_n m(\sigma^{b'})_n \right| &= \epsilon_{3,N} \geq 0 & (c) \\
\left| \frac{1}{N} \sum_{n=1}^N m(\sigma^{a''})_n m(\sigma^{b''})_n + \frac{1}{\sqrt{2}} \right| &= \frac{1}{\sqrt{2}} \epsilon_{4,N} \geq 0 & (d)
\end{aligned} \tag{6.10}$$

where the $\epsilon_{k,N}$ are expected to go to 0 in the limit $N \rightarrow \infty$ as a consequence of Eq. (5.7) for the variances and of Chebyshev's inequality (B.5)³, the quantum mechanical expectation values of the $\epsilon_{j,N}$ being equal to $1/\sqrt{N}$. We remind that the term $m(\sigma^{a'})_n$ solely denotes the measured value of the observable $\sigma^{a'}$ in the n^{th} measurement and does not indicate a functional dependence.

A contradiction occurs, if counterfactual definiteness is assumed, defined by Stapp for the specific situation considered as follows: *Regardless of which of the four alternative possible measurements is performed, there is determined or fixed, within nature, according to some underlying theoretical conception of nature, a quartet of values (r_1, r_2, r_3, r_4) , in which r_k can be identified as the value that would be obtained as the result of the measurement if the measurement k were performed* [54, p. 438]. One may set $r_1 = m(\sigma^{a'})_n$, $r_2 = m(\sigma^{a''})_n$, $r_3 = m(\sigma^{b'})_n$, $r_4 = m(\sigma^{b''})_n$. The simultaneous validity of all four equations (6.10) with one set of values $m(\sigma^{a'})_n, m(\sigma^{a''})_n, m(\sigma^{b'})_n, m(\sigma^{b''})_n$ for each n in $\{1, \dots, N\}$ is thus presumed. Under these circumstances, the following inequality can be derived from equations (6.10), cf. Eq. (8) in Ref. [54],

$$\begin{aligned}
(2 - \sqrt{2})^2 &\leq \frac{1}{N} \sum_{n=1}^N \{m_n + m(\sigma^{a''})_n + \sqrt{2} m(\sigma^{b''})_n\}^2 \\
&= \frac{2}{N} \sum_{n=1}^N \{2 + m_n m(\sigma^{a''})_n + \sqrt{2} m_n m(\sigma^{b''})_n + \sqrt{2} m(\sigma^{a''})_n m(\sigma^{b''})_n\} \\
&= \frac{2}{N} \sum_{n=1}^N \{m_n m(\sigma^{a''})_n + \sqrt{2} \left(\frac{1}{\sqrt{2}} + m_n m(\sigma^{b''})_n\right) + \sqrt{2} \left(\frac{1}{\sqrt{2}} + m(\sigma^{a''})_n m(\sigma^{b''})_n\right)\} \\
&= 2(\mp \epsilon_{3,N} \pm \epsilon_{2,N} \pm \epsilon_{4,N}) \leq 2(\epsilon_{2,N} + \epsilon_{3,N} + \epsilon_{4,N})
\end{aligned} \tag{6.11}$$

where the relationships $m_n = m(\sigma^{a'})_n = -m(\sigma^{b'})_n$ and $|m(\sigma)| = 1$ have been used. With increasing N , it becomes more probable that the inequality $\epsilon_{2,N} + \epsilon_{3,N} + \epsilon_{4,N} \geq (2 - \sqrt{2})^2/2 \approx 0.1715$ is violated. For finite N , the result of this analysis is not a blunt contradiction but a weaker statement of probability.

The above reasoning is based on two constituents: on quantum mechanical predictions and on measurement values. The quantum mechanical predictions used are solely based on the state ψ_0 , into which the two-particle system is prepared, and on the form (5.4) of the correlation operator, a form which is imposed by the operational definition (5.5) of the correlation between two observables. The correctness of the quantum mechanical predictions (5.6) and (5.7) for the expectation value and for the variance of spin-spin correlations is well established. Nonlocality is no topic here.

For the purpose of a proper discussion of the role of measurement values, it is important to distinguish clearly between 'measurement values' in the sense of observable values measured in actual experiments and 'measurement values' in the sense of predictions, based on some theory, for such measured values. No

³Stapp sets all $\epsilon_{k,N}$ equal to 0 right away, but attests only an approximate validity to the resulting equations.

contradiction can be derived from Eq. (6.10) if the quantities $m(\sigma)$ are really measured values, although the equations all express valid correlations, because there is no factual situation in which all four equations are simultaneously applicable. In an actual measurement series, the experimenter has to decide whether he wants to measure the spin components $\sigma^{a'}$ or $\sigma^{a''}$, and $\sigma^{b'}$ or $\sigma^{b''}$, and only *one* of the four equations (6.10a) – (6.10d) is involved. In order to cover all four equations, four different measurement series are needed. But, for instance, the measurement values $m(\sigma^{b''})_n$ appearing in equations (6.10b) and (6.10d) can then not be identified, a necessary condition in Stapp's argument, because they result from different measurements and may have different values. Counterfactual definiteness is the problematic assumption here not locality.

On the other hand, the concept of counterfactual definiteness appears quite natural if the notion 'measurement value' is understood in the sense of theoretical predictions for measured observable values: it is hard to see why not *all* possible predictions should be available at once. However, this does not concern the statistical interpretation of quantum mechanics because only probability distributions are then predicted not observable values measured in individual experiments; remember that quantum mechanics assigns the spin state $\rho_0 = \frac{1}{2} I$ to the isolated particles a and b . Consequently, Stapp outlines a local stochastic theory of classical type yielding simultaneous predictions $r_n(\mathbf{e}^{a'})$, $r_n(\mathbf{e}^{a''})$, $r_n(\mathbf{e}^{b'})$, $r_n(\mathbf{e}^{b''})$ for the measured values $m_n(\sigma^{a'})$, $m_n(\sigma^{a''})$, $m_n(\sigma^{b'})$, $m_n(\sigma^{b''})$ in the n^{th} experimental run. Inserting the predictions $r_n(\cdot)$ into the correlation formulae (6.10) exactly represents the situation of counterfactual definiteness as defined above although Stapp pretends to do without. Assuming counterfactual definiteness implicitly also assumes an underlying classical theory.

Stapp concludes from the resulting contradiction: *If one accepts these ideas, then the incompatibility with quantum predictions must apparently be attributed to a failure of the locality idea that no causal influence of any kind can act outside the forward light cone* [54, p. 446]. We argue against it that Stapp's stochastic theory providing simultaneous predictions for all observable values in an individual experimental run must be a classical theory in the sense of Sec. 2.5 not a quantum theory, and, hence, that no conclusions concerning quantum theories can rightly be drawn. The failure is, as is the failure of some conclusions drawn from Bell's theorem, to argue against quantum mechanics with a theory of classical type. The incompatibility of such theories with quantum mechanics is not surprising in regard of the results obtained in Chap. 3.

Chapter 7

Galilei invariance and the quantum mechanics of nonrelativistic particles

Galilei and Lorentz invariance of physical space-time are formal specifications of universality. It is shown in this chapter that the physics of a free, nonrelativistic particle is completely determined by the quantum mechanical formalism and by the requirement of Galilei invariance. A necessary pragmatic prerequisite is the possibility to parametrize physical space and time *i.e.* the existence of reliable rules and watches allowing the definition of reference frames. Physical space is homogeneous and isotropic if measured with reliable rules and physical time is homogeneous if measured with reliable watches. These symmetries *define* the notion 'reliable' in the present context, and it is an empirical fact that reliable rules and watches exist.

A special class of reference frames are *inertial frames* in relation to which a free particle is not accelerated. The origins of two inertial frames move with constant relative velocity and the angle between an arbitrary fixed direction in frame 1 and an arbitrary fixed direction in frame 2 is constant in time. With Galilei invariance, it is, furthermore, assumed that measured values of space and time intervals do not depend on which frame the rules and watches are at rest (frame invariance of space and time intervals). Special relativity reveals that these assumptions are approximately valid only if the relative velocities between inertial frames are small compared to the velocity of light.

It is shown in Chap. 3 that the noncommutativity of the observable algebra is the decisive feature discriminating quantum and classical theories. In this chapter, the origin of this feature is revealed for a simple physical system, the particle: the observables of a nonrelativistic particle are intimately related to the generators of 1-parameter subgroups of the Galilei group and inherit their noncommutative structure. The following considerations are based on the work of T. F. Jordan [55] and its elaboration by L. E. Ballentine [25, Chap. 3]. Although the idea is not new, the presentation given here is close to physics and comparatively elementary. Only some basic facts about Lie groups and their representation theory are required. The same topic is discussed *e.g.* by J.-M. Levy-Leblond [56], but with a much more mathematical approach.

7.1 The Galilei group

A *Galilei transform* is a map of physical space-time onto itself such that spacial and temporal distances are preserved and inertial frames are mapped to inertial frames. An inertial frame is defined by an origin O of space-time and a right trihedral of space directions $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. A space-time point P is parametrized by its space distance $OP = \sum_{j=1}^3 x_j \mathbf{e}_j \equiv \mathbf{x}_P$ and its time distance t from the origin. Consider now a second inertial frame with origin O' and directional trihedral $\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3$. The space distance OO' equals $\mathbf{v}t + \mathbf{a}$, the corresponding time distance equals s . The quantities \mathbf{v} , \mathbf{a} , and s are constants. The directions \mathbf{e}'_j are given

by $\mathbf{e}'_j = R \mathbf{e}_j$, $j \in \{1, 2, 3\}$, where R is a constant orthogonal matrix with determinant 1. The vector OP is transformed to the vector $O'P' = \sum_{j=1}^3 x_j \mathbf{e}'_j \equiv \mathbf{x}_{P'}$. The vector $\mathbf{x}_{P'}$ has the same length as \mathbf{x}_P and the same orientation with respect to the directions \mathbf{e}'_j as \mathbf{x}_P has with respect to the directions \mathbf{e}_j . A general Galilei transform is thus given by

$$\mathbf{x}_{P'} = R \mathbf{x}_P + \mathbf{v} t_P + \mathbf{a} \quad (7.1)$$

$$t_{P'} = t_P + s \quad (7.2)$$

Active transforms $P \rightarrow P'$ are considered here: $\mathbf{x}_{P'}, t_{P'}$ are the space-time coordinates of the transformed point P' with respect to a fixed coordinate system.

A Galilei transform can be represented as a matrix operation on a 4-dimensional subspace of \mathbb{R}^5 by setting, with respect to a fixed coordinate system,

$$\begin{pmatrix} R\mathbf{x} + \mathbf{v}t + \mathbf{a} \\ t + s \\ 1 \end{pmatrix} = \begin{pmatrix} R, \mathbf{v}, \mathbf{a} \\ \mathbf{0}^T, 1, s \\ \mathbf{0}^T, 0, 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ t \\ 1 \end{pmatrix} \quad (7.3)$$

where $\mathbf{0}^T$ denotes the transposed of the column vector $\mathbf{0}$. Galilei transforms form a group, the *Galilei group* with matrix multiplication as the group product. The unit matrix is the neutral element and the inverse element is given by

$$\begin{pmatrix} R, \mathbf{v}, \mathbf{a} \\ \mathbf{0}^T, 1, s \\ \mathbf{0}^T, 0, 1 \end{pmatrix}^{-1} = \begin{pmatrix} R^{-1}, -R^{-1}\mathbf{v}, R^{-1}(\mathbf{v}s - \mathbf{a}) \\ \mathbf{0}^T, 1, -s \\ \mathbf{0}^T, 0, 1 \end{pmatrix} \quad (7.4)$$

The matrix product and the the formation of the invers are smooth operations. The Galilei group is thus a *matrix Lie group* \mathbb{G} characterized by ten parameters (\mathbf{v} , \mathbf{a} , s , and the three Euler angles determining R [57, Sec. 4.4]). The theory of matrix Lie groups is developed with due mathematical rigor by B. C. Hall in Ref. [58]. We content ourselves here to compile the essential structural elements.

7.1.1 The Lie algebra of a matrix Lie group

A central notion used for the analysis of matrix Lie groups is the *matrix exponential* e^X of a real or complex $n \times n$ matrix X given by the familiar power series

$$e^X = \sum_{m=0}^{\infty} \frac{X^m}{m!} \quad (7.5)$$

The sum converges for any such matrix with respect to a given matrix norm $\|A\|$ with $\|cA\| = |c| \|A\|$, $\|A + B\| \leq \|A\| + \|B\|$, $\|AB\| \leq \|A\| \|B\|$. If A is an invertible $n \times n$ matrix then

$$e^{AXA^{-1}} = A e^X A^{-1} \quad (7.6)$$

Further important features of the matrix exponential are proved in Ref. [58, Sec. 2.1, 2.2].

The *Lie algebra* \mathfrak{g} of a matrix Lie group \mathbb{G} is the set of all matrices K such that e^{sK} is in \mathbb{G} for all real numbers s [58, p. 39]. It is the set of all generators of smooth 1-parameter subgroups in \mathbb{G} . The main properties of the Lie algebra of a matrix Lie group are [58, Sec. 2.6]:

- Let K be an Element of \mathfrak{g} , A an invertible element of \mathbb{G} . Then, AKA^{-1} is in \mathfrak{g}
- Let K and L be elements of \mathfrak{g} . Then

1. rK is in \mathfrak{g} for all real numbers r
 2. $K + L$ is in \mathfrak{g}
 3. The *bracket* or *commutator* $[K, L] \equiv KL - LK$ is in \mathfrak{g}
- $K = \left. \frac{d}{ds} e^{sK} \right|_{s=0}$

Properties 1 and 2 reveal that \mathfrak{g} is a real vector space.

The map $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ has the following properties [58, p. 53]:

1. $[\cdot, \cdot]$ is bilinear
2. $[L, K] = -[K, L]$ for all $L, K \in \mathfrak{g}$ (skew symmetry) (7.7)
3. $[K, [L, M]] + [L, [M, K]] + [M, [K, L]] = 0$ for all $L, K, M \in \mathfrak{g}$ (Jacobi identity)

Let n be the dimension of a finite-dimensional real Lie algebra \mathfrak{g} and let K_1, K_2, \dots, K_n be a basis for \mathfrak{g} considered as a vector space. The commutator $[K_\mu, K_\nu]$ can then be written uniquely in the form

$$[K_\mu, K_\nu] = \sum_{\lambda=1}^n c_{\mu\nu}^\lambda K_\lambda \quad (7.8)$$

for each index pair μ and ν . The real constants $c_{\mu\nu}^\lambda$ are called the *structure constants* of \mathfrak{g} with respect to the chosen basis. The skew symmetry of the bracket and the Jacobi identity imply that [58, Sec. 2.8.1]

$$c_{\mu\nu}^\lambda + c_{\nu\mu}^\lambda = 0 \quad (7.9)$$

and

$$\sum_{\sigma=1}^n c_{\mu\nu}^\sigma c_{\sigma\lambda}^\rho + c_{\nu\lambda}^\sigma c_{\sigma\mu}^\rho + c_{\lambda\mu}^\sigma c_{\sigma\nu}^\rho = 0 \quad (7.10)$$

for all values of the indices μ, ν, λ , and ρ .

