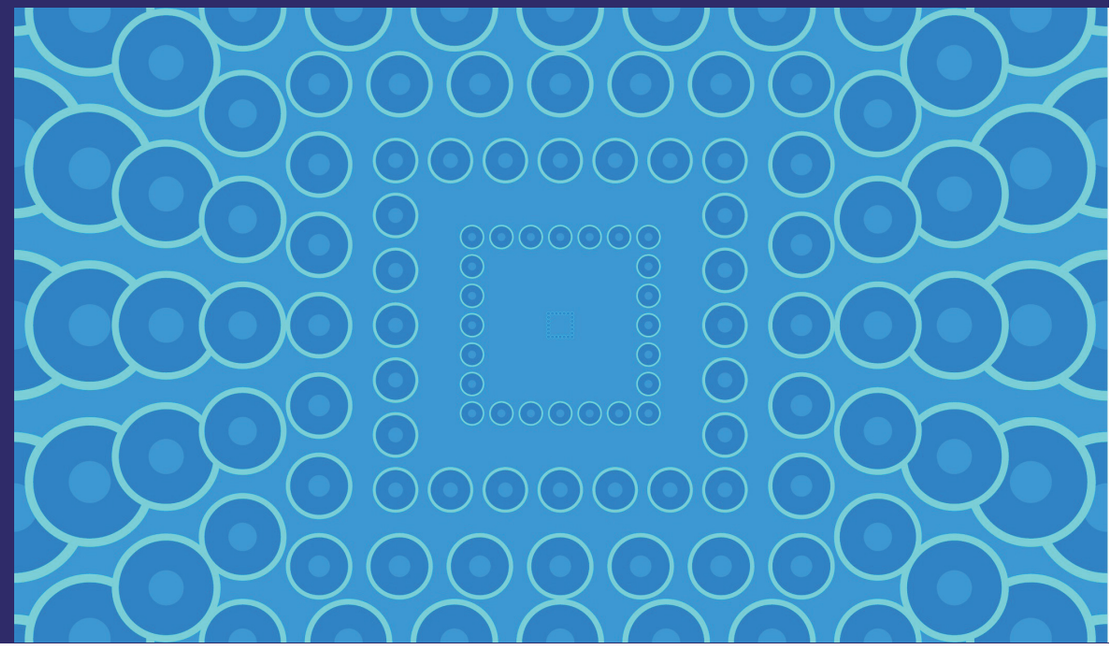
A deductive physical theory should in principle be a pure mathematical theory together with an identification of certain quantities/concepts ("observables") in the theory and corresponding observable entities in the real world. This identification – the "interpretation" of the theory – should be unproblematic, both for the theoretician and the experimentalist. A general basis for a deductive physical theory, comprising both classical and quantum physics in a unified way, is proposed. The theory is based on successive confidence estimates on quantum-mechanical wave functions corresponding to space-localizations of particles. This allows a direct and simple way of describing both macroscopic and microscopic phenomena by means of the same basic concepts. Central in the axiomatics of the outlined theory is a concept called equiangular sequences of projection operators. It describes a successive sequence of "collapses of the wave function". The proposed theory gives a basis for a general theory of irreversible processes based directly on quantum mechanics. It gives an alternative definition of entropy and an alternative derivation of entropy increase in irreversible processes.

Tomas Blomberg was born in 1940. He studied and has been teaching mathematics and theoretical physics at the University of Stockholm. He has also been working with IT-systems as programmer and systems analyst. His licentiate of philosophy dissertation was about bound states in S-matrix theory.

978-3-659-68549-1

Deductive Theoretical Physics

Blomberg



Tomas Blomberg

Principles of Deductive Theoretical Physics

A Proposal for a General Theory Based on Successive Confidence Estimates on Quantum-Mechanical Wave Functions

LAP LAMBERT Academic Publishing

Tomas Blomberg

Principles of Deductive Theoretical Physics

Tomas Blomberg

Principles of Deductive Theoretical Physics

A Proposal for a General Theory Based on Successive Confidence Estimates on Quantum-Mechanical Wave Functions

LAP LAMBERT Academic Publishing

Impressum / Imprint
Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
Alle in diesem Buch genannten Marken und Produktnamen unterliegen warenzeichen-, marken- oder patentrechtlichem Schutz bzw. sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Inhaber. Die Wiedergabe von Marken, Produktnamen, Gebrauchsnamen, Handelsnamen, Warenbezeichnungen u.s.w. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutzgesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.
Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Coverbild / Cover image: www.ingimage.com

Verlag / Publisher:
LAP LAMBERT Academic Publishing
ist ein Imprint der / is a trademark of
OmniScriptum GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland / Germany
Email: info@lap-publishing.com

Herstellung: siehe letzte Seite /
Printed at: see last page
ISBN: 978-3-659-68549-1

Copyright © 2015 OmniScriptum GmbH & Co. KG
Alle Rechte vorbehalten. / All rights reserved. Saarbrücken 2015

Layout: TEXNOT Anders Källström, Vänge 2014

Contents

Preface to the 2014 Publication

This publication of mathematical speculations and writings made during the 1970:s and early 1980:s contains the general principles of a proposed deductive approach to theoretical physics and an outline of a mathematical theory. A rigorous and self-contained exposition of the three most basic concepts of the theory, containing the definitions and theorems quoted in Part II, is given in *Successive Confidence Estimates on Solutions to the Many-Particle Schrödinger Equation. Basic Concepts* which is included in this book as Part III. The reference lists are incomplete in the sense that I have not been in a position to do ordinary studies of and make ordinary references to other existing works related to the present work.

* * *

First of all I wish to thank Lars E. Henriksson. It was Lars who, together with and supported by Peje Löfgren, took the initiative to publish these old mathematical speculations and writings of mine. Lars performed and managed all that administration work that was necessary. Without Lars' great, due to my own shortcomings necessary, patience, this publication would not have been realized.

I wish to thank Peje Löfgren for many years of profound mathematical discussions. Many of these have bearing on my own work. Let me just give one example. A central concept in the present theory is that of finite approximations. I owe much to Peje's own work on this subject.

I wish to thank Jan Pilotti. Jan has read Parts I and II of the manuscript critically and has pointed out several important corrections and elucidations.

I also wish to thank Anders Källström for help with the mathematical typesetting of the manuscript.

To all I wish to express my great gratitude.

Stockholm August 2014

Tomas Blomberg

Preface

The principal object of the following theory is to treat and develop theoretical physics as a deductive science.

Common theoretical physics, although deductive in certain parts or steps, generally displays an apparent lack of deductiveness. There are excellent examples of completely deductive theories such as e.g. Newtonian point mechanics, but they appear as small isolated and widely separated islands when considered in relation to our complete physical knowledge. Practically every theoretical discussion frequently introduces extra, often implicit, assumptions depending on the specific problem under concern, without deriving the validity of these extra assumptions from basic postulates. In some cases these extra assumptions are more or less obvious or natural (although they might be difficult to prove). In some cases they are rather doubtful. However, it is a remarkable fact that derivations of such extra assumptions are often missing in the literature even in cases where it would be a straightforward task to work out a rigorous proof. For example, we will not take for granted, but derive rigorously, the fact that light propagates along straight lines and with the "velocity of light" c. A rigorous and complete formulation and proof of this statement needs a thorough consideration of confidence estimates on wave packets. Although quite nontrivial, it is obtained by a rather elementary exercise in Fourier transform theory. It ought to be found in any thorough theoretical discussion on light.

A similar situation appears when one theory is a special case of another more general theory. There are seldom any attempts to derive in a more definite way the validity of the basic principles of the special theory from the more general theory. The task of working out such derivations, connecting different theories, is of a central interest in the following theory. Let us note that such a derivation is not only a matter of formality. It is in fact intimately related to the problem of finding the exact conditions under which the special theory is applicable and such conditions have an immediate physical significance. Often the special theory appears as an approximation of the more general theory and one also wants to know the degree of accuracy of this approximation, which also has an obvious physical significance. The lack of derivations discussed means that important physical questions are left outside the theoretical treatment.

A consequence of this general lack of deductiveness is also that it makes it practically impossible to apply the mathematical method effectively. The central position of proofs in the mathematical method is intimately connected to the function of mathematics as an "art of computation". Computations are in fact examples of the deductive method. There is no principal difference between a proof, which in a deductive way leads to a qualitative prediction and a computation, which in a deductive way leads to a quantitative prediction. Thus, we see again that deductiveness is not only a question of formality but has a practical importance. It is only when we have complete deductiveness that we can fully exploit the power of the theoretical method.

Parallel to the lack of rigorous proofs there is in common theoretical physics a pervading lack of precise definitions of important concepts used in the theories.

This indicates that the apparent lack of deductiveness is connected to general conceptual problems of theoretical physics. In order to develop a general deductive theoretical physics we have to solve the following two problems:

1) Establish a general conceptual basis for deductive theoretical physics.

2) Establish a formulation of quantum mechanics with general and unproblematic applicability to physical problems.

These two problems are closely connected since the solution of one of them presupposes a solution of the other. Classical theoretical physics is composed of a set of disconnected theories, mechanics, the electromagnetic theory, thermodynamics, etc., and the only theory which offers the possibility of a general theory, encompassing these classical theories, is quantum mechanics. On the other hand we claim that a solution of the controversial conceptual problems of quantum mechanics presupposes a general conceptual deductive framework. Our proposal for solving these two problems is the embedding of the Schrödinger equation formalism in a general "physico-logical" structure which we shall call "stochastic event structure". This structure provides basic concepts for direct descriptions both of classical and quantum phenomena in a unified and objectivistic way. The Schrödinger equation then complements this descriptive structure with a general dynamics generalizing and encompassing the classical theories.

The theory proposed in part II below is at the same time a mathematical theory and a physical theory. As a physical theory it has of course a phenomenological character. Thus, the mathematical theory below is suggested by speculations on quantum mechanics, which in turn has its origin in the physical phenomenology, and the purpose of the theory is to describe the physical reality. As any physical theory it then ultimately stands or falls depending on its further success in describing, analyzing and predicting physical phenomena. We can thus distinguish three different steps in the development of a physical theory.

1) Axiomatize the theory. This means that we establish the basic mathematical concepts which are to describe the basic physical concepts of the theory and establish in a mathematical form the basic laws connecting these concepts. The axiomatization thus results in a specific mathematical theory.