7.1.2 A basis for the Lie algebra of the Galilei group

A general space rotation can be generated by consecutive rotations around the 3 coordinate axes $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. A basis for the Lie algebra \mathfrak{g} of the Galilei group can thus be constructed with the following ten 1-parameter subgroups of \mathbb{G} :

Space-time transformations	1-parameter subgroups	(7.11)
Translations along the coordinate axes, $j \in \{1, 2, 3\}$: $x_j \rightarrow x_j + a_j, R = I, \mathbf{v} = \mathbf{0}, s = 0$	$\exp(a_j \Pi_j)$	
Rotations around the coordinate axes, $j \in \{1, 2, 3\}$: $\mathbf{x} \rightarrow R_j(\theta_j) \mathbf{x}, \mathbf{v} = \mathbf{0}, \mathbf{a} = \mathbf{0}, s = 0$	$\exp(\theta_j \Lambda_j)$	
Transformation to a moving system, $j \in \{1, 2, 3\}$: $x_j \rightarrow x_j + v_j t, R = I, \mathbf{a} = \mathbf{0}, s = 0$	$\exp(v_j \Upsilon_j)$	
Time translation: $t \rightarrow t + s, R = I, \mathbf{v} = \mathbf{0}, \mathbf{a} = \mathbf{0}$	$\exp(sH)$	

The symbol H denotes here the Greek letter Eta.

With the defining equation $K = \frac{d}{ds} e^{sK} \big|_{s=0}$ of the generator of a 1-parameter subgroup and the matrices for rotations around the coordinate axes given by

$$\begin{aligned} R_1(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \\ R_2(\theta) &= \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \\ R_3(\theta) &= \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned} \quad (7.12)$$

the following matrices are obtained for the generators of the 1-parameter subgroups

$$\begin{aligned} \Pi_1 &= \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Pi_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Pi_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ \Lambda_1 &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Lambda_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Lambda_3 = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ \Upsilon_1 &= \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Upsilon_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Upsilon_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ H &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned} \quad (7.13)$$

To calculate the 45 commutators for this basis of \mathfrak{g} is an elementary exercise in matrix operations. The commutators are, with $j, k, l \in \{1, 2, 3\}$,

$$\begin{aligned} (a) \quad [\Pi_j, \Pi_k] &= 0 & (f) \quad [\Upsilon_j, \Pi_k] &= 0 \\ (b) \quad [\Lambda_j, \Lambda_k] &= \sum_{l=1}^3 \epsilon_{jkl} \Lambda_l & (g) \quad [\Pi_j, H] &= 0 \\ (c) \quad [\Upsilon_j, \Upsilon_k] &= 0 & (h) \quad [\Lambda_j, H] &= 0 \\ (d) \quad [\Lambda_j, \Pi_k] &= \sum_{l=1}^3 \epsilon_{jkl} \Pi_l & (i) \quad [\Upsilon_j, H] &= \Pi_j \\ (e) \quad [\Lambda_j, \Upsilon_k] &= \sum_{l=1}^3 \epsilon_{jkl} \Upsilon_l \end{aligned} \quad (7.14)$$

ϵ_{jkl} is the totally antisymmetric tensor, whose only elements different from 0 are $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = -\epsilon_{132} = -\epsilon_{321} = -\epsilon_{213} = 1$.

7.2 Unitary representations

Group representations are the suited tool to implement the consequences of symmetry groups, in the present context of the Gallilei group, into the theoretical description of physical systems. In the case of quantum mechanics in the Schrödinger picture, the action of an element G of the symmetry group \mathbb{G} is represented by a map $\psi \rightarrow \psi_G$ on Hilbert space vectors ψ . Such a map should preserve the parts of the Hilbert space geometry with physical significance. Since ψ and $e^{ir}\psi$, $r \in \mathbb{R}$, describe the same physical situation, a minimal requirement on the map is that it preserves the modulus of the scalar product of a vectors ϕ and ψ

$$|(\phi_G, \psi_G)| = |(\phi, \psi)| \quad (7.15)$$

Wigner's theorem states that such a map must be either a *unitary* transform with $(\phi_G, \psi_G) = (\phi, \psi)$ and $(c_1\phi + c_2\psi)_G = c_1\phi_G + c_2\psi_G$ or an *antiunitary* transform with $(\phi_G, \psi_G) = (\phi, \psi)^*$ and $(c_1\phi + c_2\psi)_G = c_1^*\phi_G + c_2^*\psi_G$, the asterisk denoting complex conjugation. For a proof of Wigner's theorem see [59]. Only unitary transforms can serve for a representation of the smooth Galilei group because the square of an antiunitary transform is unitary.

An *unitary representation of a Lie group* \mathbb{G} is a smooth map U of \mathbb{G} onto a smooth subgroup $\mathbb{U}_{\mathbb{G}}$ of all unitary transforms \mathbb{U} on some Hilbert space \mathcal{H} that preserves the group structure of \mathbb{G}

$$U : G \in \mathbb{G} \rightarrow U(G) \in \mathbb{U}_{\mathbb{G}} \quad (7.16)$$

$$U(G_2 G_1) = U(G_2) U(G_1) \quad (7.17)$$

U is a *Lie group homomorphism* [58, p. 17, 22]. For such a Lie group homomorphism, a linear map $u : \mathfrak{g} \rightarrow \mathfrak{u}_{\mathbb{G}}$ from the Lie algebra \mathfrak{g} of \mathbb{G} to the Lie algebra $\mathfrak{u}_{\mathbb{G}}$ of $\mathbb{U}_{\mathbb{G}}$ exists such that for all elements $K \in \mathfrak{g}$

$$U(e^K) = e^{iu(K)} \quad (7.18)$$

The map u has the following properties:

1. $u(GKG^{-1}) = U(G) u(K) U(G)^{-1}$ for all elements $G \in \mathbb{G}$ and $K \in \mathfrak{g}$
2. $u([K, L]) = i[u(K), u(L)]$ for all $K, L \in \mathfrak{g}$
3. $i u(K) = \frac{d}{ds} U(e^{sK})|_{s=0}$

These properties of a *Lie algebra homomorphism* are demonstrated in Ref. [58, p. 45]. Point 1 is a consequence of Eq. (7.6). Point 3 indicates how to calculate elements $u(K)$ in \mathfrak{u} . Point 2 follows from point 3 and from the identity

$$[K, L] = \frac{d}{ds} (e^{sK} L e^{-sK})|_{s=0} \quad (7.20)$$

As a consequence, the map u preserves Eq. (7.8) and the the values of the structural constants $c_{\mu\nu}^\lambda$.

Stone's theorem states that any continuous, 1-parameter, unitary group $U(s)$, with s in \mathbb{R} , and $U(s_1 + s_2) = U(s_1)U(s_2)$, has the form $U(s) = e^{isA}$, where A is a self-adjoint operator [60, Sec. 137]. The Lie algebra \mathfrak{u} consists, therefore, of self-adjoint operators.

As for unitary representations of the Galilei group, we set, with $j \in \{1, 2, 3\}$,

$$\begin{aligned} u(\Pi_j) &= -P_j \equiv K_j \\ u(\Lambda_j) &= -L_j \equiv K_{j+3} \\ u(\Upsilon_j) &= Y_j \equiv K_{j+6} \\ u(H) &= H \equiv K_{10} \end{aligned} \quad (7.21)$$

the quantities K_l denoting elements of the Lie algebra \mathfrak{u} from now on. The minus signs in the definitions of \mathbf{P} and \mathbf{L} conforms with the dynamical significance of these quantities established in Sec. 7.5. With the above definitions and from Eq. (7.14), the structural equations for the basic commutators of the unitary representation of the Galilei group are

$$\begin{aligned}
(a) \quad \frac{1}{i} [P_j, P_k] &= 0 & (f) \quad \frac{1}{i} [Y_j, P_k] &= 0 \\
(b) \quad \frac{1}{i} [L_j, L_k] &= \sum_{l=1}^3 \epsilon_{jkl} L_l & (g) \quad \frac{1}{i} [P_j, H] &= 0 \\
(c) \quad \frac{1}{i} [Y_j, Y_k] &= 0 & (h) \quad \frac{1}{i} [L_j, H] &= 0 \\
(d) \quad \frac{1}{i} [L_j, P_k] &= \sum_{l=1}^3 \epsilon_{jkl} P_l & (i) \quad \frac{1}{i} [Y_j, H] &= P_j \\
(e) \quad \frac{1}{i} [L_j, Y_k] &= \sum_{l=1}^3 \epsilon_{jkl} Y_l
\end{aligned} \tag{7.22}$$

The only non-vanishing structure constants of this Lie algebra basis are, $j, k, l \in \{1, 2, 3\}$,

$$\begin{aligned}
(7.22d) \quad &\implies c_{j+3, k}^l = c_{j, k+3}^l = \epsilon_{jkl} \\
(7.22b) \quad &\implies c_{j+3, k+3}^{l+3} = \epsilon_{jkl} \\
(7.22e) \quad &\implies c_{j+3, k+6}^{l+6} = c_{j+6, k+3}^{l+6} = \epsilon_{jkl} \\
(7.22i) \quad &\implies c_{j+6, 10}^j = -c_{10, j+6}^j = 1
\end{aligned} \tag{7.23}$$

7.3 Unitary ray representations

At this point, it is important to remind that the Hilbert space vectors ψ and $e^{ir}\psi$, $r \in \mathbb{R}$, describe the same physical situation. The *projection* P_ψ projecting onto ψ and defined by $P_\psi \phi = ((\psi, \phi)/(\psi, \psi))\psi$, not ψ itself, is in a one to one correspondence with a pure quantum mechanical state s , the expectation value of an observable A being given by the trace formula $\langle A \rangle_s = \text{Tr} (P_\psi A)$.

The unitary map $U : \psi \rightarrow U\psi$ induces a map $P_\psi \rightarrow P_{U\psi}$ between the corresponding projections. It is

$$P_{U\psi} = U P_\psi U^{-1} \tag{7.24}$$

as readily follows from the definition of P_ψ . As a consequence, the unitary group representations U , Eq. (7.16), and $U' : G \in \mathbb{G} \rightarrow U'(G) \in \mathbb{U}_{\mathbb{G}}$ are physically equivalent if $U'(G) = e^{ir(G)}U(G)$ with $r(G) \in \mathbb{R}$ and $r(G_2 G_1) = r(G_2) + r(G_1)$ for all elements G in \mathbb{G} . The last condition ensures that U' is a Lie group homomorphism if U is one. This relationship between $U(G)$ and $U'(G)$ establishes an equivalence relation $U' \sim U$ between unitary representations of \mathbb{G} and a consequent subdivision of the set $\{\mathbb{U}_{\mathbb{G}}\}$ of all unitary representations into equivalence classes $\tilde{\mathbb{U}}_{\mathbb{G}}$. The true object of the present inquiry are, therefore, *unitary ray representations* [61]

$$\tilde{U} : G \in \mathbb{G} \rightarrow \tilde{U}(G) \subset \tilde{\mathbb{U}}_{\mathbb{G}} \tag{7.25}$$

Consider now a smooth 1-parameter subgroup $G(s)$ and its unitary ray representation $\tilde{U}(G(s)) \equiv \tilde{U}(s)$. Two representatives $U(s) = e^{iKs}$ and $U'(s) = e^{iK's}$ are members of the same class if $U'(s) = e^{irs}U(s)$ for all values of s and for some real number r because of the conditions $U(s_1 + s_2) = U(s_1)U(s_2)$ and $U'(s_1 + s_2) = U'(s_1)U'(s_2)$, respectively. The unitary representations $U(s)$ and $U'(s)$ are equivalent if and only if their generators K and K' differ by a real multiple of identity

$$K' = K + rI \tag{7.26}$$

This signifies that a unitary ray representation of a 1-parameter subgroup determines its generator K only up to a real constant.

Let now K_μ and $K'_\mu = K + r_\mu I$ denote two equivalent sets of representatives of a basis of $\tilde{\mathfrak{u}}_{\mathbb{G}}$. Clearly, it is

$$[K'_\mu, K'_\nu] = [K_\mu, K_\nu] \quad (7.27)$$

and the Jacobi identity is equally valid for the sets $\{K_\mu\}$ as for $\{K'_\mu\}$. However, Eq. (7.8) must be generalized to

$$[K_\mu, K_\nu] = \sum_{\lambda=1}^n c_{\mu\nu}^\lambda K_\lambda + r_{\mu\nu} I \quad (7.28)$$

with $r_{\nu\mu} = -r_{\mu\nu}$, and accordingly for the K'_μ . Comparison of the two resulting equations, together with Eqns. (7.26) and (7.27), yields

$$\Delta r_{\mu\nu} \equiv r'_{\mu\nu} - r_{\mu\nu} = - \sum_{\lambda=1}^n c_{\mu\nu}^\lambda r_\lambda \quad (7.29)$$

The coefficients $r_{\mu\nu}$ are restricted by the Jacobi identities, which yields

$$\sum_{\lambda=1}^n c_{\mu\nu}^\lambda r_{\lambda\sigma} + c_{\nu\sigma}^\lambda r_{\lambda\mu} + c_{\sigma\mu}^\lambda r_{\lambda\nu} = 0 \quad (7.30)$$

and equally for the $r'_{\mu\nu}$. These equations characterize the values $r_{\mu\nu}$ of a consistent set of representatives. It is a consequence of Eq. (7.10) that the $\Delta r_{\mu\nu}$ of Eq. (7.29) also fulfil Eq. (7.30).

The equations (7.30) are invariant under permutations of the free parameters μ, ν, σ . For the Gallilei group, $n = 10$, this implies that one has $10 \times 9 \times 8/3! = 120$ different cases to consider. Based on Eqns. (7.23), this elementary but somehow tedious exercise yields the values of $r_{\mu\nu}$ listed in the table below:

$r_{\mu\nu}$										
ν	1	2	3	4	5	6	7	8	9	10
μ										
1	0	0	0	0	x_3	$-x_2$	$-\xi$	0	0	0
2	0	0	0	$-x_3$	0	x_1	0	$-\xi$	0	0
3	0	0	0	x_2	$-x_1$	0	0	0	$-\xi$	0
4	0	x_3	$-x_2$	0	x_6	$-x_5$	0	x_9	$-x_8$	0
5	$-x_3$	0	x_1	$-x_6$	0	x_4	$-x_9$	0	x_7	0
6	x_2	$-x_1$	0	x_5	$-x_4$	0	x_8	$-x_7$	0	0
7	ξ	0	0	0	x_9	$-x_8$	0	0	0	x_1
8	0	ξ	0	$-x_9$	0	x_7	0	0	0	x_2
9	0	0	ξ	x_8	$-x_7$	0	0	0	0	x_3
10	0	0	0	0	0	0	$-x_1$	$-x_2$	$-x_3$	0

revealing that only 42 of the 100 coefficients $r_{\mu\nu}$ can be different from 0. There are linear interdependencies between these 42 coefficients such that only 10 linear independent quantities, $\xi, x_1, x_2, \dots, x_9$, remain.

The evaluation of Eq. (7.29) yields a second table with 9 free parameters r_1, r_2, \dots, r_9 :

$\Delta r_{\mu\nu}$										
ν	1	2	3	4	5	6	7	8	9	10
μ										
1	0	0	0	0	$-r_3$	r_2	0	0	0	0
2	0	0	0	r_3	0	$-r_1$	0	0	0	0
3	0	0	0	$-r_2$	r_1	0	0	0	0	0
4	0	$-r_3$	r_2	0	$-r_6$	r_5	0	$-r_9$	r_8	0
5	r_3	0	$-r_1$	r_6	0	$-r_4$	r_9	0	$-r_7$	0
6	$-r_2$	r_1	0	$-r_5$	r_4	0	$-r_8$	r_7	0	0
7	0	0	0	0	$-r_9$	r_8	0	0	0	$-r_1$
8	0	0	0	r_9	0	$-r_7$	0	0	0	$-r_2$
9	0	0	0	$-r_8$	r_7	0	0	0	0	$-r_3$
10	0	0	0	0	0	0	r_1	r_2	r_3	0

(7.32)

Assume now that a coherent set of representatives $\{K_{\mu\nu}\}$ produces the table (7.31). By passing to the coherent set $\{K'_{\mu\nu}\}$ with $r_\lambda = x_\lambda$, $\lambda = 1, 2, \dots, 9$ all entries of Tab. (7.31) can be set to zero except the entries $\pm\xi$ which remain unchanged. $\{K'_{\mu\nu}\}$ then fulfils the commutation relations of Eq. (7.22) except that (7.22f) must be replaced by

$$\frac{1}{i} [Y_j, P_k] = \xi \delta_{jk}, \quad \xi \in \mathbb{R} \quad (7.33)$$

This generates a 1-parameter set of nonequivalent unitary ray representations of the Galilei group, parametrized by ξ .

7.4 Unitary ray representations of the Galilei group in $\mathbb{L}^2(\mathbb{R}^3)$

In the *Schrödinger picture*, physical dynamics is represented, in the most simple case, by the wave function $\psi(\mathbf{x}, t)$, a complex function being normalized to $\int d^3x |\psi(\mathbf{x}, t)| = 1$ for all times t i.e. ψ is a member of the Hilbert space $\mathbb{L}^2(\mathbb{R}^3)$. The generalization to wave functions $\psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)$ is obvious.

Let $x \equiv (\mathbf{x}, t)$. A unitary representation of the Galilei group in $\mathbb{L}^2(\mathbb{R}^3)$ is then given by

$$(U(G)\psi)(x) \equiv \psi_G(x) = \psi(G^{-1}x) \equiv \psi_I(G^{-1}x) \quad (7.34)$$

for all members G of \mathbb{G} , the inverse map $G^{-1}x$ being required in the argument of ψ_I because of the representation condition $U(G_2 G_1) = U(G_2) U(G_1)$. The active point of view is taken here: $\psi_G(x)$ denotes the wave function of a transformed object with respect to a fixed coordinate system.

The transition to the *Heisenberg picture* is performed, as usual, via the identities

$$\langle A \rangle_{\psi_G} = (\psi_G, A \psi_G) = (U(G) \psi_I, A U(G) \psi_I) = (\psi_I, U(G)^{-1} A U(G) \psi_I) \equiv \langle A_G \rangle_{\psi_I} \quad (7.35)$$

for an arbitrary wave function $\psi_I(x)$ yielding

$$A_G = U(G)^{-1} A_I U(G) \quad (7.36)$$

The operator A_G does not depend on the choice of the representative $U(G)$ of the unitary ray representation.

7.4.1 The generators \mathbf{P} and \mathbf{L}

The generators \mathbf{P} and \mathbf{L} are now readily evaluated. From Tab. (7.11) and the relationship (7.19.3) with $u(\Pi_1) = -P_1$ and $s = a_1$, one obtains

$$\mathbf{i} \left(-P_1 \psi \right) (\mathbf{x}, t) = d_{a_1} \psi(x_1 - a_1, x_2, x_3, t)|_{a_1=0} = -(\partial_{x_1} \psi) (\mathbf{x}, t) \quad (7.37)$$

or

$$P_j = \frac{1}{\mathbf{i}} \partial_{x_j} \quad (7.38)$$

Analogously, with Eq. (7.12), $R_3^{-1}(\theta) = R_3(-\theta)$, $u(\Lambda_3) = -L_3$, $s = \theta$:

$$\begin{aligned} \mathbf{i} \left(-L_3 \psi \right) (\mathbf{x}, t) &= d_{\theta_3} \psi(x_1 \cos \theta_3 + x_2 \sin \theta_3, -x_1 \sin \theta_3 + x_2 \cos \theta_3, x_3, t)|_{\theta_3=0} \\ &= (x_2 \partial_{x_1} \psi - x_1 \partial_{x_2} \psi) (\mathbf{x}, t) \end{aligned} \quad (7.39)$$

i.e. $L_3 = X_1 P_2 - X_2 P_1$ or, in general,

$$L_l = \sum_{j,k=1}^3 \epsilon_{jkl} X_j P_k \quad (7.40)$$

where the position operator \mathbf{X} is the multiplication operator $(\mathbf{X}\psi)(\mathbf{x}, t) = \mathbf{x}\psi(\mathbf{x}, t)$.

The basic commutators

$$\frac{1}{\mathbf{i}} [X_j, X_k] = 0 \quad (7.41)$$

$$\frac{1}{\mathbf{i}} [P_j, P_k] = 0 \quad (7.42)$$

$$\frac{1}{\mathbf{i}} [X_j, P_k] = \delta_{jk} \quad (7.43)$$

imply with Eq. (7.40) that

$$\frac{1}{\mathbf{i}} [L_j, P_k] = \sum_{l=1}^3 \epsilon_{jkl} P_l \quad (7.44)$$

$$\frac{1}{\mathbf{i}} [L_j, X_k] = \sum_{l=1}^3 \epsilon_{jkl} X_l \quad (7.45)$$

$$\frac{1}{\mathbf{i}} [L_j, L_k] = \sum_{l=1}^3 \epsilon_{jkl} L_l \quad (7.46)$$

The commutation relations (7.22a), (7.22b), and (7.22d) are fulfilled with this choice for \mathbf{P} and \mathbf{L} .

7.4.2 Time translations and time derivatives

Consider now time translations. Let us denote by $\psi_s(\mathbf{x}, t)$ the wave function of a system whose history starts at time s instead at time 0, the initial conditions of the system described by ψ_0 being delayed to time s . This transition is described by

$$\psi_s(\mathbf{x}, t) \equiv \psi_0(\mathbf{x}, t - s) = (e^{iHs}\psi_0)(\mathbf{x}, t) \quad (7.47)$$

Taking the derivative with respect to s yields

$$(iH\psi_0)(\mathbf{x}, t) = (d_s\psi_s|_{s=0})(\mathbf{x}, t) = d_s\psi_0(\mathbf{x}, t - s)|_{s=0} = -(\partial_t\psi_0)(\mathbf{x}, t) \quad (7.48)$$

Eq. (7.48) does not define the generator H of time translations in contrast to the similar situation of Eq. (7.37) because $i\partial_t$ is no operator on $\mathbb{L}^2(\mathbb{R}^3)$. The task to evaluate the generator of time translations as an operator on $\mathbb{L}^2(\mathbb{R}^3)$ is tackled in Sec. 7.4.4. But Eq. (7.48) is the Schrödinger equation $d_t\psi_0 = -iH\psi_0$ which governs the time evolution of the wave function. Its integrated form

$$\psi_0(t) = e^{-iHt}\psi_0(0) \quad (7.49)$$

implies

$$\psi_s(t + s) = \psi_0(t) = e^{-iHt}\psi_0(0) = e^{-iHt}\psi_s(s) \quad (7.50)$$

revealing the homogeneity of physical time: the time evolution is independent of the chosen initial time s of the process if H is independent of time.

In the Heisenberg picture, Eq. (7.50) corresponds to

$$A_s(t + s) = e^{iHt}A_s(s)e^{-iHt} \quad (7.51)$$

from which the Heisenberg equation

$$d_t A_s(t + s)|_{t=0} = \frac{1}{i} [A_s(s), H] \quad (7.52)$$

results. The effect of a time translation s is, in the Heisenberg picture, given by Eq. (7.36) yielding

$$A_s(t) = e^{-iHs}A_0(t)e^{iHs} \quad (7.53)$$

Together with Eq. (7.51) at $s = 0$, one deduces $A_s(t) = A_0(t - s)$. The initial value $A_0(0)$ is mapped to $A_s(s)$, the initial value in the time translated system, and the subsequent time evolution of the time translated system is identical with the one of the original system. This again expresses the homogeneity of physical time.

7.4.3 There is no unitary representation of the Galilei group in $\mathbb{L}^2(\mathbb{R}^3)$

Before the generators \mathbf{Y} and H can be evaluated, it has to be stated that the Galilei group has no unitary representation of the form (7.34) in $\mathbb{L}^2(\mathbb{R}^3)$ because the commutation relations (7.22) cannot be met with self-adjoint operators on $\mathbb{L}^2(\mathbb{R}^3)$. For it follows from statement 2. of Appendix E and from Eqns. (7.22f) and (7.22g) that Y_j and H both are functions of \mathbf{P} alone. The commutator (7.22i) must, therefore, be equal to 0, and cannot be equal to P_j . Unitary ray representations are, however, possible because Eq. (7.22f) can then be replaced by Eq. (7.33) with $\xi \neq 0$.

7.4.4 The generators \mathbf{Y} and H for unitary ray representations

The task consists now to calculate the remaining generators \mathbf{Y} and H as self-adjoint operators on $\mathbb{L}^2(\mathbb{R}^3)$ by means of the relationships (7.22) but (7.22f) replaced by (7.33). We look for solutions which are polynomials in the components of \mathbf{X} and \mathbf{P} or can be approximated by such polynomials. The generators $\mathbf{L} = \mathbf{X} \wedge \mathbf{P}$ are examples. The components X_j and P_k are considered as building blocks for the generators \mathbf{Y} and H similarly to the situation in classical Hamilton mechanics where all observables are real functions of the coordinates q_j, p_k of phase space. The task is much facilitated with the following formal trick.

Commutators as formal derivatives

The basic commutators (7.41), (7.42), and (7.43) imply, together with the relationships,

$$[A, BC] = B[A, C] + [A, B]C \quad \text{and} \quad [AB, C] = A[B, C] + [A, C]B \quad (7.54)$$

that, for any polynomial $T(\mathbf{X}, \mathbf{P})$, the brackets

$$\frac{1}{i} [X_j, T] \equiv \partial_{P_j} T \quad (7.55)$$

and

$$\frac{1}{i} [T, P_k] \equiv \partial_{X_k} T \quad (7.56)$$

act as formal derivatives. The basic commutators give

$$\partial_{P_j} X_k = 0, \quad \partial_{X_k} P_j = 0, \quad \partial_{P_j} P_k = \delta_{jk}, \quad \partial_{X_k} X_j = \delta_{jk} \quad (7.57)$$

Eq. (7.54) yields multiplication rules in the noncommutative polynomial algebra:

$$\partial_{P_j}(T_1 T_2) = T_1 \partial_{P_j} T_2 + (\partial_{P_j} T_1) T_2 \quad (7.58)$$

$$\partial_{X_k}(T_1 T_2) = T_1 \partial_{X_k} T_2 + (\partial_{X_k} T_1) T_2 \quad (7.59)$$

As a consequence, the equations

$$\frac{1}{i} [X_j, T_1] = \partial_{P_j} T_1 = 0 \quad (7.60)$$

$$\frac{1}{i} [T_2, P_k] = \partial_{X_k} T_2 = 0 \quad (7.61)$$

imply that T_1 is independent of P_j and T_2 is independent of X_k , cf. Appendix E. It is, furthermore, $\partial_{P_j} I = 0$ and $\partial_{X_k} I = 0$ implying that

$$\partial_{P_j} P_j^n = n P_j^{n-1}, \quad \partial_{X_k} X_k^n = n X_k^{n-1} \quad (7.62)$$

for all entire values n . The Jacobi identity (7.7.3) entails that multiple partial derivatives can be interchanged, $\partial_{X_j} \partial_{P_k} T = \partial_{P_k} \partial_{X_j} T$ and so on.

In the following, we will mainly deal with polynomials of \mathbf{P} with real coefficients,

$$R(P_1, P_2, P_3) = \sum_{n_1, n_2, n_3} r_{n_1 n_2 n_3} P_1^{n_1} P_2^{n_2} P_3^{n_3}, \quad r_{n_1 n_2 n_3} \in \mathbb{R} \quad (7.63)$$

They are uniquely determined by the sequence $\{r_{n_1 n_2 n_3} / n_j = 0, 1, 2, \dots\}$ of their coefficients. The operation $\frac{1}{i} [X_j, R]$ induces some linear map on the coefficient sequence, the same map as does the partial derivative ∂_{p_j} on the real polynomial $R(p_1, p_2, p_3)$, $p_j \in \mathbb{R}$. This ensures, within the space of polynomials $R(P_1, P_2, P_3)$, that the formal derivative ∂_{P_j} acts in the same manner as does the ordinary partial derivative ∂_{p_j} on the space of the real polynomials. In particular, the chain rule is also valid for the formal derivatives.

The generator \mathbf{Y}

Tentative application of the relationship (7.19.3) yields

$$i(Y_1 \psi)(\mathbf{x}, t) = \partial_{v_1} \psi(\mathbf{x} - \mathbf{v}t, t) \big|_{\mathbf{v}=\mathbf{0}} = -t(\partial_{x_1} \psi)(\mathbf{x}, t) \quad (7.64)$$

or $\mathbf{Y} = -t\mathbf{P}$. But this result conforms only with unitary representations because it is then $\frac{1}{i} [Y_j, P_k] = 0$.