2) Develop this mathematical theory.

3) Compare results obtained in the mathematical theory with the physical reality.

It is important for the deductiveness of the theory that we are in a position where steps 1 and 3 present no problems and we can deal mathematically with step 2 in a free way, undisturbed by unformalized physical questions. We claim that the theory proposed below meets this demand.

The purpose of the following exposition is to give the general principles of the theory i.e. establish step 1) above. For the mathematical development of the theory, step 2) above, we refer to the self-contained, purely mathematical exposition given in Part III.

Introduction

In the following we shall give the conceptual foundations of a mathematical formulation of quantum mechanics based on confidence estimates instead of mean values and density operators, used in conventional quantum mechanics and quantum statistical mechanics.

A mathematical formulation of this theory – an approximation theory of L^2-functions of several variables, applied to sequences of interrelated subspaces of solutions to the many-particle Schrödinger equation – is given in Part III. The exposition in Parts I and II can be considered as a physical motivation and a mathematical outline of this theory. For proofs of theorems cited below and for further development of the mathematical technique needed in this theory, we refer to Part III.

The main purpose of this exposition is to describe the basic principles of the theory, in the following called "the confidence theory".

A second and complementary purpose is to discuss the difference between the theory and the conventional formulation of quantum theory and statistical mechanics. Although we reject the Copenhagen interpretation and the formalism built on it, we shall, due to its present overwhelming position, recapitulate it in chapter 2 and criticize it in chapter 3. Conventional statistical mechanics is discussed in chapter 4 (4.1).

The two main purposes of the confidence theory are the following:

1) To modify the conventional theory to give an unambiguous, deductive theory.

2) To propose a general theoretical basis for the treatment of, generally non-stationary, macroscopic systems.

The most important of these is the second. We consider the first purpose, although it has its own conceptual interest, mainly as a means to achieve the (more pragmatical) second purpose.

Since the exposition in Parts I and II is mainly conceptual, it is important to emphasize the mathematical technical character of the confidence theory. The confidence theory is not in first hand a philosophical-logical discussion on the subject "quantum mechanics without the observer". It consists of a mathematical technique, the above mentioned approximation theory, whose main motivation is the second purpose stated above.

Chapters 1–4 mainly have the purpose of supporting the heuristic derivation of the confidence theory given in chapter 5. An axiomatic exposition of the theory is given in chapters 6 and 7. The systematic exposition of the theory given in chapters 6–7 is "self-contained" in the sense that it does not formally or logically presuppose chapters 1–5.

By a confidence estimate we shall mean an estimate of the form

$$\int_R |\psi(x,t)|^2 \, dx \geq 1 - \varepsilon$$

where $\psi(x,t)$ is a normalized wave function of the time variable t and the configuration space variables x for a set of elementary particles. ε should be a very

small positive number and $1-\varepsilon$ is called the "confidence level". According to the statistical interpretation of the wave function, the above estimate means that the particles are, with practical certainty (probability $\geq 1-\varepsilon$) confined to the region R at time t. A reason for using confidence estimates rather than exact localization statements comes from the fact that a wave function $\psi(x,t)$, localized exactly at time t_1 to a region R_1 (i.e. vanishing outside R_1), will, according to the Schrödinger equation, generally spread out in space so that it cannot be localized exactly to any finite region at another time t_2. On the other hand, we can under certain assumptions obtain estimates

$$\int_{R_2} |\psi(x,t_2)|^2\, dx \geq 1-\varepsilon$$

at time t_2 with finite region R_2 and very small ε.

The most general basic question of a physical theory is the study of the macroscopic distribution of matter in space at different instants of time. Even if we are studying an experiment observing a single elementary particle, the situation can and should ultimately be described by macroscopic, directly observable, quantities. Thus, the task of establishing an interpretation of the quantum-mechanical wave functions and the task of describing and understanding macroscopic processes from an underlying atomistic point of view are closely connected.

The macroscopic distribution of matter in space can be instantaneously described by the localization of the configuration space variables for the constituting elementary particles to suitable intervals or regions, which are small from a macroscopic, but large from a microscopic point of view. For a quantum-mechanical wave function this means that it has (at a given instant of time) its support in that region i.e. vanishes outside the region. The set of all such wave functions constitute a (closed) subspace of the Hilbert space of wave functions. We shall describe the macroscopic distribution of matter by means of sequences of such subspaces (or equivalently by their corresponding projection operators).

By the preceeding argument, the use of confidence estimates will allow us to describe, with a sufficiently high degree of accuracy, the macroscopic behaviour of systems by consequently using only such subspaces (projection operators). The confidence theory is a theory developed consequently along these lines.

The confidence theory thus gives, in a direct way, a connection between wave functions and macroscopic quantities and therefore presents an alternative to the ensemble (density operator) methods of statistical mechanics. We shall criticize the conventional (classical and quantum) statistical mechanics in chapter 4 and propose an alternative theory based directly on phase-space region localizations. The confidence theory is a quantum-mechanical generalization of this phase-space region theory.

Parts I and II is an attempt to describe the theory (and its relation to the conventional theory) in qualitative, intuitive, verbal terms. Due to the structural and conceptual complexity of the subject, it consists of a network of different aspects and more or less precise arguments. A completely rigorous discussion can of course only be given in an axiomatized mathematical exposition (see Part III).

Part I

Heuristic Derivation

Chapter 1

Interpretations of the Wave Function

In all classical physical theories, the concept of state is fundamental. The state of the system under concern is described by basic physical quantities of the theory (e.g. positions of particles, components of fields as functions of space coordinates etc.) and the change of state in time is governed by the fundamental dynamical equations of the theory. This means that we have a simple and obvious one-to-one correspondence between basic quantities in the theory and tangible, observable quantities in the physical reality. Such a correspondence is at hand in all classical theories – classical mechanics, continuum mechanics, the electromagnetic field theory, classical macroscopic thermodynamics and also in the special and the general theory of relativity. The existence of such a correspondence means that there is no problem with the interpretation of the theory.

The situation is quite different in quantum mechanics. Here one is faced with a situation of having a differential equation (the Schrödinger equation) for a certain wave function, but the interpretation is no longer trivial. It soon turned out that there is no one-to-one correspondence between a wave function and the consecutive states of the system. The spreading of the wave function in scattering processes, together with the indivisible nature of elementary particles, forced a statistical interpretation of the wave function and the considering of stochastic transitions between different wave functions.

Thus, one very soon found the correct dynamical equation of quantum mechanics, namely the Schrödinger equation, but its interpretation raised considerable difficulties. It is quite obvious, that the Schrödinger equation is correct in some sense despite the difficulties of interpretation. It is the basic dynamical equation both in the conventional and in the present theory.

In the following, we shall discuss three different interpretations of the wave function:

Interpretation 1: "The primitive statistical interpretation".

This is the original interpretation of quantum theory. It consists primarily of Born's interpretation of $|\psi(x,t)|^2$ for given time t as a probability density for the configuration space variable x. Together with this, one has some kind of corre-

spondence between physical quantities and operators. One expression for this correspondence is that the possible values of a physical quantity (e.g. the energy) are given by the eigenvalues of the corresponding operator. From the configuration space probability distribution above, one can derive (or at least make plausible) other distributions, e.g. that $|\hat{\psi}(k,t)|^2$, where $\hat{\psi}$ denotes the Fourier transform, is in some sense a probability density for the wave-number k, connected with the momentum variable p by means of the de' Broglie relation $p = \hbar k$. (See e.g. Bohm (2)).

Interpretation 2: The Copenhagen Interpretation

This interpretation, described in chapter 2, constitutes the "conventional formulation" of quantum mechanics.

Interpretation 3: The Interpretation of the Confidence Theory

In this theory, the interpretation is established by means of the concept of "stochastic event structure".

Interpretation 1 can easily be criticized for being vague and incomplete. Obviously it does not constitute a complete satisfactory theory. However, one important positive credit to it should be pointed out. It is sufficient for many practical purposes (e.g. calculation of bound state energies or scattering cross-sections) and then gives correct results. We shall now point out two basic problems of interpretation of wave functions.

The first we shall here call "the state-observable dualism". In the mathematical formalism occurs both elements, namely wave functions ψ:s, and corresponding Hilbert space vectors "$\boldsymbol{\Psi}$:s", (see section 7.5 below), which have the "character of states" and elements, operators, x, $p_x = i\hbar\partial_x$ etc., corresponding to real-valued physical quantities ("observables"). In the classical theories, the states are just described by real-valued physical quantities. In quantum mechanics, the two concepts has split into separate mathematical concepts. This is really a dualism since states and observables should describe the same physical reality. It is obviously a conflict. Both cannot stand in a one-to-one relation to reality since there is no obvious general one-to-one relation between themselves, i.e. between the "$\boldsymbol{\Psi}$:s" ("states") and the operators ("observables"), (although they are closely related in different ways). Dualisms are not preferable from a reductionist's point of view. Here it even means a manifest inconsistency which has to be resolved.

For the second we shall here use the term "the arbitrariness problem". It consists of three closely connected parts – the "arbitrariness of states", the "arbitrariness of observables" and the "arbitrariness of collapses (transitions)". The first two mean simply that there is no restriction in the formalism (theory) on which momentary wave functions are allowed or which operators as describing observables are allowed. Perhaps the most striking and well-known consequence of the arbitrariness of states is the Schrödinger cat paradox (see below). The crucial point in this paradox is that a state which is a superposition of a living and a dead cat is not forbidden by the theory.

The occurrence of stochastic quantum transitions implies that the change of state of a system from one time to another is not completely described by the

time-dependent Schrödinger equation. One also has to consider so called "collapses" of the wave functions i.e. we have one Hilbert space vector Ψ_1, before the transition and another Ψ_2 after. The transition probability is then given by the square-modulus of the scalar product

$$p = |\langle \Psi_1, \Psi_2 \rangle|^2$$

(where Ψ_1 and Ψ_2 are properly normalized and taken at the same time). The arbitrariness problem can then also be expressed as the arbitrariness of collapses". There is no rule that tells us which collapses can and will occur.