The evaluation of the generator Y_j for unitary ray representations starts with Eq. (7.33). Setting

$$Y_j = \xi X_j + Z_j \quad (7.65)$$

one obtains $[Z_j, P_k] = 0$. The Z_j are thus functions of \mathbf{P} . Eq. (7.22c) and $\xi \neq 0$ imply, furthermore, that

$$\partial_{P_j} Z_k = \partial_{P_k} Z_j \quad (7.66)$$

To alleviate notation, we set $\partial_{P_j} \equiv \partial_j$ in the rest of this section. It follows from Eqns. (7.22e) and (7.45) that

$$\begin{aligned} \sum_l \epsilon_{jkl} Z_l &= \frac{1}{i} [L_j, Z_k] = \sum_{mn} \epsilon_{mnj} \frac{1}{i} [X_m P_n, Z_k] \\ &= \sum_{mn} \epsilon_{mnj} \frac{1}{i} [X_m, Z_k] P_n = \sum_{mn} \epsilon_{mnj} (\partial_m Z_k) P_n \end{aligned} \quad (7.67)$$

or

$$\begin{aligned} j, k = 2, 3 : \quad Z_1 &= P_1 \partial_3 Z_3 - P_3 \partial_1 Z_3 \implies Z_1 + P_3 \partial_3 Z_1 = P_1 \partial_3 Z_3 \\ j, k = 3, 1 : \quad Z_2 &= P_2 \partial_1 Z_1 - P_1 \partial_2 Z_1 \implies Z_2 + P_1 \partial_1 Z_2 = P_2 \partial_1 Z_1 \\ j, k = 1, 2 : \quad Z_3 &= P_3 \partial_2 Z_2 - P_2 \partial_3 Z_2 \implies Z_3 + P_2 \partial_2 Z_3 = P_3 \partial_2 Z_2 \end{aligned} \quad (7.68)$$

The relationships $\partial_j Z_k = \partial_k Z_j$ have been used for the second step.

Note now that for each term $P_1^{n_1} P_2^{n_2} P_3^{n_3}$

$$P_k \partial_k (P_1^{n_1} P_2^{n_2} P_3^{n_3}) = n_k P_1^{n_1} P_2^{n_2} P_3^{n_3}, \quad n_k \geq 0 \quad (7.69)$$

which implies that Z_j must have the form

$$Z_j = P_j W_j(P_1, P_2, P_3) \quad (7.70)$$

With these substitutions, Eq. (7.68) become

$$\begin{aligned} W_1 - W_3 + P_3 \partial_3 (W_1 - W_3) &= 0 \\ W_2 - W_1 + P_1 \partial_1 (W_2 - W_1) &= 0 \\ W_3 - W_2 + P_2 \partial_2 (W_3 - W_2) &= 0 \end{aligned} \quad (7.71)$$

Using again Eq. (7.69), one concludes that $W_1 - W_3 = W_2 - W_1 = W_3 - W_2 = 0$ or

$$W_1 = W_2 = W_3 = W \quad (7.72)$$

It follows from Eq. (7.66) that $P_k \partial_j W = P_j \partial_k W$ i.e. that $\partial_j W \propto P_j$ and, finally, that W is a polynomial in $\mathbf{P}^2 \equiv P_1^2 + P_2^2 + P_3^2$. To summarize: Eqns. (7.33), (7.22c), (7.22e) imply that

$$Y_j = \xi X_j - \tau P_j - P_j \sum_{n=1}^{\infty} \tau_n \mathbf{P}^{2n} \quad (7.73)$$

with real constants $\xi \neq 0$, τ and τ_n .

The generator H

Eq. (7.22g) entails that H is a function of \mathbf{P} . With Eq. (7.40) and Eq. (7.22h), one obtains

$$0 = \sum_{j,k,l} \epsilon_{jkl} \frac{1}{i} [X_j P_k, H(\mathbf{P})] = \sum_{j,k,l} \epsilon_{jkl} \frac{1}{i} [X_j, H(\mathbf{P})] P_k = \sum_{j,k,l} \epsilon_{jkl} P_k \partial_{P_j} H \quad (7.74)$$

for $l \in \{1, 2, 3\}$, which implies that $P_k \partial_j H = P_j \partial_k H$ and, further, that $H = H(\mathbf{P}^2)$. Eqns. (7.22i) and (7.73) yield, finally, that $\xi \partial_{P_j} H = P_j$ for $j \in \{1, 2, 3\}$. The solution of this set of partial differential equations is

$$H = \frac{\mathbf{P}^2}{2\xi} + e_0 I \quad (7.75)$$

where e_0 is a real constant.

7.4.5 Actions of 1-parameter unitary subgroups generated by H , \mathbf{P} , \mathbf{L} , \mathbf{Y} on polynomials $T(\mathbf{X}, \mathbf{P})$

The actions of 1-parameter unitary subgroups e^{isK} on polynomials $T(\mathbf{X}, \mathbf{P})$ are now evaluated. From Eq. (7.36), it follows that the transformed polynomial T_s is given by

$$T_s = \exp\left(\frac{s}{i} K\right) T_0 \exp\left(-\frac{s}{i} K\right) \quad (7.76)$$

It suffices to evaluate this formula only for the components X_j and P_j because $T_0 = T_0^{(1)} T_0^{(2)}$ implies $T_s = T_s^{(1)} T_s^{(2)}$.

The following identity is helpful:

$$A(x) \equiv e^{xB} A e^{-xB} = A + x[B, A] + \frac{x^2}{2!} [B, [B, A]] + \frac{x^3}{3!} [B, [B, [B, A]]] \dots \equiv e^{x \text{ad}_B} A \quad (7.77)$$

where

$$\text{ad}_B A \equiv [B, A] \quad (7.78)$$

To prove this identity, integrate the differential equation $d_x A(x) = [B, A(x)]$, obtained from the definition of $A(x)$, from 0 to x and solve by iteration. Eq. (7.77) may also be checked by direct calculation [58, p. 61].

Let us now consider the subgroup e^{isH} . The action of this subgroup is determined by Eq. (7.22g) implying, first, that \mathbf{P} is constant in time in the initial system '0', secondly, that the value of \mathbf{P} remains the same in the time translated system 's', and, thirdly, that the value of H remains the same in the space translated system 'a'. As for the action of e^{isH} on the component X_j , Eqns. (7.76) and (7.77) yield

$$\begin{aligned} (X_j)_s &= \exp\left(\frac{s}{i} H\right) (X_j)_0 \exp\left(-\frac{s}{i} H\right) = (X_j)_0 + \frac{s}{2\xi} \frac{1}{i} [\mathbf{P}^2, (X_j)_0] \\ &= (X_j)_0 - \frac{s}{2\xi} \partial_{P_j} \mathbf{P}^2 = (X_j)_0 - \frac{s}{\xi} P_j \end{aligned} \quad (7.79)$$

All together, one gets

$$T_s(\mathbf{X}, \mathbf{P}) = \exp\left(\frac{s}{i} H\right) T_0(\mathbf{X}, \mathbf{P}) \exp\left(-\frac{s}{i} H\right) = T_0\left(\mathbf{X} - \frac{s}{\xi} \mathbf{P}, \mathbf{P}\right) \quad (7.80)$$

Setting $K = -P_j$ and $s = a_j$ in Eq. (7.76), one obtains from Eq. (7.77) that $(P_k)_{a_j} = (P_k)_0$ and $(X_k)_{a_j} = (X_k)_0 + a_j \delta_{jk} I$ or

$$T_{\mathbf{a}}(\mathbf{X}, \mathbf{P}) = e^{i(\mathbf{a}, \mathbf{P})} T_0(\mathbf{X}, \mathbf{P}) \exp^{-i(\mathbf{a}, \mathbf{P})} = T_0(\mathbf{X} + \mathbf{a}I, \mathbf{P}) \quad (7.81)$$

Consider now $K = -L_3$ and $s = \theta$. It follows from Eqns. (7.44) and (7.45) that $i[L_3, X_3] = 0$, $i[L_3, P_3] = 0$, and, for $j \in \{1, 2\}$,

$$\begin{aligned} i[L_3, X_j] &= X_1 \delta_{2,j} - X_2 \delta_{1,j} \\ i^2[L_3[L_3, X_j]] &= -X_j \end{aligned} \quad (7.82)$$

and

$$\begin{aligned} i[L_3, P_j] &= P_1 \delta_{2,j} - P_2 \delta_{1,j} \\ i^2[L_3[L_3, P_j]] &= -P_j \end{aligned} \quad (7.83)$$

The sum (7.77) in Eq. (7.76) being performed yields, in a coordinate free notation,

$$T_{\theta, \mathbf{e}}(\mathbf{X}, \mathbf{P}) = e^{i\theta(\mathbf{e}, \mathbf{L})} T_{0,0}(\mathbf{X}, \mathbf{P}) \exp^{-i\theta(\mathbf{e}, \mathbf{L})} = T_{0,0}(R(\theta, \mathbf{e}) \mathbf{X}, R(\theta, \mathbf{e}) \mathbf{P}) \quad (7.84)$$

where $R(\theta, \mathbf{e})$ denotes a rotation by an angle θ around the axis \mathbf{e} .

Consider, finally, $K = Y_j \equiv \xi X_j - \tau P_j - F_j(\mathbf{P})$ and $s = v_j$. It is then

$$\begin{aligned} \frac{1}{i} [Y_j, P_k] &= \xi \delta_{jk} \\ \frac{1}{i} [Y_j, X_k] &= \tau \delta_{jk} + \partial_{P_k} F_j \\ \frac{1}{i} [Y_j, \partial_{P_k} F_j] &= \xi \partial_{P_j} \partial_{P_k} F_j \end{aligned} \quad (7.85)$$

and so on. The transformation of \mathbf{P} follows immediately. In order to get the transformation formula for \mathbf{X} imposed by Tab. (7.11), τ must be set equal to t and τ_n equal to 0 in Eq. (7.73), which yields the expected expression

$$T_{\mathbf{v}}(\mathbf{X}, \mathbf{P}) = \exp\left(\frac{1}{i}(\mathbf{v}, \mathbf{Y})\right) T_0(\mathbf{X}, \mathbf{P}) \exp\left(-\frac{1}{i}(\mathbf{v}, \mathbf{Y})\right) = T_0(\mathbf{X} + t\mathbf{v}I, \mathbf{P} + \xi\mathbf{v}I) \quad (7.86)$$

and the definitive formula

$$\mathbf{Y} = \xi\mathbf{X} - t\mathbf{P} \quad (7.87)$$

for the generator \mathbf{Y} of proper Galilei transforms. Note that the expression for H , Eq. (7.75), does not depend on this choice for \mathbf{Y} .

7.5 The quantum mechanics of a particle without inner degrees of freedom

Until now, \mathbf{X} denotes some position in space described with a fixed reference frame and the quantities \mathbf{P} , \mathbf{L} , H are mere generators of 1-parameter unitary transformations on the Hilberspace $\mathbb{L}^2(\mathbb{R}^3)$ engendered by unitary ray representations of 1-parameter subgroups of Galilei transforms. Physical systems enter the scene if \mathbf{X} is considered as the position of a particle. The quantities \mathbf{P} , \mathbf{L} , and H are self-adjoint operators according to Stone's theorem and may thus be understood also as observables of the particle. This *dynamical meaning* of \mathbf{P} , \mathbf{L} , H is readily established [55] by calculating, in the Heisenberg picture, the velocity operator \mathbf{V} defined by the following identities between expectation values

$$\text{Tr}(\rho_s \mathbf{V}(t)) \equiv d_t \text{Tr}(\rho_s \mathbf{X}(t)) = \text{Tr}(\rho_s d_t \mathbf{X}(t)) \quad (7.88)$$

for all states s . It is, therefore, $\mathbf{V} = d_t \mathbf{X}$. The Heisenberg equation yields

$$V_j = d_t X_j = \frac{1}{i} [X_j, H] = \partial_{P_j} H = \frac{1}{\xi} P_j \quad (7.89)$$

or $\mathbf{P} = \xi \mathbf{V}$ implying $H = (\xi/2) \mathbf{V}^2 + e_0 I$. The dynamical meanings of \mathbf{P} , \mathbf{L} , and H are indeed the ones suggested by the notations chosen for the generators namely the momentum, the angular momentum, and the kinetic energy of a particle.

The quantity ξ was introduced as a real number parametrizing a set of nonequivalent unitary ray representations of the Galilei group. If one wishes to identify ξ with the mass of a particle then the range of ξ is restricted to positive values and the momentum \mathbf{P} cannot be equal to $\frac{1}{i} \nabla$ but at most proportional to it, which is obvious already from the physical dimensions of the quantities involved. The momentum operator must, therefore, be defined as

$$\mathbf{P} = \frac{\hbar}{i} \nabla \quad (7.90)$$

where $\hbar = h/(2\pi)$, h being Planck's constant. It follows that

$$m = \hbar \xi \quad (7.91)$$

and the generators \mathbf{L} , \mathbf{Y} , and H must also be multiplied by \hbar in order to get their dynamical counterparts. Finally, it suggests itself to replace $\frac{1}{i} [\cdot, \cdot]$ by the new bracket symbol

$$\{A, B\} \equiv \frac{1}{i\hbar} (AB - BA) \quad (7.92)$$

for an economical notation of the fundamental commutators

$$\begin{array}{ll} \text{(a)} \quad \{X_j, X_k\} = 0 & \text{(g)} \quad \{L_j, Y_k\} = \sum_{l=1}^3 \epsilon_{jkl} Y_l \\ \text{(b)} \quad \{X_j, P_k\} = \delta_{jk} & \text{(h)} \quad \{Y_j, P_k\} = m \delta_{jk} \\ \text{(c)} \quad \{P_j, P_k\} = 0 & \text{(i)} \quad \{P_j, H\} = 0 \\ \text{(d)} \quad \{L_j, L_k\} = \sum_{l=1}^3 \epsilon_{jkl} L_l & \text{(j)} \quad \{L_j, H\} = 0 \\ \text{(e)} \quad \{L_j, X_k\} = \sum_{l=1}^3 \epsilon_{jkl} X_l & \text{(k)} \quad \{Y_j, H\} = P_j \\ \text{(f)} \quad \{L_j, P_k\} = \sum_{l=1}^3 \epsilon_{jkl} P_l & \end{array} \quad (7.93)$$

between the observables \mathbf{X} , $\mathbf{P} = \frac{\hbar}{i} \nabla$, $\mathbf{L} = \mathbf{X} \wedge \mathbf{P}$, $H = \mathbf{P}^2/(2m) + e_0 I$, and $\mathbf{Y} = m\mathbf{X} - t\mathbf{P}$. With this result, the complete physical description of a free, nonrelativistic particle without spin is obtained, uniquely based on the general formal structure of quantum mechanics and on Galilei invariance.