In case of the cat paradox one has a thought situation with an initial state (a living cat together with a stochastic poisoning apparatus) which, under a period of isolation evolves, according to the time-dependent Schrödinger equation, into a state Ψ which is a superposition of a living (Ψ_1) and a dead (Ψ_2) cat. Eventually and after the time of isolation it appears whether the cat is living or dead so the wave function has collapsed into either Ψ_1 or Ψ_2. The arbitrariness then appears as the question of when and why the collapse occurs.

The main point of the critique, given below, of the Copenhagen interpretation is that it does not solve the "state-observable duality" and the "arbitrariness problem". These problems are in the following considered as "key problems" and will be used in chapter 5 as "hints" in a heuristic motivation of the confidence theory.

Chapter 2

Recapitulation of the Conventional Quantum-mechanical Theory

The incompleteness and vagueness of interpretation 1 has lead to the development of the Copenhagen interpretation, which replaces the primitive statistical interpretation of the wave function by a more general and formalized theory.

It assumes that states of a system are described by wave functions (properly "unit rays" in the Hilbert space of wave functions) and that observables (real-valued physical quantities) are described by hermitian operations.

In order to get a connection between states and observables, one has introduced the concept of measurement. A stochastic quantum transition is then considered as forced by the measurement of some quantity. E.g. in a scattering of a particle, the detection of the particle as coming out in some special direction can be considered as a measurement of its position at a suitable instant of time after the scattering. This gives an important formalization of the transition concept. It is always connected with and described by a hermitian operator. The wave function after the transition (i.e. measurement) then has to be an eigenstate of the hermitian operator with eigenvalue equal to the value obtained by the measurement.

The conventional quantum-mechanical formalism is inseparably tied to the concept of measurement. Formally it is a theory of measurements rather than a theory of objective changes of states of a system, and in order to apply the theory, every situation has in principle to be considered in some way as a measurement.

In order to see what the conventional formulation of quantum mechanics means conceptually, we shall briefly describe an axiomatization of it. Let us then first note that the part of time-evolution which is described by the time-dependent Schrödinger equation can be conveniently handled by using the Heisenberg picture (see section 7.6 below), which means that the observables and the corresponding operators are labeled by (i.e. functions of) the time-parameter. The state vector Ψ is then independent of the time parameter.

The conceptual basis of the conventional theory is a mathematical structure

$$(S, O, p) \tag{1}$$

where S is a set of states and O is a set of observables. A measurement is then a pair

$$M = (s,\ A) \tag{2}$$

where s is a state in S and A is an observable in O. For every measurement M,

$$p_M(\Omega) = p(s,\ A,\ \Omega)$$

is a probability measure on subsets Ω of the real line. $M = (s, A)$ means that we measure the quantity A on a system which is in the states s (before the measurement). $p_M(\Omega)$ means the probability that the result of the measurement of A gives a value a in the set Ω. Thus, states s, observables A, measurements M, probabilities p and possible values $a\ldots$ of an observable constitute the conceptual basis of the theory, through which the interpretation, i.e. the contact with the physical reality, is given.

To these abstract, conceptual axioms then comes a set of Hilbert space axioms specifying the mathematical representation of S, O and p. Thus S is supposed to be the set of all unit rays in a Hilbert space (or more generally density operators, see chapter 4). To every observable is associated a hermitian (self-adjoint) operator. It is convenient (but no loss of generality) to restrict the observables to so called yes-no questions Q, for which the corresponding operators are projection operators (which has 1 and 0 as the only eigenvalues). Thus O is a set of yes-no questions and to every Q in O is associated a projection operator P_Q. If $s = s_\Psi$ is the unit ray spanned by the unit vector Ψ, then

$$p(s \mid Q) = p(s_\Psi, Q, 1) = \|P_Q \Psi\|^2 \tag{3}$$

is the probability for getting the result 1 = "yes", measuring Q on s.

Chapter 3

Critique of the Conventional Quantum-mechanical Theory

The critique given in this chapter is not only intended as a critique directed to the conventional theory, but is also intended to give arguments supporting the heuristic motivation of the confidence theory given in chapter 5.

3.1 The State-Observable Dualism

It is reasonably the state-observable dualism which is the origin of the central position that the concept of measurement has received in the conventional theory. This concept is introduced in order to get a connection between states and observables. However, these two concepts are both basic in the conventional formalism and this is much more clearly stated there than in the case of interpretation 1. Sometimes (i.e. before a measurement) we describe our system by means of the concept of state and sometimes (i.e. when considering the result of a measurement) we describe our system by means of values of physical quantities (the obtained eigenvalues of the measured observables). Properly, one describes the system by means of observables only at the moment of observation. The state after the measurement is of little concern in the general basic theory (see next section) and one generally does not even assume that the measurement is repeatable ("measurement of 1:st kind"). It can equally well be a non-repeatable measurement ("measurement of 2:nd kind") as e.g. in the case of a detection of a photon.

Thus, the conventional formalism does not solve the state-observable dualism. Rather it confirms it since as was stated in chapter 2, a measurement is conceptually a pair (s, A) and thus presupposes both the state and the observable concept. Obviously, a measurement is not a "relation" between the concepts of state and observable of such a kind that it can reduce these concepts on each other.

Let us also note in this connection that the generalization to ensembles (density operators), by introducing an extra, non quantum-mechanical statistical distribution into the state concept, obviously amplifies the state-observable dualism.

3.2 The State After Measurement

The measurement gives a kind of one-directed relation between states and observables. It gives a way to come from states to observables. It is natural to try to complement it by a way of getting from observables to states. We shall mention three different proposals for such a connection. None of them has, however, a status of general acceptance as a basic principle in the quantum-mechanical formalism.

The first way is the use of "complete sets of commuting observables". If we assume that the state vector after measurement is an eigenvector to the measured observables, this will give a unique prescription for the state after measurement only if we always assume that we measure a complete set of commuting observables. Having the description of macroscopic systems in mind, this assumption is obviously too restrictive.

A second way is to generalize the concept of state to ensembles (density operators) and introduce a "maximum entropy principle". We shall discuss this principle in chapter 4, where we shall reject it as a basic principle.

A third way is to introduce the concept of "ideal preparatory measurement". Consider the measurement of a yes-no observable Q described by the projection operator P_Q and let M_Q denote the closed subspace on which P_Q projects. If we measure on a state described by the state-vector Ψ and obtains the result "yes" (eigenvalue 1) then any vector in M_Q could be the result after measurement. In view of (3) of chapter 2, however, the choice $\Psi' = P_Q\Psi$ seems natural. A measurement with this prescription for the state after measurement is sometimes called an "ideal preparatory measurement". It means in a sense a minimal disturbance on the measured system. The introduction of an extra principle stating that all measurements are ideal preparatory measurements, would obviously reduce the state-observable dualism.

The lack of a general prescription for the state after measurement indicates an incompleteness in the axiomatization of the theory. It makes it principally impossible (without extra assumptions) to apply the theory to the study of successive transitions, which obviously occurs in many complex situations.

3.3 The Arbitrariness Problem

The formalism of conventional quantum mechanics does not prescribe any restriction on which states and which observables are allowed and thus preserves the arbitrariness of states and observables. The introduction of the measurement concept means that collapses of wave functions and observables are both connected to this concept. Thus, the arbitrariness of collapses coincides with that of observables into what we shall call "the arbitrariness of measurements". Any observable can be measured at any time by applying a suitable measuring equipment.

Carrying matters to an extreme, any course of events is possible. Take any prescribed sequence of states $\Psi_1, \Psi_2, \ldots$at the times $t_1, t_2, \ldots$and let $P_1, P_2, \ldots$be the projection operators projecting on the one-dimensional subspaces spanned by $\Psi_1, \Psi_2, \ldots$respectively. Then, measuring P_1 at time t_1, and so on would then, with some positive probability, lead to the sequence $\Psi_1, \Psi_2 \ldots$which thus describes a

possible course of events. (If two consecutive Ψ_i and Ψ_{i+1} are orthogonal, we can intercalate an extra state, e.g. $\Psi_i + \Psi_{i+1}$, in order to get a positive probability. In fact, by intercalating sufficiently many Ψ:s one could get a total probability arbitrarily close to 1.)

This is indisputably what the formalization looks like. The arbitrariness apparently becomes problematic if we try to apply quantum mechanics as it stands to macroscopic phenomena. Now, the measuring apparatus is ultimately a macroscopic system and this obviously puts strong practical limitations on which measurements are possible. In fact it should be pointed out that the Copenhagen interpretation presupposes the existence of a macroscopic physical world in which all measurements are registered. The formally total arbitrariness is therefore strongly reduced when the physical properties of the measuring apparatus are taken into account.

On the other hand, these physical properties of the measurement apparatus are not formalized in the theory, where a measurement is abstracted into a pair (s, A). Thus, the arbitrariness means an incompleteness of the theory.

To obtain a complete general physical theory, we have the following three alternatives:

Either (first alternative) strengthen the theory with extra rules which, explicitly or implicitly, give the same restrictions of the arbitrariness as the physical properties of the measuring apparatus gives or

(second alternative) complement the quantum mechanical theory, dealing only with the atomic part of the system, with a fully axiomatized theory for macroscopic systems. This macroscopic theory should then have a well-defined interface to the quantum mechanical formalism so that the two theories can cooperate and together constitute a complete theory.

The only further (third) alternative would be to give up the quantum-mechanical formalism for a different theory.

Let us note, that also in the first alternative, the theory must in some way include macroscopic physics in order to be accepted as a general complete theory. Perhaps the most important conclusion from these arguments is that the basic conceptual and interpretational questions of quantum theory are strongly coupled to the task of developing a general theory for macroscopic phenomena.

It is reasonable to expect the dynamics of quantum theory to be a general basis from which ultimately also macroscopic phenomena can be predicted. The commonly proposed theory of macroscopic phenomena based on quantum mechanics is quantum statistical mechanics. However, the present status of this theory is far from a complete satisfactory theory. It will be rejected in chapter 4 from general principal arguments.

The Copenhagen interpretation, by presupposing macroscopic physics, has chosen the second of the above mentioned alternatives, waiting for further development of the macroscopic theory. It therefore accepts a dualism between macroscopic and microscopic physics. Even if quantum mechanics is supposed to be a basis also for macroscopic physics, then the Copenhagen interpretation implies

that we have two quantum-mechanical theories – one (equal to that described in chapter 2, i.e. the Copenhagen formalism) for atomic phenomena and another ("dynamical quantum mechanics") at the bottom of quantum statistical mechanics. The two are obviously different since the Copenhagen interpretation is not relevant for the latter. The interpretation is there given through macroscopic quantities, some of which are described by certain mean-values (e.g. energy) and some of which are described by parameters (e.g. temperature) or other characteristic (e.g. entropy) of (non quantum-mechanical!) statistical distributions.