7.6 An isolated system consisting of two interacting particles

A physically more interesting case than one isolated particle is the system of two particles united into one combined system by some external action. The system is isolated after the external action has ceased. The positions, momenta, angular momenta, energies, and masses of the particles are denoted by $\mathbf{X}^{(i)}$, $\mathbf{P}^{(i)}$, and so forth, $i = 1, 2$. It follows from the algebraic structure of time evolution in the Heisenberg picture that, besides $\{X_j^{(i)}, X_k^{(i)}\} = 0$, $\{P_j^{(i)}, P_k^{(i)}\} = 0$, and $\{X_j^{(i)}, P_k^{(i)}\} = \delta_{jk}$, also

$$\begin{aligned} \{X_j^{(1)}, X_k^{(2)}\}(t) &= 0 \\ \{P_j^{(1)}, P_k^{(2)}\}(t) &= 0 \\ \{X_j^{(1)}, P_k^{(2)}\}(t) &= 0 \\ \{P_j^{(1)}, X_k^{(2)}\}(t) &= 0 \end{aligned} \quad (7.94)$$

for all times t if these commutators are equal to 0 at one given time. This is assumed in the following. Let us define the total mass, the total momentum, the total angular momentum, and the total energy by

$$\begin{aligned} \text{(a)} \quad m &= m^{(1)} + m^{(2)} \\ \text{(b)} \quad \mathbf{P} &= \mathbf{P}^{(1)} + \mathbf{P}^{(2)} \\ \text{(c)} \quad \mathbf{L} &= \mathbf{L}^{(1)} + \mathbf{L}^{(2)} \\ \text{(d)} \quad H &= H^{(1)} + H^{(2)} + E \end{aligned} \tag{7.95}$$

$m^{(i)}$, $\mathbf{P}^{(i)}$, $\mathbf{L}^{(i)}$, $H^{(i)}$ denote the corresponding quantities of the free particle (i) and E the interaction energy. The quantities m , \mathbf{P} , \mathbf{L} are mere definitions at the moment. The quantities $X_j^{(i)}$, $P_k^{(i)}$ are considered as building blocks of the observable algebra of the combined system. The interaction energy E is thus a function of these basic observables.

It turns out that the quantities m , \mathbf{P} , \mathbf{L} and H describe the external observables of the combined system in a manner that has to be specified now. Eqns. (7.95a) and (7.95b) entail the existence of a specific space variable characterizing the total system namely its center of mass coordinate \mathbf{X} :

$$\mathbf{P} \equiv m \, d_t \mathbf{X} = \mathbf{P}^{(1)} + \mathbf{P}^{(2)} = m^{(1)} \, d_t \mathbf{X}^{(1)} + m^{(2)} \, d_t \mathbf{X}^{(2)} \tag{7.96}$$

or

$$\mathbf{X} = \frac{m^{(1)}}{m} \mathbf{X}^{(1)} + \frac{m^{(2)}}{m} \mathbf{X}^{(2)} \tag{7.97}$$

if scale invariance is assumed.

The total angular momentum \mathbf{L} is now decomposed into

$$\mathbf{L} = \mathbf{L}^{(1)} + \mathbf{L}^{(2)} \equiv \mathbf{L}^{\text{cm}} + \mathbf{L}^{\text{rel}} \tag{7.98}$$

with

$$\mathbf{L}^{\text{cm}} \equiv \mathbf{X} \wedge \mathbf{P} \tag{7.99}$$

implying that

$$\begin{aligned} \mathbf{L}^{\text{rel}} &= \mathbf{X}^{(1)} \wedge \mathbf{P}^{(1)} + \mathbf{X}^{(2)} \wedge \mathbf{P}^{(2)} - \mathbf{X} \wedge \mathbf{P} \\ &= (\mathbf{X}^{(2)} - \mathbf{X}^{(1)}) \wedge \left(\frac{m^{(1)}}{m} \mathbf{P}^{(2)} - \frac{m^{(2)}}{m} \mathbf{P}^{(1)} \right) \\ &\equiv \mathbf{X}^{\text{rel}} \wedge \mathbf{P}^{\text{rel}} \end{aligned} \tag{7.100}$$

where

$$\mathbf{X}^{\text{rel}} = \mathbf{X}^{(2)} - \mathbf{X}^{(1)} \tag{7.101}$$

and

$$\mathbf{P}^{\text{rel}} = \frac{m^{(1)}}{m} \mathbf{P}^{(2)} - \frac{m^{(2)}}{m} \mathbf{P}^{(1)} = \frac{m^{(1)}m^{(2)}}{m} (\mathbf{V}^{(2)} - \mathbf{V}^{(1)}) = \mu \, d_t \mathbf{X}^{\text{rel}} \tag{7.102}$$

The quantity

$$\mu \equiv \frac{m^{(1)}m^{(2)}}{m} \tag{7.103}$$

denotes the *reduced mass* of the relative system.

It follows from Eqns. (7.94) and the above definitions that

$$\begin{aligned}
\{X_j, X_k\} &= 0 & \{X_j^{\text{rel}}, X_k^{\text{rel}}\} &= 0 \\
\{P_j, P_k\} &= 0 & \{P_j^{\text{rel}}, P_k^{\text{rel}}\} &= 0 \\
\{X_j, P_k\} &= \delta_{jk} & \{X_j^{\text{rel}}, P_k^{\text{rel}}\} &= \delta_{jk} \\
\{X_j, X_k^{\text{rel}}\} &= 0 & \{P_j, X_k^{\text{rel}}\} &= 0 \\
\{X_j, P_k^{\text{rel}}\} &= 0 & \{P_j, P_k^{\text{rel}}\} &= 0
\end{aligned} \tag{7.104}$$

with $j, k \in \{1, 2, 3\}$. These equations also imply that

$$\{L_j^{\text{cm}}, L_k^{\text{rel}}\} = 0 \tag{7.105}$$

It is, finally,

$$\begin{aligned}
H &= H^{(1)} + H^{(2)} + E = \frac{\|\mathbf{P}^{(1)}\|^2}{2m^{(1)}} + \frac{\|\mathbf{P}^{(2)}\|^2}{2m^{(2)}} + E \\
&= \frac{\|\mathbf{P}\|^2}{2m} + \frac{\|\mathbf{P}^{\text{rel}}\|^2}{2\mu} + E \equiv H^{\text{cm}} + H^{\text{rel}}
\end{aligned} \tag{7.106}$$

where H^{cm} is the kinetic energy of the center of mass system and the interaction energy E is attributed to H^{rel} . The coordinates $\mathbf{X}^{(i)}$ and the momenta $\mathbf{P}^{(i)}$ can be expressed as linear combinations of \mathbf{X} , \mathbf{X}^{rel} and of \mathbf{P} , \mathbf{P}^{rel} , respectively. We may, therefore consider E as a function of \mathbf{X} , \mathbf{P} , \mathbf{X}^{rel} , and \mathbf{P}^{rel} .

One easily verifies that the components of \mathbf{P} , \mathbf{L} , and $\mathbf{Y} = m\mathbf{X} - t\mathbf{P}$ fulfill the commutation relationships (7.93c) – (7.93h), *i.e.* $\frac{1}{\hbar}(\mathbf{e}, \mathbf{P})$, $\frac{1}{\hbar}(\mathbf{e}, \mathbf{L})$, and $\frac{1}{\hbar}(\mathbf{e}, \mathbf{Y})$ generate the unitary ray representations of the corresponding 1-parameter subgroups of the Galilei group. The remaining equations (7.93i) – (7.93k) involving the total energy H are valid in the case of noninteracting particles, $E = e_0 I$. The total energy H is then invariant under translations, rotations, and proper Galilei transforms and H , \mathbf{P} , \mathbf{L} are constants of motion.

We assume now the validity of these equations also for interacting particles and investigate the consequences for the structure of the interaction energy E . It is with Eq. (7.93i)

$$0 = \{P_j, H\} = \{P_j, H^{\text{cm}}\} + \{P_j, H^{\text{rel}}\} = \{P_j, H^{\text{rel}}\} = \{P_j, E\} \tag{7.107}$$

yielding

$$\{P_j, E\} = \partial_{X_j} E = 0 \tag{7.108}$$

and with Eq. (7.93k)

$$\begin{aligned}
P_j &= \{Y_j, H\} = \{mX_j - tP_j, H^{\text{cm}} + H^{\text{rel}}\} = m\{X_j, H^{\text{cm}}\} + \{mX_j - tP_j, E\} \\
&= P_j + m\{X_j, E\}
\end{aligned} \tag{7.109}$$

yielding

$$\{X_j, E\} = \partial_{P_j} E = 0 \tag{7.110}$$

which means that E depends only on the relative quantities

$$E = E(\mathbf{X}^{\text{rel}}, \mathbf{P}^{\text{rel}}) \tag{7.111}$$

and that

$$\{H^{\text{cm}}, H^{\text{rel}}\} = 0 \tag{7.112}$$

i.e. H^{cm} and H^{rel} are both constants of motion.

It follows, finally, with Eq. (7.93j) that

$$\begin{aligned} 0 &= \{L_j, H\} = \{L_j^{\text{cm}}, H\} + \{L_j^{\text{rel}}, H\} \\ &= \{L_j^{\text{cm}}, H^{\text{cm}}\} + \{L_j^{\text{cm}}, H^{\text{rel}}\} + \{L_j^{\text{rel}}, H\} = \{L_j^{\text{rel}}, H\} \end{aligned} \quad (7.113)$$

i.e. \mathbf{L}^{cm} and \mathbf{L}^{rel} are both conserved quantities. Furthermore

$$\begin{aligned} 0 &= \{L_j, H\} = \{L_j, H^{\text{cm}}\} + \{L_j, H^{\text{rel}}\} \\ &= \{L_j^{\text{cm}}, H^{\text{cm}}\} + \{L_j, E\} = \{L_j, E\} \end{aligned} \quad (7.114)$$

and, therefore,

$$\begin{aligned} E(R_{\theta, \mathbf{e}} \mathbf{X}^{\text{rel}}, R_{\theta, \mathbf{e}} \mathbf{P}^{\text{rel}}) &= \exp\left(\frac{i}{\hbar} \theta(\mathbf{e}, \mathbf{L})\right) E(\mathbf{X}^{\text{rel}}, \mathbf{P}^{\text{rel}}) \exp\left(-\frac{i}{\hbar} \theta(\mathbf{e}, \mathbf{L})\right) \\ &= E(\mathbf{X}^{\text{rel}}, \mathbf{P}^{\text{rel}}) \end{aligned} \quad (7.115)$$

for arbitrary rotations. The interaction energy only depends on the rotation invariant quantities $\|\mathbf{X}^{\text{rel}}\|^2$, $\|\mathbf{P}^{\text{rel}}\|^2$ and $(\mathbf{X}^{\text{rel}}, \mathbf{P}^{\text{rel}}) + (\mathbf{P}^{\text{rel}}, \mathbf{X}^{\text{rel}})$. This is a reduction from 12 to 3 variables as possible independent arguments of E . Conversely, if $E = E(\|\mathbf{X}^{\text{rel}}\|^2, \|\mathbf{P}^{\text{rel}}\|^2, (\mathbf{X}^{\text{rel}}, \mathbf{P}^{\text{rel}}) + (\mathbf{P}^{\text{rel}}, \mathbf{X}^{\text{rel}}))$ then \mathbf{X} , \mathbf{P} , \mathbf{L} , $\mathbf{Y} = m\mathbf{X} - t\mathbf{P}$, and H fulfil all commutation relations (7.93), and are, therefore, related to the generators of 1-parameter subgroups of an unitary ray representation of the Galilei group.

To summarize: all center of mass observables commute with the corresponding relative quantities, and the observables \mathbf{P} , \mathbf{L}^{cm} , H^{cm} , \mathbf{L}^{rel} , H^{rel} are constant in time. The isolated system of two interacting particles decomposes in a natural way into two noninteracting subsystems, the center of mass system and the relative system. The commutation relationships (7.44), (7.45), and (7.46) between the components of \mathbf{X} , \mathbf{P} , and \mathbf{L} separately apply to the center of mass and to the relative quantities. H^{cm} generates the time evolution of the center of mass system, H^{rel} of the relative system.

The center of mass system can be considered as a free particle if, in the hitherto valid definition of a particle, the constant kinetic energy H^{cm} is completed with a constant internal energy $H^{\text{int}} \equiv H^{\text{rel}}$ to the total energy $H = H^{\text{cm}} + H^{\text{int}}$ and the constant angular momentum $\mathbf{L}^{\text{cm}} = \mathbf{X} \wedge \mathbf{P}$ is completet with a constant internal angular momentum or spin $\mathbf{S} \equiv \mathbf{L}^{\text{rel}}$ to the total angular momentum $\mathbf{L} = \mathbf{L}^{\text{cm}} + \mathbf{S}$. Both internal observables commute with \mathbf{X} and \mathbf{P} and are constants of motion. The spin transforms like a vector

$$\{L_j, S_k\} = \sum_{l=1}^3 \epsilon_{jkl} S_l \quad (7.116)$$

which implies, together with $\{L_j^{\text{cm}}, S_k\} = 0$, that

$$\{S_j, S_k\} = \sum_{l=1}^3 \epsilon_{jkl} S_l \quad (7.117)$$

Appendix A

Elements of probability theory

In this Appendix, central notions and concepts of probability theory are compiled. We follow the axiomatics of A. N. Kolmogorov which is the generally accepted basis of probability theory and allows to treat discrete and continuous probability distributions on an equal footing (for a reference to probability theory see e.g. [62]).

Let us recall first the basic definitions of set theory [63]:

The *void* or *empty* set, denoted \emptyset , is the set with no element. If every element of a set σ_1 is also an element of the set σ_2 , the σ_1 is a *subset* of σ_2 , denoted $\sigma_1 \subseteq \sigma_2$. $\sigma_1 = \sigma_2$ means $\sigma_1 \subseteq \sigma_2$ and $\sigma_2 \subseteq \sigma_1$. The *union* of the sets σ_1 and σ_2 , denoted $\sigma_1 \cup \sigma_2$, is the set of all elements belonging either to σ_1 or to σ_2 or to both. The *intersection* of σ_1 and σ_2 , denoted $\sigma_1 \cap \sigma_2$, is the set of all elements which belong to both, σ_1 and σ_2 . Two sets are *disjoint* if their intersection is void. Let σ be a subset of Ω . The *complement of σ relative to Ω* is the totality of elements $\omega \in \Omega$ not contained in σ . It is denoted by $\sigma \setminus \Omega$ or by $\bar{\sigma}$. The *difference* of two sets $\sigma_1 \subseteq \Omega$ and $\sigma_2 \subseteq \Omega$, denoted $\sigma_1 - \sigma_2$, is the totality of elements $\omega \in \sigma_1$ which are not contained in σ_2 : $\sigma_1 - \sigma_2 = \sigma_1 \cap \bar{\sigma}_2$.