In the following sections we shall review the critics already given by discussing certain central questions in this connection and giving some concluding remarks.

3.4 The Concept of Measurement in the Conventional Theory

A peculiar property of a "measurement" in the conventional theory is that it disturbs the system measured. In fact it forces the initial state to collapse into an eigenstate of the measured observable. A concept with this property is not what is originally meant with the word "measurement". To avoid this misuse of the word "measurement" and thus retain its basic common sense meaning we shall use the term "conventional quantum-mechanical measurement" for the former concept.

The arguments in the preceding section indicate that the division into measured object and measuring apparatus is in some way unsuitable. The arbitrariness shows that too much of the physical conditions is left to the apparatus and has to be complemented by extra assumptions. The measuring apparatus is also a physical system and should be handled on the same footing as the object system. Irrespectively of if we use the quantum theory formalism or not, the stochastic transition in an experimental situation is forced by the complete physical arrangement of object and apparatus together. The occurrence of stochastic transition is an expression for a fundamental indeterminacy of quantum physics. The existence of this indeterminacy is here considered as an experimental fact.

What is specific for "a measuring situation" in difference to an arbitrary course of events is that it is arranged in such a way that there is a coupling (correlation) between some event(s) in the "object" and some event(s) in the "measuring apparatus" so that the latter event(s) can be interpreted as a registration of the former. This general definition of "measurement" can be used both in classical and quantum physics and has logically nothing to do with disturbances or stochastic phenomena.

What is specific for a "conventional quantum-mechanical measurement" is that it is a combination of a forced stochastic transition and a registration (of some property of the state after the transition). Thus, the common statement that a measurement on a microscopic system always disturbs the system, is due to this confusing mixture of registration and forced stochastic transition and an unsuitable use of the word "measurement" for this mixture.

The overstressed central position of the concept of measurement means that every situation has to be considered, in some way, as a measurement in order to apply the formalism. This is not very natural. Instead one would prefer a theory

which describes objective courses of events in a system irrespectively of whether it is arranged as a measurement or not.

3.5 The Real Occurrence of Transitions

A very important characteristic of the conventional formalism is its inability to describe the real occurrence of transitions (collapses of wave functions). This is shown clearly in e.g. Jauch (3), where a distinction is made between what he calls "events" and "data". There, "event" is the abstract, potential concept of what could happen. The real occurrence of an event is called a "datum". The formalized theory (essentially that described in chapter 2) then only deals with "events" (described by yes-no observables) and their probabilities.

The arbitrariness of collapses has led to many confusing discussions on when and why the collapse of the wave function occurs. One opinion in this connection (held among others by Bohr) is that the transition has become an objective fait accompli when the atomic measurement signal has been amplified in an irreversible way into the macroscopic part of the measuring apparatus. The not fully clear and convincing character of this statement, leaning on sophisticated technical questions of irreversible statistical mechanics, has lead to controversial discussions on subjectivity and of the ultimate role of the observer.

In the conventional formalism, the real occurrence of transitions (Jauchs "data") are thus pushed aside to the question of interpretation and we are left with a deterministic dynamics. The real occurrence of transitions is an observable phenomenon which has such an obvious physical content that it should reasonably be less step-motherly treated by the formalization of the theory.

3.6 Conclusion

Much of the critics of the conventional theory, given here and in other places, could be rejected by the argument that, if we apply it in the right way, it works. (With this pragmatic attitude one can equally well do with interpretation 1 and the choice between the two interpretations seems to be a question of taste).

However, if one seeks for a complete unified physical theory describing both atomic and general macroscopic phenomena, the incompleteness of conventional quantum mechanics indisputably rules it out as a candidate. Thus we reject the conventional Copenhagen formalism for being unsuitable and incomplete rather than incorrect.

The incompleteness reduces it practical value. The obviously unsuitable character of its basic concepts should give a hint to what has to be changed.

Chapter 4

An Alternative to Statistical Mechanics

4.1 Conceptual Critique of Different Schools in Statistical Mechanics

4.1.1 The Kinetic Theory of Matter

In the kinetic theory of matter statistical arguments are used in combination with mechanical considerations concerning colliding molecules. The theory has led to a good principal understanding of the behavior of macroscopic systems from a molecular point of view and has also led to a series of good qualitative predictions. However, the statistical arguments are used in a vague and heuristic way. The most ambitious and systematic branch of the kinetic theory is the theory of the Boltzmann equation. It is a natural approach based on a conceptually well-defined amplitude (an occupation number distribution in the 6-dimensional, "one-particle", phase space). It has however met with two serious difficulties. Firstly, there exists no rigorous derivation of the Boltzmann equation from the underlying mechanics. Secondly, the equation itself is difficult to handle with available mathematical methods without making dubious approximations.

Although statistical arguments play an essential role in the kinetic theory, the kinetic theory is traditionally considered as distinct from "statistical mechanics". The distinction is that the kinetic theory makes essential use of the mechanical properties of the system studied while statistical mechanics tries to minimize the use of mechanical assumptions, making maximal use of statistical methods. This distinction is in one respect unfortunate. The obviously non-rigorous character of the kinetic methods has lead to the misconception that statistical mechanics is the more fundamental theory. In fact this distinction indicates the incompleteness of statistical mechanics.

There exist a lot of problems which are more or less successfully treated by the kinetic theory, but which lie completely outside the scope of statistical mechanics. Thus statistical mechanics takes too little account of the underlying dynamics when judged as a general theory of macroscopic phenomena. Such a theory must give a theoretical framework also for typical kinetic problems.

4.1.2 Boltzmann Statistics

It is characteristic for the kinetic theory that it considers distributions of small systems which occur as multitude parts in a large macroscopic system. Many of the "probabilities" occurring in the kinetic theory can be re-expressed in terms of averages or in terms of distributions of matter (particles) in different states of motion. Thus, many of the distributions in the kinetic theory are properly not statistical distributions but rather mass distributions, some of which are directly coupled to macroscopically observable distributions such as the distribution of total mass in space $\rho(x)$ and the distribution in space of collective velocity $v(x)$.

In contrast to the "Boltzmann statistics", the "Gibbs statistics" considers instead a distribution (ensemble) of the state of the total system. The latter approach is motivated by a wish to take the interaction of the small subsystems into account. In fact, the weak point of the kinetic theory is that this interaction is not based on and deduced from the underlying mechanics, but is handled by vague intuitive "statistical" arguments and assumptions. On the other hand, the methods of the "Gibbs statistics" are far removed from the concrete mass distribution concepts and do not either present a method for proving the heuristically obtained results of the kinetic theory.

4.1.3 The Ergodic Theory

The basic assumption of the ergodic theory is that macroscopic quantities are given by (infinite) time averages of corresponding mechanical quantities for the microscopic system. The main problem of the theory is to prove the equality of time averages and phase-space averages. The reason for introducing phase-space averages is that the calculation of phase-space averages is a task accessible with common mathematical methods (which is not the case for time averages).

The most immediate, apparent drawback of this theory is the use of infinite time averages and the corresponding limitation to stationary (equilibrium) states. It is not easy to see how the theory could be generalized to describe non-stationary, time-varying courses of events such as e.g. irreversible thermodynamic processes.

A critique of the use of time averages will be given below.

4.1.4 Statistical Ensembles

Except for the kinetic theory and discussions concerning Boltzmann statistics, all schools of statistical mechanics use statistical ensembles as the central mathematical concept. Mathematically, an "ensemble" is a probability distribution on the phase-space of the mechanical system. In the ergodic theory, this probability distribution is introduced formally as a way of introducing statistical terminology and methods for the handling of phase-space averages. (See Khinchin (4), chap. IV). In the other schools it is the basic concept. (The use of the term "statistical mechanics" for the theory of macroscopic systems based on microscopic mechanics is an expression for this). A macroscopic "state" is described theoretically by an ensemble and macroscopic quantities are supposed to be expectation values (mean values) of mechanical quantities over this ensemble (compare the "state-observable dualism in quantum theory described in chapter 1).

When discussing a specific macroscopic state, one has to make an assumption about the form of the corresponding statistical distribution. The only practically interesting ensembles used to describe systems in equilibrium are the microcanonic and the canonic ensembles. However, there are several schools giving different arguments for the use of a specific form of the statistical distribution. These schools are connected with different attitudes concerning the meaning of the probabilities occurring in the ensemble concept.

4.1.5 Different Concepts of Probability

In the ergodic theory, the nature of the probabilities occurring in the ensemble concept is quite clear. They are not true probabilities but are introduced formally as auxiliary quantities used in the computations. They have no need for interpretation.

As mentioned above, "statistical mechanics" is characterized by the tendency to minimize the mechanical assumptions and drawing as many conclusions as possible from statistical arguments. In the exposition on the foundations of statistical mechanics given by Gibbs (5), one has taken an attitude to the ensemble concept which means that the purpose with the theory is to construct a "statistical model" for thermodynamics rather than derive it from first principles. Thus, e.g., the canonical ensemble is introduced for simplicity rather than being motivated from some general principles. The concept of probability is in this exposition (and also in many considerably later expositions) very primitive as the term "ensemble" indicates.

The "a priori school" (see e.g. Tolman (6)) introduces a basic principle – "the principle of equal a priori probabilities in phase space" (complemented in quantum mechanics by a principle of random a priori phases). The idea is that we are considering a system of which we have only partial knowledge, and after putting the restrictions we know on the state of the system, we are left with a set of states (phase space points) all satisfying these restrictions. The principle then gives a rule of weighting these states.

Equal a priori probabilities in phase space is natural due to the special properties of the phase space measure. It presupposes, however, that the conditions are so formulated that they lead to a set of initial states. On the other hand, the ensemble which the principle determines the form of is used to derive mean values for the description of physical states. Thus we have a "dualism of states" (similar to "the state observable dualism" of quantum theory). The states are both characterized by mean values and by "conditions representing our partial knowledge of the system", leading to phase space regions. There is no principle on how to get from former to the latter.

A way of avoiding this "state dualism" is given by "the maximum entropy principle" (see e.g. Jaynes (7)). One then introduces the "entropy of a statistical distribution" $p(x)$ (in this case the ensemble) by an expression

$$H = -k \int p(x) \log p(x)\, dx$$

and determines the distribution as that which maximizes the entropy under given subsidiary conditions. These subsidiary conditions then expresses that the mean values of certain quantities, e.g. the energy, are given. Thus one starts with mean values and ends up with other mean values calculated from the determined distribution (ensemble).

Let us point out that there is a large ambiguity in the choice of the form of the distribution for a given macroscopic state. Many distributions give the same relations between macroscopic quantities. Especially the microcanonic and the canonic ensembles are equivalent in describing relations between thermodynamic equilibrium variables. This ambiguity of the distributions makes the attempts to give arguments for using a special form of the distribution questionable. Too much meaning has been given to the distributions. The often stated proposition that the microcanonic ensemble describes an adiabatic (heat isolated) system and the canonical ensemble describes a system in thermodynamic equilibrium with a heat bath, is an example of this. Such a distinction between isolated systems and systems in thermal equilibrium is completely theoretical and physically meaningless.

The á priori principle and especially the maximum entropy principle has lead to considerations on the concept of "subjective probabilities" (see Jaynes (7)). Such probabilities is an expression for our lack of knowledge of the system. They are not "true probabilities" having a frequency interpretation.

In any case, the probabilities in an ensemble describing a macroscopic system can not be "true probabilities" having a frequency interpretation. One can give two arguments for this. Firstly, the ensemble is a probability distribution in a space with an extremely large number of dimensions. To determine this distribution experimentally by using a repeated sequence of experiments on the system is not only practically but also principally impossible. It would contradict the whole idea of thermodynamic irreversibility as being connected to the principal impossibility of measuring the exact state of a macroscopic system containing a large number of constituents. Thus, the probability distribution in an ensemble describing a macroscopic system is an unphysical (unmeasurable) quantity.

Secondly, our lack of knowledge of the exact state of our system under concern is always connected with a simultaneous lack of knowledge of its relation to the environment. This leads to an uncertainty in the distribution itself. This is in accordance with the above mentioned ambiguity of the distribution.

The conclusion is that the probability distribution in an ensemble is a concept introduced for computational reasons, having no physical interpretation.

4.1.6 The Use of Mean Values

A characteristic feature of all the above mentioned schools of statistical mechanics is the use of mean values to describe macroscopic quantities. However, the connection between a macroscopic quantity and a suitable mean value in the formalism is introduced by using vague and intuitive arguments rather than being the consequence of a deductive theory.

Some schools (see e.g. Jaynes (7))consider statistical mechanics as a matter of statistical inference theory. The statistical inference theory is very controversial and we seem to have another instance of "how to lie with statistics".

In the ergodic theory, the use of time averages is commonly motivated by introducing the concept of a "macroscopic measurement", whose properties and functioning is however not further analyzed. (The situation is in fact quite similar to that in the Copenhagen formulation of quantum mechanics). One then argues that this measurement is not performed instantaneously, but requires a certain interval of time and that what is measured is the time average of the quantity under concern. That the basic interpretation of the theory relies on a concept of "macroscopic measurement", which is not further analyzed is completely unsatisfactory from a general principal point of view, since the main purpose of the theory was just to describe and understand macroscopic processes of which a measurement is a special instance.

It seems difficult to give a physical motivation for the use of mean values. None of the schools offers such a motivation and it always has to be assumed that the mean values describe macroscopic quantities. The only motivation seems to be that of computational simplicity.

4.1.7 Critique of the Ensemble Concept

The purpose of the ensemble (and the mean values taken over it) is to describe the coarse properties of a system. That it fulfills this purpose is by no means self-evident.

The ensemble describes the mean course of events of a large set of systems but a real course of events is described by one single system. This presents a conceptual jump. It seems rough to take the mean-value over all possible states. Some of them must have very exceptional behavior far from that of the macroscopic system to be described. That these "exceptional" states are so few that they do not contribute to the mean value does not follow from a statistical argument but is a dynamical property of large systems ultimately to be proved.

Thus, although ensembles really happens to give a good description of the coarse properties of a system (at least for equilibrium states), it seems difficult to give a theoretical motivation for this. None of the mentioned schools give a satisfactory motivation. This indicates that the ensemble is not a suitable concept in an atomistic theory for macroscopic systems based on first principles.

A further argument leading to this conclusion will be given in the next section.

4.1.8 Non-equilibrium States and Irreversible Processes

Perhaps the most apparent indication for the incompleteness of statistical mechanics is its difficulties to describe non-stationary systems. It is not obvious how a non-equilibrium state should be described by an ensemble. A theory including also non-equilibrium states is necessary for the understanding of the conditions for equilibrium.

In order to describe the increase of entropy in irreversible processes one has to introduce extra assumptions on "coarse-grained" distributions (and then prove a H-theorem). (The introduction of coarse-grained distributions is motivated by an argument on macroscopic measurements and we again have a parallel to the Copenhagen formalism of quantum mechanics.). Although this is reasonable in

some sense, the resulting formalism is not very elegant and is far from a closed and well-developed general theory.

The ensemble concept was invented to describe macroscopic states. The necessity of "coarse-graining" shows, however, that the ensemble is not sufficient to describe macroscopic states. This clearly indicates that the ensemble concept is unsuitable in a general theory.

4.2 An Event Theory of Collective Phenomena

Having rejected statistical mechanics, we shall now give an alternative, direct, non-statistical (non-probabilistic) approach. We shall use the term "the event theory of collective phenomena" to denote this theory, thus avoiding the term "statistical mechanics" which we shall use to denote the (rejected) probabilistic approach. We avoid the term "statistical" since this term is too much associated with probability theory.

The discussion will be very brief, with emphasis on the conceptual part. It will constitute a complement to the heuristic arguments for the confidence theory formulation of quantum mechanics, which is a generalization of, and thus incorporates, the event theory of collective phenomena.

4.2.1 Mechanics as Basis for Macroscopic Phenomena

We shall consider the following as our basic object in this context: *To develop a theory of macroscopic systems based on some given microscopic mechanics*.

This is a purely dynamical problem. In spite of the fact that a macroscopic system is composed of a very large number of particles, we shall try to handle this problem in a very direct way. Although much of the ideas will be taken over to the quantum-mechanical treatment, we will, for simplicity, first reason as if classical mechanics for some atomistic particles constitute the basis for macroscopic physics. The generalization to quantum mechanics will be discussed in section 4.2.11.

As a simple example, showing that it is possible to handle problems of this kind by direct methods we can take the theorem of motion of the center of mass. For some body, macroscopically considered in some context as a particle, we can use this theorem to derive rigorously the macroscopic behavior from a given underlying supposed microscopic mechanics, irrespectively of how many microscopic particles the body is composed of. We shall consider this simple example as a prototype when trying to handle more complicated questions e.g. the motion of a continuous mass distribution.

If mechanics is to be the basis, then we must take as our basic ideas that:

1) In a concrete real situation, what we are observing is the macroscopic behavior of one single possible motion of the mechanical system. We are not observing a set ("ensemble") of systems.

2) There exists a macroscopic behavior for at least some microscopic motions i.e. the macroscopic quantities are defined as characteristics of the microscopic

motion and the relations between macroscopic characteristics follows from the underlying microscopic dynamics.

3) In spite of the empirical evidence for the existence of a macroscopic behavior, this existence should also ultimately be proved to be a consequence of the underlying mechanics.

Point 3) is important both from the conceptual point of view and for the deductiveness of the theory. Thus e.g. it is a question to be handled by the theory to decide whether a certain portion of matter under certain circumstances can adequately be described as a laminar liquid motion.

4.2.2 General and Dynamic-depending Results

We will not here try to achieve the goal completely, but instead divide the way to the goal into two steps:

1) Establish a general conceptual basis for the description of macroscopic systems and derive general results under certain general dynamical assumptions.

2) Prove these dynamical assumptions from the underlying dynamics in specific cases.

As an example of "dynamic assumption" we can take the assumption that a certain portion of matter can be described by a continuous distribution of mass density, velocity and inner energy. A first step in formulating dynamical assumptions will generally consist in stipulating a purely kinematic description of the system. It can then be extended to include assumptions on e.g. equations of motion and constitutive equations.

This division of the complete problem is very practical and in fact necessary. Step 1 must be worked out completely before we have any chance to attack the much more difficult step 2 in a rational way.

Extracting step 1 from the complete goal makes it possible to rigorously formulate and prove such general results as classical thermodynamics. The dynamical assumptions, on the other hand, gives clearly formulated basic questions to be investigated for different systems with specifically stipulated dynamics. This work belongs to the application of the general theory and we shall here restrict ourselves to a discussion of the general part of the theory i.e. step 1. It should be noted, however, that the dynamical assumptions, even if unproven for the moment, are generally very natural or reasonable, as the example above shows. They are often essentially identical to some axiomatic assumptions in a corresponding classical theory.

4.2.3 The Concept of Approximation as Basis for the Theory

In our prototype example of motion of a macroscopic particle, the dynamical assumption is the assumption that a certain portion of matter is (as a consequence of the underlying mechanics) held together to form a body which at every instant of the motion is confined to a small region in space. Under this assumption we

can then rigorously prove the validity of the equations of motion of the body considered as a particle. Step 1 then means a general derivation of the equations of particle mechanics at the macroscopic level. It should be noted that the obtained equations are in general approximate to the extent that the body can be considered as a point since the total external force on an extended body has to be approximated by some corresponding point-particle value. Thus, the derivation also gives an estimate of this approximation i.e. how the accuracy of the point-particle description depends on the small diameter of the body.

We can now see the basic role of the concept of approximation in this context. In fact we can consider the whole question of describing and deriving the macroscopic properties of a system as a question of approximation. We shall define macroscopic quantities in terms of the microscopic quantities and the derived relations between these macroscopic quantities will then generally only be approximate (although the "errors" are small from a macroscopic point of view).

This appearance of approximations in the theory leads us to consider quantities which are confined to certain regions in the corresponding state space rather than consider them as having exactly known values. Subsets in the state space of the microscopic systems are also the natural concept to use in view of the fact that a macroscopic state is described by a much smaller number of variables than the complete microscopic state. This can also be considered as a kind of approximation. We only know the microscopic state partially or approximately.