The basic term of probability theory is the *random experiment*, the possible outcomes ω of which form the *sample space* or the *space of elementary events* Ω . A subset $\sigma \in \Omega$ is called a (*random*) *event*: "*The outcome of the random experiment lies in σ* ". The set Σ of random events fulfills the following requirements:

1. Ω is an element of Σ .
2. If σ_1 and σ_2 are subsets of Ω and elements of Σ , then their complements $\bar{\sigma}_i$ with respect to Ω , the union $\sigma_1 \cup \sigma_2$, and the intersection $\sigma_1 \cap \sigma_2$ are elements of Σ .
3. If the countably infinite subsets $\sigma_1, \sigma_2, \dots$ of Ω are all elements of Σ , then their union $\sigma_1 \cup \sigma_2 \cup \dots \equiv \bigcup_{n=1}^{\infty} \sigma_n$ and their intersection $\sigma_1 \cap \sigma_2 \cap \dots \equiv \bigcap_{n=1}^{\infty} \sigma_n$ are elements of Σ .

These three conditions specify a σ -*field*. Together with Ω , the empty set \emptyset is an element of Σ . Ω is the *certain event*, \emptyset the *impossible event*.

The *probability* is a real function P on Σ with the following properties (Axioms of Kolmogorov):

1. The probability $P(\sigma)$ of any event $\sigma \in \Sigma$ is a real number with $0 \leq P(\sigma) \leq 1$.
2. $P(\Omega) = 1$.
3. Let $\sigma_1, \sigma_2, \dots$ denote a finite or a countably infinite set of pairwise disjoint random events, $\sigma_n \cap \sigma_m = \emptyset$ for all $n \neq m$. Then $P(\bigcup_n \sigma_n) = \sum_n P(\sigma_n)$: P is σ -*additive*.

A function $P : \Sigma \rightarrow [0, 1]$ with these properties is a *probability measure* and the triplet (Ω, Σ, P) a *probability space*.

It is $\sigma_1 \cup \sigma_2 = \sigma_1 \cup (\sigma_2 - (\sigma_1 \cap \sigma_2))$ and $\sigma_2 = (\sigma_1 \cap \sigma_2) \cup (\sigma_2 - (\sigma_1 \cap \sigma_2))$. The righthand sides are unions of disjoint sets. Therefore, $P(\sigma_1 \cup \sigma_2) = P(\sigma_1) + P(\sigma_2 - (\sigma_1 \cap \sigma_2))$ and $P(\sigma_2) = P(\sigma_1 \cap \sigma_2) + P(\sigma_2 - (\sigma_1 \cap \sigma_2))$ and together

$$P(\sigma_1 \cup \sigma_2) = P(\sigma_1) + P(\sigma_2) - P(\sigma_1 \cap \sigma_2) \quad (\text{A.1})$$

The *conditional probability* $P(\sigma | \sigma_1)$ is the probability of the event σ under the condition that the event σ_1 with $P(\sigma_1) > 0$ occurs or has occurred. It is

$$P(\sigma | \sigma_1) = \frac{P(\sigma \cap \sigma_1)}{P(\sigma_1)} \quad (\text{A.2})$$

The conditional probability $P(\sigma | \sigma_1)$ is a probability with probability space $(\sigma_1, \Sigma \cap \sigma_1, P(\sigma \cap \sigma_1)/P(\sigma_1))$. Two random events σ_1 and σ_2 are *mutually independent* if $P(\sigma_1 | \sigma_2) = P(\sigma_1)$, $P(\sigma_2) > 0$, or $P(\sigma_2 | \sigma_1) = P(\sigma_2)$, $P(\sigma_1) > 0$: the occurrence of σ_2 does not change the probability of σ_1 and vice versa. The two conditions amount to the criterion $P(\sigma_1 \cap \sigma_2) = P(\sigma_1) P(\sigma_2)$ for the independence of the random events σ_1 and σ_2 .

The case $\Omega = \mathbb{R}$, the elementary events being real numbers, is of particular importance for our purposes. The set of random events is then the *Borel σ -field* \mathcal{B} of real *Borel sets* σ . \mathcal{B} is generated by the open intervals (a, b) or $(a, +\infty)$ or $(-\infty, a)$, or equivalently by the closed intervals $[a, b]$, or by the semiclosed intervals $(a, b]$ or $[a, b)$ or $(-\infty, a]$ or $[a, +\infty)$ [63, p. 33].

A *stochastic variable* X is a measurable real function on the sample space Ω of a probability space (Ω, Σ, P) . X being *measurable* means that the *full inverse* $X^{-1}(\sigma) \equiv \{\omega \in \Omega / X(\omega) \in \sigma\}$ lies in Σ for any real Borel set σ . The stochastic variable X induces a real probability measure P_X on the Borel σ -field \mathcal{B} by $P_X(\sigma) \equiv P(X^{-1}(\sigma))$, $\sigma \in \mathcal{B}$, i.e. $(\mathbb{R}, \mathcal{B}, P_X)$ is a probability space. With respect to this probability space, the stochastic variable X is simply $X : x \in \Omega = \mathbb{R} \rightarrow x \in \mathbb{R}$. Instead of $P_X(\sigma)$ one commonly writes $P(X \in \sigma)$, or for simple cases, e.g. for $\sigma = (-\infty, x]$, $P_X((-\infty, x]) \equiv P(X \leq x)$.

Let $f : x \in D_f \rightarrow f(x) \in \Delta_f$ be a single valued, measurable, real function or *Borel function*, $D_f \subseteq \mathbb{R}$ denoting the *domain* and $\Delta_f \subseteq \mathbb{R}$ the *range* of f . The full inverse of f is primarily defined for $\sigma \subseteq \Delta_f$ but is readily extended to arbitrary Borel sets. Set $f^{-1}(\sigma) = \emptyset$ for $\sigma \subseteq \overline{\Delta_f}$ and use the identity $f^{-1}(\sigma_1 \cup \sigma_2) = f^{-1}(\sigma_1) \cup f^{-1}(\sigma_2)$. It is then $f^{-1}(\sigma) = f^{-1}(\sigma \cap \Delta_f) \subseteq D_f$ for all Borel sets σ . The probability measure of the stochastic variable $f(X)$ is given by

$$P_{f(X)}(\sigma) = P_X(f^{-1}(\sigma)), \sigma \in \mathcal{B} \quad (\text{A.3})$$

because $f(X)^{-1} = X^{-1}(f^{-1})$. The stochastic variable $f(X)$ corresponds to the function $f : x \in \Omega = \mathbb{R} \rightarrow f(x) \in \mathbb{R}$ in the probability space $(\mathbb{R}, \mathcal{B}, P_X)$.

The *distribution function* $F_X(x)$ is defined by

$$F_X(x) = P_X((-\infty, x]) \equiv P(X \leq x) \quad (\text{A.4})$$

and characterized by the following properties:

1. $0 \leq F_X(x) \leq 1$
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$
3. The function $F_X(x)$ is continuous from the right : $\lim_{\epsilon \downarrow 0} F_X(x + \epsilon) = F_X(x)$

$F_X(x)$ is *discontinuous* at $x = x_0$ if $\lim_{\epsilon \downarrow 0} (F_X(x_0 + \epsilon) - F_X(x_0 - \epsilon)) = r_0 > 0$. A distribution function is discontinuous at most at countably infinite values of its argument [62, p. 123].

The probability measure can be reconstructed from its distribution function because it is $P_X((x_1, x_2]) = F_X(x_2) - F_X(x_1)$ for $x_1 < x_2$. The probability of arbitrary measurable intervals can then be obtained by using σ -additivity.

The *expectation value* $E(X)$ of the stochastic variable X is given by the Stieltjes integral [60, Chapter III]

$$E(X) = \int_{-\infty}^{\infty} x dF_X(x) \quad (\text{A.6})$$

and the *variance* $\text{Var}(X)$ by

$$\text{Var}(X) \equiv E\left((X - E(X))^2\right) = E(X^2) - E(X)^2 \quad (\text{A.7})$$

The expectation value of the stochastic variable $f(X)$ equals

$$E(f(X)) = \int_{-\infty}^{\infty} f(x) dF_X(x) \quad (\text{A.8})$$

The nonnegative *probability density* $p_X(x)$ of a stochastic variable X is defined by $dF_X(x) = p_X(x) dx$. The density $p_X(x)$ contains delta distributions at the discontinuity points of the distribution function $F_X(x)$. It is $\int_{-\infty}^{+\infty} p_X(x) dx = \int_{-\infty}^{+\infty} dF_X(x) = F_X(+\infty) - F_X(-\infty) = 1$.

The above terminology may be generalized to a set of N stochastic variables $\mathbf{X} = (X_1, X_2, \dots, X_N)$ acting on a common real sample space $\Omega^N = \mathbb{R}^N = \{(r_1, r_2, \dots, r_N)\}$. The Borel σ -field Σ^N is then generated by the subsets $a_1 < r_1 < b_1, a_2 < r_2 < b_2, \dots, a_N < r_N < b_N$, the probability $P_{\mathbf{X}}(\sigma)$ is a σ -additive probability measure on the subsets $\sigma \subset \Sigma^N$. A set $\underline{Y} \equiv (Y_1, Y_2, \dots, Y_M)$ is a real function of \mathbf{X} , $\underline{Y} = \underline{f}(\mathbf{X}) = (f_1(\mathbf{X}), f_2(\mathbf{X}), \dots, f_M(\mathbf{X}))$ if the full inverse $\underline{f}^{-1}(\sigma)$ is an element of Σ^N for every Borel set $\sigma \subset \mathbb{R}^M$. The identity

$$P_{\underline{Y}}(\sigma) = P_{\mathbf{X}}(\underline{f}^{-1}(\sigma)), \quad \sigma \in \Sigma^M \quad (\text{A.9})$$

defines a probability space $(\mathbb{R}^M, \Sigma^M, P_{\underline{Y}})$.

The distribution function Eq. (A.4) is generalized to

$$F_{\mathbf{X}}(\mathbf{x}) = P_{\mathbf{X}}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_N \leq x_N) \quad (\text{A.10})$$

If $P_{\mathbf{X}}(\cdot)$ admits a probability density $p_{\mathbf{X}}(\mathbf{x})$ then

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_N} dy^N p_{\mathbf{X}}(\mathbf{y}) \quad (\text{A.11})$$

A set of stochastic variables $\mathbf{X} = (X_1, X_2, \dots, X_N)$ is *mutually independent* if its probability distribution factorizes:

$$p_{\mathbf{X}}(x_1, x_2, \dots, x_N) = \prod_{n=1}^N p_{X_n}(x_n) \quad (\text{A.12})$$

Let us consider now a stochastic variable X with only two different outcomes $x_1 < x_2$ in a random experiment. Its probability measure is concentrated on the two points $x_1 < x_2$ or, equivalently, its distribution function is discontinuous at the points x_1 and x_2 and constant otherwise. The function $f(x) = (x - x_1)/(x_2 - x_1)$ maps $x_1 \rightarrow 0$ and $x_2 \rightarrow 1$. The probability measure of the observable $Q_X = f(X)$ is concentrated on the points 0 and 1. We may call such a stochastic variable a *proposition*. Q_X corresponds to the assertion "*The outcome of X in a random experiment is equal to x_2* ". The assertion is true if the outcome of Q_X is equal to 1. The expectation value of a proposition Q satisfies the inequality

$$0 \leq E(Q) \leq 1 \quad (\text{A.13})$$

In addition, it fulfils the equation $Q^2 = Q$ i.e.

$$P_{Q^2}(\sigma) \equiv P_Q((x^2)^{-1}(\sigma)) = P_Q(\sigma) \quad (\text{A.14})$$

for all Borel sets σ . This result is readily verified by considering the cases $\sigma \in (-\infty, 0)$, $\sigma = \{0\}$, $\sigma \in (0, 1)$, $\sigma = \{1\}$, $\sigma \in (1, +\infty)$ and using the σ -additivity of probability measures. As a consequence, the variance of a proposition Q is given by

$$\text{Var}(Q) = E(Q^2) - E(Q)^2 = E(Q) - E(Q)^2 = E(Q) (1 - E(Q)) \quad (\text{A.15})$$

$\text{Var}(Q) = 0$ is equivalent to $E(Q) = 0$ or $E(Q) = 1$.

Propositions are completely determined by their expectation values:

$$P_Q(\sigma) = \begin{cases} 0 & \text{if } 0 \notin \sigma, 1 \notin \sigma \\ E(Q) & \text{if } 0 \notin \sigma, 1 \in \sigma \\ 1 - E(Q) & \text{if } 0 \in \sigma, 1 \notin \sigma \\ 1 & \text{if } 0 \in \sigma, 1 \in \sigma \end{cases} \quad (\text{A.16})$$

The *characteristic function* $\chi_\sigma : \mathbb{R} \rightarrow \{0, 1\}$ of a Borel set σ is given by

$$\chi_\sigma(x) = \begin{cases} 1 & \text{if } x \in \sigma \\ 0 & \text{if } x \notin \sigma \end{cases} \quad (\text{A.17})$$

$\chi_\sigma(x)$ is a Borel function. For consistency, one sets $\chi_\emptyset(x) = 0$ for all $x \in \mathbb{R}$. It is then

$$\begin{aligned} \chi_{\bar{\sigma}} &= 1 - \chi_\sigma \\ \chi_{\sigma_1 \cap \sigma_2} &= \chi_{\sigma_1} \chi_{\sigma_2} \\ \chi_{\sigma_1 \cup \sigma_2} &= \chi_{\sigma_1} + \chi_{\sigma_2} - \chi_{\sigma_1} \chi_{\sigma_2} \end{aligned} \quad (\text{A.18})$$

The stochastic variables $\chi_\sigma(X)$ are propositions: "*the value of the stochastic variable X in a random experiment lies in the Borel set σ* ". The proposition is true if the value of $\chi_\sigma(X)$ is 1. The probability measure $P_X(\sigma)$ can be recovered as the expectation value of $\chi_\sigma(X)$

$$P_X(\sigma) = E(\chi_\sigma(X)) \quad (\text{A.19})$$

The *spectral decomposition of the observable A* may be defined by

$$E_A(x) \equiv \chi_{(-\infty, x]}(A) \quad (\text{A.20})$$

It has the following properties:

1. If $x_1 \leq x_2$ then $E_A(x_1) E_A(x_2) = E_A(x_2) E_A(x_1) = E_A(x_1)$
 2. $\lim_{\epsilon \downarrow 0} E_A(x + \epsilon) = E_A(x)$
 3. $\lim_{x \rightarrow -\infty} E_A(x) = 0$
 4. $\lim_{x \rightarrow +\infty} E_A(x) = 1$
 5. $E(E_A(x)) = F_A(x)$
- (A.21)

There is a striking analogy to the spectral decomposition of a self-adjoint operator on a Hilbert space [60, Sec. 120]. $E_A(x)$ is then a projection operator, the expectation value in the state s is given by the trace formula, and the distribution function $F_{A,s}(x)$ by Eq. (2.8).