This general concept of approximation, which is in concordance with the concept of neighborhood in general point-set topology, unifies the deductive and the conceptual aspects of the theory. It is the natural concept to use for the definition of "macroscopic states" and a derivation of relations between macroscopic quantities or macroscopic equations of motion cannot be rigorous and complete unless we give rigorous estimates on the corresponding degree of accuracy.

4.2.4 Macroscopic Events

We are thus lead to consider certain sets of microscopic states. These sets describe the macroscopic properties of the system. All the different states in the set have the same macroscopic property described by the definition of the set. All the states are similar and lie near each other in this sense. In the language of topology, the states lie in some neighborhood of each other, namely the neighborhood defined by the set.

To say that an element (e.g. a state) belongs to a certain set is the common mathematical way to say something about the element. From the logical point of view, a set is thus connected to a statement (about some arbitrary element belonging to the set). In our physical context, a subset in the state space of a system will then correspond to a statement about the physical system. We shall call such statements "events". An event has occurred in a certain real situation if the corresponding statement is true i.e. the state belongs to the corresponding set defining the event.

The motion of a mechanical system is in this context conveniently described by means of the phase-space (constituted by the configuration space variables together with their canonically conjugated momenta) which we shall take as our

state space. A point in the state-space is then an instantaneous state of the system and the motion is described by giving the phase-space point as function of the time variable. We shall call such a function an "orbit" and the equations of motion singles out which orbits are possible.

We shall base our theory on instantaneous events which are statements about the system at a given instant of time. An instantaneous event is then a pair $e = (\Omega, t)$, where Ω is a subset of the phase-space and t is an instant of time. The event e has occurred if the orbit passes through Ω at time t.

An instantaneous macroscopic state is generally characterized by a set of events $e_1 = (\Omega_1, t)$, $e_2 = (\Omega_2, t)$, …, $e_n = (\Omega_n, t)$, all associated to the same instant of time t. e_1 may e.g. specify the energy of the system, e_2 may specify the space region occupied by the system etc. The statement "e_1 and e_2 and …and e_n" is then also an (instantaneous) event. It is the logical conjunction $e_1 \wedge e_2 \wedge \ldots \wedge e_n$ of the events e_1, …, e_n which obviously is identical with (Ω, t), where $\Omega = \Omega_1 \cap \Omega_2 \cap \ldots \cap \Omega_n$ is the intersection of the sets Ω_1,…, Ω_n.

We shall now concretize our general abstract concept of event by giving a simple example of that we shall call "macroscopic events". These express coarse, collective statements about a system composed of a large number of particles.

We shall consider an "ideal gas" composed of n identical free particles of mass m contained in a region V in configuration space $\mathbb{R}^3$. The phase-space X of this system is constituted by the position variable $\vec{x}_i \in \mathbb{R}^3$, $i = 1, 2, \ldots, n$ together with their conjugated momenta

$$\vec{p}_i = m\vec{v}_i \in \mathbb{R}^3_p,\ i = 1, 2, \ldots, n,$$

where $\vec{v}_i$ is the velocity of the i:th particle and $\mathbb{R}^3_p$ denotes the 3-dimensional momentum space. Thus $X = \mathbb{R}^{3n} \times \mathbb{R}^{3n}_p$. The localization of the gas to the region V at time t is then described by the event $e_1 = (\Omega_1, t)$ where

$$\Omega_1 = \left\{ \vec{x}_i, \vec{p}_i; \vec{x}_j \in V \quad \text{for all } j \right\} = V^n \times \mathbb{R}^{3n}_p.$$

The energy of the system is

$$U = \frac{1}{2m} \sum_{i=1}^{n} p_i^2$$

As we shall see below, in connection with the definition of entropy, it is convenient to consider statements where the energy is confined to an interval $(U - \Delta U, U + \Delta U)$ rather than having an exact value. Thus, the energy of the system is characterized by the event $e_2 = (\Omega_2, t)$, where

$$\Omega_2 = \left\{ \vec{x}_i, \vec{p}_i; \frac{1}{2m} \sum_{i=1}^{n} p_i^2 \in (U - \Delta U, U + \Delta U) \right\}$$

Taking e_1 and e_2 together we obtain the event $e(V, U, \Delta U, t) = e_1 \wedge e_2 =$

$(\Omega_1 \cap \Omega_2, t)$ where

$$\Omega_1 \cap \Omega_2 = \left\{ \vec{x}_i, \vec{p}_i; \vec{x}_i \in V \text{ and } \frac{1}{2m} \sum p_j^2 \in (U - \Delta U, U + \Delta U) \right\}.$$

describing both the volume occupied and the total energy of the system.

In this example we have made several simplifications and idealizations. The "wall" or environment enclosing the gas is not treated from a molecular mechanical point of view and we have not considered the number of particles n to lie in a relatively small interval rather than having an exact value. The latter generalization is necessary if we want to use a set of statements to describe a space distribution of matter. We refer to chapter 6 below for treatment of variable particle numbers and to (1) for the treatment of boundary conditions.

The statement $e(V, U, \Delta U, t)$ does not say anything about how the particles and the energy are distributed within the region V. However, if the particles and the energy are distributed continuously (not necessarily uniformly) over the region, we can describe this distribution with sufficient accuracy by using a sufficiently large set $e_i(V_i, U_i, \Delta U_i, t)$, $i = 1, 2, \ldots$ of such events with sufficiently small $V_i \subset V$.

4.2.5 Deterministic Processes

In general we are not only interested in a specific instantaneous state but in the evolution of the system in time. This can generally be described by a sequence of events $e_1 = (\Omega_1, t_1)$, $e_2 = (\Omega_2, t_2)$,..., $e_n = (\Omega_n, t_n)$, expressing statements about the system at a sequence of times $t_1, t_2, \ldots, t_n$. We shall call such a sequence a "course of events". The statement that all these events have occurred in a certain real situation then obviously means that the orbit of the system passes through all the sets Ω_i at the corresponding times t_i. If there exists at least one orbit with this property, we shall say that the course of events $e_1, \ldots, e_n$ is possible.

A basic problem of dynamics is the study of how an "initial state" of the system at time t_1 develops into another state at a subsequent time t_2. Such an evolution is in many cases deterministic. We shall restrict the discussion here to deterministic evolutions, deferring the discussion of general indeterministic evolutions to the quantum-mechanical theory developed below.

Let our initial state be described by the event $e_1 = (\Omega_1, t_1)$, and the final state by the event $e_2 = (\Omega_2, t_2)$. If the evolution is deterministic, then the event e_2 must necessarily occur if the event e_1 has occurred. This means that every possible specific microscopic motion of the system, for which we know that e_1 has occurred must have the property that also e_2 occurs. Thus every orbit, which passes through Ω_1 at time t_1 also passes through Ω_2 at time t_2. We have thus defined what is to be meant with the statement "e_1 implies e_2" (in logical notation $e_1 \Rightarrow e_2$) and we have reduced the concept of deterministic evolution to the concept of implication for events.

Instead of using the notation $e_1 \Rightarrow e_2$, we shall, for reasons that will become clear in the quantum-mechanical generalization below, denote this implication $(e_1|e_2) \in T_1$. The pair $(e_1|e_2)$ will be called a "process". e_1 is called the "premise" or the "initial conditions" and e_2 will be called the "outcome". T_1 is the set all processes which have certain outcome.

We shall now introduce a convenient representation of events which is quite analogous to the introduction of the Heisenberg picture in quantum mechanics. We have made the assumption that our underlying microscopic dynamics is described by classical mechanics. From this it follows that exactly one orbit passes through a given phase-space point at a given time and that the orbits maps phase-space points (or sets) at one instant of time t_1 in a one-to-one way on phase-space points (or sets) at another time t_2. To every event $e = (\Omega, t)$ we can then in a one-to-one way associate a pair (M, t), where M is the set of all orbits for which e occurs i.e. passes through Ω at time t. The advantage of this representation is that events $e_1, e_2, \ldots$ at different times $t_1, t_2, \ldots$ can be represented by subsets $M_1, M_2, \ldots$ of a common space namely the set of all orbits. A course of events $(e_1, e_2, \ldots, e_n)$ is then possible if and only if the intersection $M_1 \cap \ldots \cap M_n$ of the corresponding M:s is nonempty and $(e_1|e_2) \in T_1$ is equivalent to $M_1 \subset M_2$. The course of events $(e_1, e_2, \ldots, e_n)$ is an event i.e. a statement about the physical system which expresses the same thing as the logical conjunction $e_1 \wedge e_2 \wedge \ldots \wedge e_n$. This "composite" event, which is non-instantaneous if some of the t_i are different, is represented by the intersection $M_1 \cap M_2 \cap \ldots \cap M_n$.

We have thus extended our "logic of events" to events at different times and to non-instantaneous events. Processes $(e_1, \ldots, e_m|e'_1, \ldots, e'_n)$ where the initial condition $(e_1, \ldots, e_m)$ and the outcome $(e'_1, \ldots, e'_n)$ are courses of events can then be given a meaning. We now see that $(e_1, \ldots, e_m|e'_1, \ldots, e'_n) \in T_1$ if and only if $M_1 \cap M_2 \cap \ldots \cap M_m \subset M'_1 \cap M'_2 \cap \ldots \cap M'_n$ for the corresponding M:s. This statement expresses the dynamical rule that if we know that the initial conditions $e_1, \ldots, e_m$ are satisfied (has occurred) then we know that the course of events $e'_1, \ldots, e'_n$ will then necessarily happen (occur).

4.2.6 Irreversibility, Causality and the Direction of Time

Before we can apply our general theory of processes to macroscopic processes, we have to make two complements to it in order to handle the problem of irreversibility. First, we shall, in this section, introduce a "principle of causality", which introduces an asymmetry with respect to forward and backward time direction. Second, we shall, in section 4.2.9 generalize our concept of process to ε-processes in order to take into account the fact that there may be exceptional states in the phase-space, which do not develop in a "regular macroscopic way".

The underlying microscopic dynamics is reversible since we have assumed it to be classical mechanics. Thus it follows that if $e_1 = (\Omega_1, t_1)$, $e_2 = (\Omega_2, t_2), \ldots, e_n = (\Omega_n, t_n)$ is a possible course of events, then the backward (time-reflected) course of events $e'_n = (\Omega_n, -t_n), \ldots, e'_1 = (\Omega_1, -t_1)$ is also possible. (Our macroscopic events will generally be symmetric with respect to a change of all velocities v to $-v$.)

This is true if we consider a closed system of whatever size it may be. However, in any real situation, we have to take into account also the environment of the system under concern and its interaction with the system. Even if our system under concern is more or less completely isolated during a period of time, its initial state comes out as consequence of some kind of preparation by its preceding interaction with other surrounding systems. Thus, a description of a sufficiently large part of the environment of our system during a sufficiently large preceding period of time

is necessary in order to guarantee the initial state of the system. Such a description can be made by a (possibly large) sequence of events $e_1, \ldots, e_m$ which we call the preparing initial conditions of the system and a process $(e_1, \ldots, e_m | e'_1, \ldots, e'_n) \in T_1$ then describes how these initial conditions forces the course of events $e'_1, \ldots, e'_n$ in our system to occur. e'_1 can then e.g. be an internal initial state of the system if it is isolated from that time.

Thus we see the importance of the environment. A time reflection can not be made only for a limited part of the universe which is our "system under concern". The time reflection must be made for the whole or at least for a very large part of universe and under a considerable period of time.

From our everyday experience it is very natural to consider the initial conditions to be given before the course of events considered. This is our intuitive idea of causality that cause comes before action. It is generally realized that causality is an ultimate condition for life as we know it and our general perception of the world, our possibility of making observations and possibility of communication.

Another reason for taking the conditions to be given before their action comes from the empirical irreversibility of many macroscopic processes such as e.g. heat conduction. A process according to the classical heat conduction equation is empirically possible but all empirical evidence says that the opposite time-reflected process, which satisfies a heat equation with a minus sign on the time derivative is impossible. Causality is in this case a natural mathematical consequence of the form of the equation. The initial value problem for the heat equation is stable and well-posed in the forward time direction but extremely unstable and ill-posed in the backward time direction. This indicates the importance of stability in this context.

There exists unstable processes such as e.g. condensing in an undercooled gas, which to some degree could look like an irreversible process going backward in time. In a turbulent gas flow we have instabilities roughly similar to that of a backward heat equation. However, the macroscopic laws governing these phenomena are not completely symmetrical with respect to time-reflection. Even if reversible convective mechanisms are dominant in a highly developed turbulent flow, there are also dissipative forces which act in an irreversible damping way. Thus even if our underlying microscopic dynamics is reversible, we must end up with a macroscopic theory whose equations of motion are manifestly unsymmetrical with respect to time-reflection. Although they may admit instabilities to develop to some degree, they contain dissipative mechanisms which results in an overall stability (in the forward time direction).

Our overwhelming experience of causality is an expression for the fact that our whole world is so organized that all "preparations" must act in forward time direction or, equivalently, that all predictions must be made by ultimately using a sufficiently comprehensive knowledge of the previous history of the system under concern and its environment rather than using a correspondingly comprehensive knowledge of its future, which we generally do not have. It is directly connected to a high overall stability in forward predictions together with a high "overall instability in backward predictions" which generally characterizes macroscopic systems, and which in turn is connected to the general irreversibility characterizing

macroscopic systems.

We shall now introduce the basic postulate, based on experience, that all physical situations must ultimately be described by processes $(e_1, \ldots, e_m | e'_1, \ldots, e'_n)$, such that the events e_j in the initial conditions are associated to times t_j, which are before all the times t'_k of the outcome events e'_k i.e. $t_j < t'_k$ for all $j = 1, \ldots, m$, $k = 1, \ldots, n$. We shall call this postulate "the principle of causality". In the following, we shall let T_1 denote the set of processes fulfilling this time-ordering condition.

In the next sections we shall discuss how thermodynamic irreversibility can be described and understood from a molecular mechanical point of view as a consequence of this principle of causality. Before we shall do this, we shall make a general comment on the fact that our world is organized according to this causality.

Thus, the whole of our (observable) universe is in a situation, which, from a mechanical point of view is characterized by a special kind of motion with a highly developed asymmetry with respect to forward and backward time direction. From a pure mechanical point of view such a situation seems accidental. It is only one of a set of possible motions of the universe.

We can imagine a general chaotic, unstable universe without any manifest causality in either time direction. This universe could then perhaps for some reason undergo an expansive evolution, which results in a far developed separation and condensation into a large set of stable macroscopic systems. This stability can then, reasonably, by the following argument occur only in one time direction.

In such a universe, it will be a common process that systems come into contact with each other and remain so for some subsequent period of time. The juxtaposition of two macroscopic systems with e.g. different temperature, would then, due to mixing in a system composed of a large number of particles, with extremely high certainty lead to a more uniform distribution of temperature in the composite system. The inverse process, where a portion of matter with uniform temperature distribution by itself undergoes an evolution which leads to different temperatures in different parts (which then might be separated for some reason), is on the other hand from a molecular mechanical point of view extremely exceptional and improbable. This example makes probable the existence of high instabilities in the backward time direction. A high general overall stability of macroscopic systems in one time direction is thus, by mechanical properties of systems composed of a large number of particles, necessarily connected to a corresponding general high instability in the opposite time direction. (We shall discuss this effect of large number of particles more fully in the following sections).

We have seen that the existence of extreme instabilities in the backward time direction in certain processes is a consequence of the existence of macroscopic systems, which are stable in the forward time direction. Thus causality and irreversibility is essentially reduced to the existence of stable macroscopic systems.

On the other hand, we have given no explanation why parts of our universe should condense into separated stable systems by giving out energy to the surrounding universe. Thus we see that if we try to reduce the apparent macroscopic irreversibility to the underlying reversible microscopic mechanics, we have to introduce extra assumptions on the organization of the environment, which are of a cosmological nature. This reductionism is thus obtained at the cost of introducing

cosmological assumptions which are difficult to formalize. Instead we shall, by introducing the principle of causality as a basic, irreducible postulate, eliminate these cosmological questions and obtain a local deductive theory.

4.2.7 The Entropy of a Macroscopic System

In the discussion of irreversibility below, the volumes of the regions in phase-space corresponding to macroscopic events will play an important role. The logarithm of the phase-space volume is, apart from a constant factor k (the Boltzmann constant), under certain circumstances equal to the entropy of the system. We shall denote the volume of the phase-space region Ω by $\mu(\Omega)$. μ is then a measure i.e. an additive set function. We shall also write $\mu(e)$ for $\mu(\Omega)$ if e is the event for $e = (\Omega, t)$.

For our ideal gas example above we get for $e = e(V, U, \Delta U, t)$

$$\mu(e) = V^n S_{3n} \Big\{ \big[2m(U + \Delta U)\big]^{3n/2} - \big[2m(U - \Delta U)\big]^{3n/2} \Big\}$$

where v is the 3-dimensional volume of V and

$$S_n = \frac{2\pi^{n/2}}{n\Gamma\left(\frac{n}{2}\right)}$$

is the volume of the unit sphere in n dimensions.

We shall now assume n to be very large and that

$$\frac{1}{n} \ll \frac{\Delta U}{U} \ll 1$$

and make use of the fact that the volume content of a sphere in n dimensions with n large is essentially concentrated to a thin layer along the surface. We then get

$$(U + \Delta U)^{3n/2} - (U - \Delta U)^{3n/2} = U^{3n/2} \Big[\big(1 + \frac{\Delta U}{U}\big)^{3n/2} - \big(1 - \frac{\Delta U}{U}\big)^{3n/2} \Big]$$

where

$$\big(1 + \frac{\Delta U}{U}\big)^{3n/2} \approx e^{\frac{\Delta U}{U} \cdot \frac{3n}{2}} \gg 1$$

and

$$\big(1 - \frac{\Delta U}{U}\big)^{3n/2} \approx e^{-\frac{\Delta U}{U} \cdot \frac{3n}{2}} \ll 1.$$

Thus the term $(U - \Delta U)^{3n/2}$ is very small relative to $(U + \Delta U)^{3n/2}$ and can be neglected. Taking the logarithm we then get

$$\log \mu(e) \approx n \log V + \log S_{3n} + \frac{3n}{2} \log[2m(U + \Delta U)].$$

In this expression ΔU gives a relatively very small contribution since by assumption

$$\log(U + \Delta U) - \log U \approx \frac{\Delta U}{U} \ll 1$$

and hence

$$\log \mu(e) \approx n \log V + \log S_{3n} + \frac{3n}{2} \log(2mU)$$

is (approximately) independent of ΔU

Thus, by admitting an "uncertainty" $\Delta U \gg \frac{U}{n}$ in the energy, we are able to define an entropy

$$S = k \log \mu(e)$$

of our event e, which only depends on the volume V, the energy U and the number of particles n (alternatively on the macroscopic mass $M = n \cdot m$) of the system. Note that S is only approximately defined since we have made two approximations in order to get a result independent of ΔU.

The quantity ΔU can be chosen arbitrarily to a large extent and has no physical significance in itself. Rather it is to be considered as a technical device in our approximation technique ultimately leading to relations between quantities with obvious "physical significance" such as energy (U) and entropy (S).

The obtained formula for the entropy gives a well-defined value of the entropy of the system confined to the region V only if the particles and the energy are uniformly distributed over V. If this distribution is non-uniform, we have to divide the region into sufficiently small regions $V_1, V_2, \ldots$ such that the distribution within each is (sufficiently) uniform. The entropy S of the total system in V is then given by the sum

$$S = \sum S_i$$

of the entropies S_i of the subsystems confined to the regions V_i.

We shall make an important observation concerning the comparison of the phase-space volumes of a non-uniform and a uniform distribution. Let our space region V with volume v be divided into two parts V_1 and V_2 with volumes v_1 and v_2 and let us assume $v_1 = v_2$ for simplicity. If there are m particles in V_1 , then this will contribute to the phase-space volume with the factor

$$\binom{n}{m} v_1^m v_2^{n-m} = \binom{n}{m} v_1^n$$

where we have taken into account that the n identical particles can be divided into one group with m and one group with $n-m$ elements in $\binom{n}{m}$ ways. By the binomial theorem we have

$$v^n = (v_1 + v_2)^n = \Big[\sum_{m=0}^{n} \binom{n}{m} \Big] v_1^n$$

and by the properties of the binomial coefficients for large n we have that nearly all of the volume v^n comes from terms with m close to $n/2$. Thus there exists a

relatively small interval $I_1 = \left(\frac{n}{2} - \Delta m, \frac{n}{2} + \Delta m\right)$ and a very small $\varepsilon > 0$ such that

$$v_1^n \sum_{m \notin I_1} \binom{n}{m} < \varepsilon v^n.$$

For n large, ε will be extremely small.