Appendix B

The Chebyshev inequality

Consider a nonnegative stochastic variable X with probability density $p_X(x)$. It is $p_X(x) = 0$ for $x < 0$. For any $\epsilon > 0$, the following inequalities are obvious

$$E(X) = \int_0^\infty x p_X(x) dx \geq \int_\epsilon^\infty x p_X(x) dx \geq \epsilon \int_\epsilon^\infty p_X(x) dx \equiv \epsilon P(X \geq \epsilon)$$

or $P(X \geq \epsilon) \leq E(X)/\epsilon$. Substituting $|X - r|$ for X , where r is a real constant, yields, with $\alpha > 0$,

$$P(|X - r| \geq \epsilon) = P(|X - r|^\alpha \geq \epsilon^\alpha) \leq \frac{E(|X - r|^\alpha)}{\epsilon^\alpha}$$

Setting $r = E()$ and $\alpha = 2$ gives

$$P(|X - E()| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2} \quad (\text{B.1})$$

one of the forms of Chebyshev's inequality.

Let us now denote by $\mathbf{X} = (X_1, X_2, \dots, X_N)$ a set of N mutually independent copies of the stochastic variable X , *i.e.* $p_{X_n}(x) = p_X(x)$ for all $n = 1, 2, \dots, N$ and $p_{\mathbf{X}}(x_1, x_2, \dots, x_N) = \prod_{n=1}^N p_X(x_n)$ and let

$$\bar{X} \equiv \frac{1}{N} \sum_{n=1}^N X_n \quad (\text{B.2})$$

Direct calculation yields

$$E(\bar{X}) = E(X) \quad (\text{B.3})$$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{N} \quad (\text{B.4})$$

Therefore from Eq. (B.1)

$$P\left(\left|\frac{1}{N} \sum_{n=1}^N X_n - E(X)\right| \geq \epsilon\right) \leq \frac{\text{Var}(X)}{N\epsilon^2} \quad (\text{B.5})$$

Appendix C

Simultaneous eigenstates for Mermin's observables $A_1 - A_4$

The topic of this appendix is, first, to evaluate a system of common normalized eigenvectors of the observables A_1, A_2, A_3, A_4 , and, secondly, to show that the state of each particle, considered as isolated, equals $\rho_0 = \frac{1}{2} I$.

We start with the eigenvectors $|+\rangle$ and $|-\rangle$ of σ_3 with eigenvalues ± 1 . It is $\sigma_1|+\rangle = |-\rangle$ and $\sigma_1|-\rangle = |+\rangle$. The eight different products $|\pm\rangle^a \otimes |\pm\rangle^b \otimes |\pm\rangle^c \equiv |\pm\pm\pm\rangle$ form an orthonormal basis $\{\phi_l\}$, $l \in \{1, 2, \dots, 8\}$, of the Hilbertspace $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ of the total system. These eight basis vectors and the actions of the observables A_1, \dots, A_4 on them are listed in the table below

l	ϕ_l	$A_1\phi_l$	$A_2\phi_l$	$A_3\phi_l$	$A_4\phi_l$
1	$\phi_1 = +++\rangle$	ϕ_1	ϕ_5	ϕ_6	ϕ_7
2	$\phi_2 = - ++\rangle$	$-\phi_2$	ϕ_3	ϕ_4	$-\phi_8$
3	$\phi_3 = + - +\rangle$	$-\phi_3$	ϕ_2	$-\phi_8$	ϕ_4
4	$\phi_4 = + + -\rangle$	$-\phi_4$	$-\phi_8$	ϕ_2	ϕ_3
5	$\phi_5 = - - +\rangle$	ϕ_5	ϕ_1	$-\phi_7$	$-\phi_6$
6	$\phi_6 = - + -\rangle$	ϕ_6	$-\phi_7$	ϕ_1	$-\phi_5$
7	$\phi_7 = + - -\rangle$	ϕ_7	$-\phi_6$	$-\phi_5$	ϕ_1
8	$\phi_8 = - - -\rangle$	$-\phi_8$	$-\phi_4$	$-\phi_3$	$-\phi_2$

(C.1)

Let now

$$\psi = \sum_{l=1}^8 r_l \phi_l \quad (C.2)$$

be a simultaneous eigenvector of A_1, A_2, A_3, A_4 *i.e.*

$$A_k \psi = a_k \psi, \quad k = 1, \dots, 4, \quad a_k = \pm 1 \quad (C.3)$$

The expansion coefficients r_l can be chosen real, because only real matrices are involved. The following table shows the effects of the four eigenvalue equations on the coefficients r_l in Eq. (C.2):

$A_1\psi = a_1\psi$	$A_2\psi = a_2\psi$	$A_3\psi = a_3\psi$	$A_4\psi = a_4\psi$
$r_1 = a_1r_1$	$r_5 = a_2r_1$	$r_6 = a_3r_1$	$r_7 = a_4r_1$
$-r_2 = a_1r_2$	$r_3 = a_2r_2$	$r_4 = a_3r_2$	$-r_8 = a_4r_2$
$-r_3 = a_1r_3$	$r_2 = a_2r_3$	$-r_8 = a_3r_3$	$r_4 = a_4r_3$
$-r_4 = a_1r_4$	$-r_8 = a_2r_4$	$r_2 = a_3r_4$	$r_3 = a_4r_4$
$r_5 = a_1r_5$	$r_1 = a_2r_5$	$-r_7 = a_3r_5$	$-r_6 = a_4r_5$
$r_6 = a_1r_6$	$-r_7 = a_2r_6$	$r_1 = a_3r_6$	$-r_5 = a_4r_6$
$r_7 = a_1r_7$	$-r_6 = a_2r_7$	$-r_5 = a_3r_7$	$r_1 = a_4r_7$
$-r_8 = a_1r_8$	$-r_4 = a_2r_8$	$-r_3 = a_3r_8$	$-r_2 = a_4r_8$

(C.4)

Choose now $a_1 = 1$ and $r_1 = r$. It follows

$$r_2 = r_3 = r_4 = r_8 = 0, \quad r_5 = a_2r, \quad r_6 = a_3r, \quad r_7 = a_4r$$

and $a_1a_2a_3a_4 = -1$. Similarly, with $a_1 = -1$ and $r_2 = r$, one obtains

$$r_1 = r_5 = r_6 = r_7 = 0, \quad r_3 = a_2r, \quad r_4 = a_3r, \quad r_8 = -a_4r$$

and again $a_1a_2a_3a_4 = -1$. The eight simultaneous eigenvectors ψ_j are, therefore, determined by the eigenvalues a_1, a_2 , and a_3 :

j	ψ_j	$a_{1,j}$	$a_{2,j}$	$a_{3,j}$	$r_{j,1}$	$r_{j,2}$	$r_{j,3}$	$r_{j,4}$	$r_{j,5}$	$r_{j,6}$	$r_{j,7}$	$r_{j,8}$	$a_{4,j}$
1	ψ_1	+	+	+	r	0	0	0	r	r	$-r$	0	-
2	ψ_2	+	+	-	r	0	0	0	r	$-r$	r	0	+
3	ψ_3	+	-	+	r	0	0	0	$-r$	r	r	0	+
4	ψ_4	+	-	-	r	0	0	0	$-r$	$-r$	$-r$	0	-
5	ψ_5	-	+	+	0	r	r	r	0	0	0	$-r$	+
6	ψ_6	-	+	-	0	r	r	$-r$	0	0	0	r	-
7	ψ_7	-	-	+	0	r	$-r$	r	0	0	0	r	-
8	ψ_8	-	-	-	0	r	$-r$	$-r$	0	0	0	$-r$	+

(C.5)

It follows from $\langle \psi_j | \psi_j \rangle = 1$ that of r can be set equal to $\frac{1}{2}$.

The expectation values of the observables σ_k^ξ with $k = \{1, 2, 3\}$ and $\xi = \{a, b, c\}$ are now readily evaluated by observing that all ψ_j are linear combinations either of $\phi_1, \phi_5, \phi_6, \phi_7$ or of $\phi_2, \phi_3, \phi_4, \phi_8$.

$\phi_1 = +++>$	$\phi_2 = -++>$
$\phi_5 = --+>$	$\phi_3 = +-+>$
$\phi_6 = -+->$	$\phi_4 = ++->$
$\phi_7 = +-->$	$\phi_8 = --->$

In each of these two groups and for all particle pairs (a, b) , (a, c) , (b, c) , the combinations of signs $(++)$, $(+-)$, $(-+)$, and $(--)$ appear exactly once and, for the remaining particle, the signs $+$ and $-$ appear twice. Therefore for $\langle \psi_j | I \otimes I \otimes \sigma_k | \psi_j \rangle$ with $j \in \{1, 2, 3, 4\}$,

$$\begin{aligned}
\langle \psi_j | I \otimes I \otimes \sigma_k | \psi_j \rangle &= \sum_{l,m \in \{1,5,6,7\}} r_{j,l} r_{j,m} \langle \phi_l | I \otimes I \otimes \sigma_k | \phi_m \rangle = \sum_{l \in \{1,5,6,7\}} r_{j,l}^2 \langle \phi_l | I \otimes I \otimes \sigma_k | \phi_l \rangle \\
&= 2r^2 (\langle + | \sigma_k | + \rangle + \langle - | \sigma_k | - \rangle) = \frac{1}{2} \text{Tr}(\sigma_k) = 0
\end{aligned}
\tag{C.6}$$

and analogously for the other cases. For every ψ_j , the state of each isolated particle equals

$$\rho_0 = \frac{1}{2} I \quad (\text{C.7})$$

Or, with respect to the reduced observable algebra $\{A \otimes I \otimes I\} \cup \{I \otimes B \otimes I\} \cup \{I \otimes I \otimes C\}$, each state ψ_j is equivalent to the product state $\rho_0 \otimes \rho_0 \otimes \rho_0$.

Appendix D

The time evolution of S_1 and S_2

The task consists in evaluating the integrated Heisenberg equations (4.15) for the spin variables S_1 and S_2 . The time interval $0 \leq t < T$ is considered where the Hamilton operator \check{H} is constant. It is advantageous to choose \check{H} in the form

$$\check{H} = \frac{\mathbf{P}(0)^2}{2m} - \gamma X_3(0) S_3(0)$$

because $S_1(0)$ and $S_2(0)$ commute with $\mathbf{P}(0)$ and $X_3(0)$. The components $P_1(0)$ and $P_2(0)$ commute with all other observables occurring and $\check{U}(t)$ factorizes into a product

$$e^{-\frac{i}{\hbar} \frac{1}{2m} (P_1^2(0) + P_2^2(0))t} \check{U}_{\text{red}}(t)$$

In the time evolution equation (4.15) for the observables, the first factor is cancelled by the corresponding term in $\check{U}^{-1}(t) = \check{U}(-t)$. Only the part $\check{U}_{\text{red}}(t)$ is relevant for the time evolution of $S_1(t)$ and $S_2(t)$. \check{U}_{red} is given by

$$\check{U}_{\text{red}}(t) = e^{it(A+B\sigma_3)} \quad (\text{D.1})$$

with the substitutions

$$A = -\frac{1}{\hbar} \frac{P_3(0)^2}{2m} \quad (\text{D.2})$$

$$B = \frac{\gamma}{2} X_3(0) \quad (\text{D.3})$$

$$\boldsymbol{\sigma}(t) = \frac{2}{\hbar} \mathbf{S}(t) \quad (\text{D.4})$$

$\boldsymbol{\sigma}(0) \equiv \boldsymbol{\sigma}$ are the Pauli spin matrices (3.21); $\boldsymbol{\sigma}$ commutes with A and B and it is $\sigma_3^{2n} = I$ and $\sigma_3^{2n+1} = \sigma_3$.

Because σ_3 commutes with A and B , the power series expansion of the exponential function reveals that the expressions

$$\frac{1}{2} \left(e^{it(A+B\sigma_3)} \pm e^{it(A-B\sigma_3)} \right)$$

contain only even (plus sign) or odd (minus sign) powers of σ_3 . It follows that

$$\check{U}_{\text{red}}(t) = E^+(t) + E^-(t) \sigma_3 \quad (\text{D.5})$$

with

$$E^\pm(t) = \frac{1}{2} (U^+(t) \pm U^-(t)) \quad (\text{D.6})$$

where the unitary operators $U^\pm(t)$ are given by

$$U^\pm(t) = e^{it(A \pm B)} \quad (\text{D.7})$$

Set now

$$\sigma_{\pm}(t) \equiv \sigma_1(t) \pm i\sigma_2(t) \quad (\text{D.8})$$

An easy calculation yields, with $U_P(t)$ given by Eq. (4.7), that

$$U_P^{-1}(t) \sigma_{\pm}(0) U_P(t) = e^{\pm i\omega t} \sigma_{\pm}(0) \quad (\text{D.9})$$

indicating the spin precession due to the constant part B_0 of the external magnetic field. It is, further,

$$\sigma_{\pm}\sigma_3 = \mp\sigma_{\pm}, \quad \sigma_3\sigma_{\pm} = \pm\sigma_{\pm}, \quad \sigma_3\sigma_{\pm}\sigma_3 = -\sigma_{\pm}$$

Therefore

$$\begin{aligned} \sigma_{\pm}(t) &= \check{U}_{\text{red}}(-t) U_P(-t) \sigma_{\pm}(0) U_P(t) \check{U}_{\text{red}}(t) \\ &= e^{\pm i\omega t} \check{U}_{\text{red}}(-t) \sigma_{\pm}(0) \check{U}_{\text{red}}(t) \\ &= e^{\pm i\omega t} \{ (E^+(-t) + E^-(-t) \sigma_3) \sigma_{\pm}(0) (E^+(t) + E^-(t) \sigma_3) \} \\ &= e^{\pm i\omega t} \{ E^+(-t) E^+(t) - E^-(-t) E^-(t) \pm (E^-(-t) E^+(t) - E^+(-t) E^-(t)) \} \sigma_{\pm}(0) \\ &= \frac{1}{2} e^{\pm i\omega t} \{ U^+(-t) U^-(t) + U^-(-t) U^+(t) \pm (U^+(-t) U^-(t) - U^-(-t) U^+(t)) \} \sigma_{\pm}(0) \end{aligned} \quad (\text{D.10})$$

Or, with the unitary operator

$$V(t) \equiv U^+(-t) U^-(t) = e^{-it(A+B)} e^{it(A-B)} \quad (\text{D.11})$$

$V^\dagger V = I$ one obtains

$$\sigma_+(t) = V(t) e^{i\omega t} \sigma_+(0) \quad (\text{D.12})$$

$$\sigma_-(t) = V^\dagger(t) e^{-i\omega t} \sigma_-(0) \quad (\text{D.13})$$

With $\sigma_+\sigma_- = \sigma_1^2 + \sigma_2^2 + 2\sigma_3$ and $\sigma_3(t) = \sigma_3(0)$, it follows that

$$(\sigma_1^2 + \sigma_2^2)(t) = (\sigma_1^2 + \sigma_2^2)(0) = 2 \quad (\text{D.14})$$

In terms of S_1 and S_2 , it is

$$\begin{pmatrix} S_1 \\ S_2 \end{pmatrix}(t) = \begin{pmatrix} \text{Co}(t) & -\text{Si}(t) \\ \text{Si}(t) & \text{Co}(t) \end{pmatrix} \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}(0) \quad (\text{D.15})$$

where

$$\text{Co}(t) = \frac{1}{2} (V(t) + V^\dagger(t)) \quad (\text{D.16})$$

$$\text{Si}(t) = \frac{1}{2i} (V(t) - V^\dagger(t)) \quad (\text{D.17})$$

One easily verifies that $V = \text{Co} + i\text{Si}$, $\text{Co}^2 + \text{Si}^2 = I$, $[\text{Co}, \text{Si}] = 0$, and, from $S_1(t) S_2(t) = -S_2(t) S_1(t)$ and Eq. (D.14) that $S_1^2(t) = S_2^2(t) = I$ for all times t . The operators V , Co , and Si only act on the 3-component of configuration space.

To summarize: a spin precession due to the gradient of the magnetic field is superposed to the spin precession due to the homogeneous part B_0 of the magnetic field.

Appendix E

X and P are irreducible on $\mathbb{L}^2(\mathbb{R}^3)$

Let A denote a self-adjoint operator on the Hilbertspace $\mathbb{L}^2(\mathbb{R}^3)$, the vector space of square integrable, complex functions over \mathbb{R}^3 , \mathbf{X} the position operator, $(\mathbf{X}\psi)(\mathbf{x}) = \mathbf{x}\psi(\mathbf{x})$, \mathbf{P} the momentum operator, $(\mathbf{P}\psi)(\mathbf{x}) = \frac{\hbar}{i}(\nabla\psi)(\mathbf{x})$. The following statements are then valid

1. If $[A, \mathbf{X}] = \mathbf{0}$ then $A = A(\mathbf{X})$.
2. If $[A, \mathbf{P}] = \mathbf{0}$ then $A = A(\mathbf{P})$.
3. If $[A, \mathbf{X}] = \mathbf{0}$ and $[A, \mathbf{P}] = \mathbf{0}$ then $A = aI$ with some real constant a . This means that \mathbf{X} and \mathbf{P} form an irreducible set of operators on $\mathbb{L}^2(\mathbb{R}^3)$.

Proof In coordinate representation, A is an integral operator

$$(A\psi)(\mathbf{x}) = \int_{\mathbb{R}^3} d^3x' a(\mathbf{x}, \mathbf{x}') \psi(\mathbf{x}') \quad (\text{E.1})$$

The kernel $a(\mathbf{x}, \mathbf{x}') = a(\mathbf{x}', \mathbf{x})^*$ may include δ -distributions $\delta(x_j - x'_j)$ and derivatives of such δ -distributions representing multiplication and derivation operators [40]. Let $f(\mathbf{x})$ denote a real function that can be approximated by polynomials in x_1, x_2, x_3 . It is, with $[A, \mathbf{X}] = \mathbf{0}$, also $[A, f(\mathbf{X})] = 0$ or, in coordinate representation,

$$f(\mathbf{x}) \int_{\mathbb{R}^3} d^3x' a(\mathbf{x}, \mathbf{x}') \psi(\mathbf{x}') = \int_{\mathbb{R}^3} d^3x' a(\mathbf{x}, \mathbf{x}') f(\mathbf{x}') \psi(\mathbf{x}') \quad (\text{E.2})$$

for any wave function $\psi \in \mathbb{L}^2(\mathbb{R}^3)$ and any function f for which the integrals exist. Set now f equal to

$$f_\sigma(\mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^3 \exp \left(-\frac{\|\mathbf{x} - \mathbf{x}_0\|^2}{2\sigma^2} \right), \quad \sigma > 0 \quad (\text{E.3})$$

where $\lim_{\sigma \downarrow 0} f_\sigma(\mathbf{x}) = \delta^3(\mathbf{x} - \mathbf{x}_0)$. Inserting f_σ into Eq. (E.2) and passing to the limit $\sigma = 0$ yields

$$\delta^3(\mathbf{x} - \mathbf{x}_0) \int_{\mathbb{R}^3} d^3x' a(\mathbf{x}, \mathbf{x}') \psi(\mathbf{x}') = a(\mathbf{x}, \mathbf{x}_0) \psi(\mathbf{x}_0) \quad (\text{E.4})$$

and thus $a(\mathbf{x}, \mathbf{x}') = \delta^3(\mathbf{x} - \mathbf{x}') a(\mathbf{x})$. It is then

$$(A\psi)(\mathbf{x}) = a(\mathbf{x}) \int_{\mathbb{R}^3} d^3x' \delta^3(\mathbf{x} - \mathbf{x}') \psi(\mathbf{x}') = a(\mathbf{x}) \psi(\mathbf{x}) \quad (\text{E.5})$$

This proves point 1. If now, in addition, $[A, \mathbf{P}] = \mathbf{0}$ then $a\nabla\psi = \nabla(a\psi)$ implying $\nabla a = \mathbf{0}$ or $a = \text{constant}$. This proves point 3. Point 2 is demonstrated in momentum representation similarly to the proof of point 1.

QED

Bibliography

- [1] Schilpp, P. A. *Albert Einstein, Philosopher-Scientist*; The Library of Living Philosophers, Inc.: Evanston, Illinois, 1949.
- [2] Tschudi, H. R. *Helvetica Physica Acta* **1987**, 60, 363 – 383.
- [3] Mermin, N. D. *Rep. Prog. Phys.* **2019**, 82, 1 – 16.
- [4] Born, M. *Zeitschrift für Physik* **1926**, 37, 863 – 867.
- [5] Ballentine, L. E. *Rev. Mod. Phys.* **1970**, 42(4), 358 – 381.
- [6] Schrödinger, E. *Die Naturwissenschaften* **1935**, 23(48), 807 – 812.
- [7] Schrödinger, E. *Die Naturwissenschaften* **1935**, 23(49), 823 – 828.
- [8] Schrödinger, E. *Die Naturwissenschaften* **1935**, 23(50), 844 – 849.
- [9] Popper, K. R. *Quantum Mechanics without "The Observer"*. In: M. Bunge ed., *Quantum Theory and Reality*; Springer-Verlag: Berlin, 1967.
- [10] Blokhintsev, D. I. *The Philosophy of Quantum Mechanics*; Reidal: Dortrecht, Holland, 1968.
- [11] Jammer, M. *The Philosophy of Quantum Mechanics — The Interpretations of Quantum Mechanics in Historical Perspective*; John Wiley & Sons, Inc.: New York, London, Sydney, Toronto, 1974.
- [12] Kochen, S.; Specker, E. P. *Journal of Mathematics and Mechanics* **1967**, 17, 59 – 87.
- [13] Bohm, D. *Quantum Theory*; Prentice Hall: Englewood Cliffs (N.J.), 1951.
- [14] Bohm, D.; Aharonov, Y. *Physical Review* **1957**, 108, 1070 – 1076.
- [15] Einstein, A.; Podolsky, B.; Rosen, N. *Physical Review* **1935**, 47, 777 – 780.
- [16] Mermin, N. D. *Rev. Mod. Phys.* **1993**, 65(3), 803 – 815.
- [17] Bell, J. S. *Physics* **1964**, 1, 195 – 200.
- [18] Popper, K. R. *Logik der Forschung*; J. C. B. Mohr (Paul Siebeck): Tübingen, fourth ed., 1971.
- [19] Scharf, R.; Middelmann, T. *PTB-Mitteilungen* **2016**, 126(2), 5 – 15.
- [20] Gibney, E. *Nature* **2017**, 550, 312 – 313.
- [21] Mittelstraß, J., Ed. *Enzyklopädie Philosophie und Wissenschaftstheorie, Band 4*; J.B. Metzler: Stuttgart, Weimar, 1996.
- [22] von Mises, R. *Probability, Statistics, and Truth*; Macmillan Publishers: New York, 1939.

- [23] Popper, K. R. *The propensity interpretation of the calculus of probability, and the quantum theory*. In: S. Körner ed., *Observation and Interpretation*; Butterwoths Scientific Publications: London, 1957.
- [24] Mackey, G. W. *The Mathematical Foundation of Quantum Mechanics*; W. A. Benjamin: New York, 1963.
- [25] Ballentine, L. E. *Quantum mechanics A Modern Development*; World Scientific: New Jersey, second ed., 2015.
- [26] Mermin, N. D.; Schack, R. *Foundations of Physics* **2018**, 48, 1007 – 1020.
- [27] Bell, J. S. *Foundations of Physics* **1982**, 12, 989 – 999.
- [28] Belinfante, F. J. *A Survey of Hidden-Variables Theories*; Pergamon Press: Oxford, New York, Toronto, Sidney, Braunschweig, 1973.
- [29] von Neumann, J. *Mathematical foundations of quantum mechanics*; Princeton University Press: Princeton, N.J., 1955.
- [30] Bell, J. S. *Rev. Mod. Phys.* **1966**, 38, 447 – 452.
- [31] Hermann, G. *Abhandlungen der Fries'schen Schule* **1935**, 6, 75 – 152.
- [32] Bohm, D. *Physical Review* **1952**, 85, 166 – 179.
- [33] Bohm, D. *Physical Review* **1952**, 85, 180 – 193.
- [34] Schiff, L. I. *Quantum Mechanics*; McGraw-Hill Book Company, Inc.: New York, Toronto, London, second ed., 1955.
- [35] Gleason, A. M. *J. of Math. and Mech.* **1957**, 6, 885 – 893.
- [36] Jost, R. *Measures on the Finite Dimensional Subspaces of a Hilbert Space: Remarks to a Theorem by A. M. Gleason*. In: *Studies in Mathematical Physics, Essais in Honor of Valentine Bargmann*; Princeton University Press: Princeton, 1976.
- [37] Bohr, N. *Physical Review* **1935**, 48, 696 – 702.
- [38] Gerlach, W.; Stern, O. *Zeit. Phys.* **1922**, 4, 349 – 352.
- [39] Platt, D. E. *Am. J. Phys.* **1990**, 60(4), 306 – 308.
- [40] Lighthill, M. J. *Introduction to Fourier Analysis and Generalised Functions*; Cambridge University Press: Cambridge, 1958.
- [41] Fuchs, C. A.; Schack, R. *Ref. Mod. Phys.* **2013**, 85(4), 1693 – 1715.
- [42] Adler, M. *Natural Theology, Chance, and God*, in: *The Great Ideas Today*; Encyclopedia Britannica: Chicago, 1992.
- [43] Born, M. *The Born-Einstein letters*; Macmillan: London, 1971.
- [44] Bell, J. S. *Speakable and unspeakable in quantum mechanics*; Cambridge University Press: Cambridge, New York, Port Chester, Melbourne, Sidney, 1987.
- [45] Clauser, J. F.; Horne, M. A.; Shimony, A.; Holt, R. A. *Physical Review Letters* **1969**, 23(15), 880 – 884.

- [46] Aspect, A.; Dalibard, J.; Roger, G. *Phys. Rev. Lett.* **1982**, *49*, 1804 – 1807.
- [47] Salart, D.; Baas, A.; Branciard, C.; Gisin, N.; Zbinden, H. *Nature* **2008**, *454*, 861 – 864.
- [48] Hensen, B.; et al.. *Nature* **2015**, *526*, 682 – 686.
- [49] Greenberger, D. M.; Horne, M. A.; Shimony, A.; Zeilinger, A. *Am. J. Phys.* **1990**, *58*(12), 1131 – 1143.
- [50] Stapp, H. P. *Physical Review D* **1971**, *3*(6), 1303 – 1320.
- [51] A. Einstein, H. B.; Born, M. *Albert Einstein, Hedwig und Max Born: Briefwechsel 1916 - 1955*; Rowohlt Taschenbuchverlag: Reinbek bei Hamburg, 1972.
- [52] Born, I. *The Born-Einstein letters*; Walker and Company: New York, 1971.
- [53] Eberhard, P. H. *Nuovo Cimento B* **1977**, *38*, 75 – 80.
- [54] Stapp, H. P. *Found. Phys.* **1988**, *18*(4), 427 – 448.
- [55] Jordan, T. F. *American Journal of Physics* **1975**, *43*, 1089 – 1093.
- [56] Levy-Leblond, J.-M. *J. Math. Phys.* **1963**, *4*(6), 776 – 788.
- [57] Goldstein, H. *Classical Mechanics*; Addison-Wesely Publishing Company, Inc.: Reading, Mass. and London, 1959.
- [58] Hall, B. C. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*; Springer Verlag: New York, 2003.
- [59] Bargmann, V. *Journal of Mathematical Physics* **1964**, *5*(7), 862 – 868.
- [60] Riesz, F.; Nagy, B. S. *Vorlesungen über Funktionalanalysis*; VEB Deutscher Verlag der Wissenschaften: Berlin, 1968.
- [61] Bargmann, V. *Annals of Mathematics* **1954**, *59*(1), 1 – 46.
- [62] Gnedenko, B. W. *Lehrbuch der Wahrscheinlichkeitstheorie*; Verlag Harri Deutsch: Thun, Frankfurt am Main, 1997.
- [63] Choquet-Bruhat, Y.; Witt-Morette, C. D.; Dillard-Bleick, M. *Analysis, Manifolds and Physics*; North-Holland Publishing Company: Amsterdam, New York, Oxford, 1977.