Let $I_2 = (m_2 - \Delta m_2, m_2 + \Delta m_2)$ be another interval disjoint to I_1, and let $e(m \in I_1)$ and $e(m \in I_2)$ be the events corresponding to the statements that the number of particles in V_1 lies in the interval I_1 and I_2 respectively. Then $e(m \in I_2)$ contributes to the total phase-space volume with a factor which is less than ε (properly $\frac{\varepsilon}{1-\varepsilon}$) times smaller than that of $e(m \in I_1)$ which gives essentially the same contribution to the total phase-space volume as the event $e =$ "there are n particles in V".

Similar arguments hold for non-uniform distributions of the energy. Thus we come to the conclusion that the phase-space volume of a non-uniform distribution is generally extremely small in comparison to the phase-space volume of the corresponding uniform distribution.

4.2.8 The Increase of Phase-space Volume

We have seen that the orbits maps subsets of the phase-space at one instant of time in a one-to-one way on subsets of the phase-space at another instant of time. By the Liouville theorem this map preserves the phase-space volume i.e. the measure which we have denoted by μ. From this it follows that we can define a measure, which we shall also denote by μ, on sets M of orbits. Then $\mu(M) = \mu(\Omega)$ if Ω is the set through which the orbits in M passes at some arbitrary time.

We have introduced above the representation of events $e = (\Omega, t)$ by pairs (M, t). The measure on sets M then makes it possible to generalize the concept of phase-space volume to courses of events $e_1 = (\Omega_1, t_1), e_2 = (\Omega_2, t_2), \ldots, e_n = (\Omega_n, t_n)$. Thus we define

$$\mu(e_1, e_2, \ldots, e_n) = \mu(M_1 \cap M_2 \cap \ldots M_n)$$

where $(M_1, t_1), \ldots, (M_n, t_n)$ are the representations of $e_1, \ldots, e_n$. Applying this to a process
$(e_1, \ldots, e_m | e'_1, \ldots, e'_n) \in T_1$ we get, since $M_1 \cap M_2 \cap \ldots \cap M_m \subset M'_1 \cap M'_2 \cap \ldots \cap M'_n$ for the corresponding M:s implies

$$\mu(M_1 \cap \ldots \cap M_m) \leq \mu(M'_1 \cap \ldots \cap M'_n),$$

the general result

$$\mu(e_1, \ldots, e_m) \leq \mu(e'_1, \ldots, e'_n).$$

Thus our theory implies a general result on increase of phase-space volume.

4.2.9 Exceptional States and ε-Processes

Let $e_1 = (\Omega_1, t_1), e_2 = (\Omega_2, t_2), e_3 = (\Omega_3, t_3), t_1 < t_2 < t_3$ be three macroscopic events which describe a deterministic evolution of a closed system. Then e_1 implies e_2 and e_2 implies e_3. We have to explain why the system strives at equilib-

rium. If we can show that there is a general strive at more uniform distributions (for mass, energy, etc.) it both follows that uniform distributions are equilibrium states (since they cannot develop into any other state than itself) and that there is a general strive at equilibrium.

If e_1, e_2 and e_3 correspond to different states of the system which do not have uniform distributions, then e_2 should have more uniform distribution than e_1 and e_3 have more uniform distribution than e_2. By the arguments at the end of section 4.2.7 this means that $\mu(M_1) \ll \mu(M_2) \ll \mu(M_3)$ where M_1, M_2, M_3 are the M:s corresponding to e_1, e_2, e_3.

An increase of phase-space volume, and thereby a general tendency towards uniform distributions would follow from the result of the preceding section if we could apply our concept of process. We cannot, however, strictly assume that

$$M_1 \subset M_2 \subset M_3$$

for the following reason. The macroscopic situation at t_2 described by e_2 can come out as result of many different preceding situations and e_1 is only one of them. Thus only a very small subset Ω_1' of Ω_2 represents states which are results of evolutions from the situation e_1 at t_1. If $(x_1, \dots, p_1, \dots) \in \Omega_1'$ is such a state, then the velocity-reversed state $(x_1, \dots, -p_1, \dots)$ would obviously develop into a state at time $t' = t_2 + (t_2 - t_1)$ for which the event (Ω_1, t'), similar to e_1, would occur. This would generally contradict the situation e_3 at t_3 if e.g. $t_3 = t'$ and would contradict our general experience of irreversibility. We shall call such states exceptional states. These must clearly be avoided in a more correct definition of process.

Now, we can achieve this in a general and simple way, without giving a detailed dynamical specification of all different exceptional states, by observing that the set of all exceptional states will have a relatively very small measure. In our example above, the set Ω_1' of states in Ω_2 at t_2 which are results of previous states in Ω_1 at t_1 have measure $\mu(M_1)$ and $\mu(M_1) \ll \mu(M_2)$ as assumed.

By the arguments at the end of section 4.2.7, it is natural to assume that the set of all exceptional states together has a relatively very small measure. E.g. we have seen in 4.2.7 that a uniform detailed mass distribution accounts for essentially all of the phase-space volume and all other non-uniform distributions, and there will be many such, will together concur on a extremely small fraction of the possible phase-space.

Thus, instead of demanding that $M_2 \subset M_3$, we shall assume that there exists a subset of exceptional states $M_{2,\text{ex}} \subset M_2$ such that

$$\mu(M_{2,\text{ex}}) \leq \varepsilon \mu(M_2)$$

and

$$M_{2,\text{reg}} = M_2 - M_{2,\text{ex}} \subset M_3$$

i.e. the set $M_{2,\text{reg}}$ of "regular" (non-exceptional) states will necessarily result in states at t_3 for which e_3 occurs. Here ε is to be a very small positive number. A

pair $(e_2|e_3)$ satisfying this condition will be called an ε-process.

Generally, we make the following definition: $(e_1, \ldots, e_m|e'_1, \ldots, e'_n)$ is an ε-process and we write $(e_1, \ldots, e_m|e'_1, \ldots, e'_n) \in T_{1,\varepsilon}$ if $t_i < t'_k$ for all i, k where t_i, t'_k are the corresponding times of e_i, e'_k, and

$$\mu(M_1 \cap \ldots \cap M_m - M'_1 \cap \ldots \cap M'_n) \leq \varepsilon \mu(M_1 \cap \ldots \cap M_m)$$

where M_i, M'_k are the M:s corresponding to e_i, e'_k, and the minus sign denotes set subtraction i.e. $A - B$ is the set of all elements in A which are not elements in B if A and B are sets.

Our theorem above on increase of phase-space volume has to be slightly modified for ε-processes. Thus, if $(e_1, \ldots, e_m|e'_1, \ldots, e'_n) \in T_{1,\varepsilon}$, then

$$\mu(e_1, \ldots, e_m) \leq \frac{1}{1-\varepsilon} \mu(e'_1, \ldots, e'_n)$$

We have given a general rigorous definition of the concept ε-process. To get a physical theory out of it we now make the following "application postulate": We can choose a very small positive number ε such that any behavior of macroscopic systems can be completely described by ε-processes $(e_1, \ldots, e_m|e'_1, \ldots, e'_n) \in T_{1,\varepsilon}$ of macroscopic events $e_1, \ldots, e_m, e'_1, \ldots, e'_n$.

This postulate implies that we assume that the exceptional states are so extremely rare or miraculous that we can neglect them altogether. We shall call this "the principle of confidence" and $1 - \varepsilon$ will be called the confidence level. We shall use this principle, thus formulated, without any probabilistic interpretation. An introduction of a principle of equal à priori probabilities in phase-space as introduced in the à priori school of statistical mechanics would of course make our confidence principle natural and would give a probabilistic interpretation of $\mu(\Omega_{\text{ex}})$ and ε. By the arguments given at the end of section 4.1.5 above, however, we reject any attempt to give a physical meaning to a probability distribution in the phase-space of a macroscopic system.

That the exceptional states are confined to a phase-space region with very small relative measure ε is provisionally to be regarded as a mathematical conjecture to be ultimately proved (or disproved) from the underlying dynamics.

4.2.10 Further Development of the Theory

Motivated by a sequence of heuristic arguments we have given a rigorous base for step 1 of section 4.2.2. This base of events and ε-processes constitutes an axiomatic basis. A lot of further theoretical work has to be done before step 1 is completed. E.g. we have to introduce such concepts as temperature and pressure, study continuous mass distribution, boundary conditions etc. This work is essentially of a technical character and consists in working out approximation techniques for asymptotic behavior of systems at large particle numbers.

Before it is worthwhile to do this work, however, we have to generalize the theory to quantum mechanics. We will therefore finish this brief discussion on collective phenomena by giving the most essential step in this generalization.

4.2.11 The Quantum-Mechanical Substitute for Phase-Space Regions

Let us, for simplicity, discuss the case of a "one-dimensional particle". The generalization to an n-particle system in 3-dimensional configuration space will be straightforward.

Our system is then described by a configuration space variable $x \in \mathbb{R}^1$ and a momentum variable $p \in \mathbb{R}^1_p$. In the quantum-mechanical case it is, due to the uncertainty principle, impossible to give a physical meaning to a phase-space point (x, p). On the other hand, if we admit suitable uncertainties Δx and Δp in the variables, we can give a meaning to an approximate phase space point. By using the concept of confidence estimate, we can give a precise meaning to this. Thus we shall say that the system, described by the instantaneous wave function $\psi(x)$, is localized to the space interval $I_x = \left(x_0 - \frac{\Delta x}{2}, x_0 + \frac{\Delta x}{2}\right)$ if

$$\int_{I_x} |\psi(x)|^2 \, dx \geq (1 - \varepsilon_x) \int_{-\infty}^{\infty} |\psi(x)|^2 \, dx$$

Similarly, we shall say that the system is localized to the momentum-space interval
$I_p = \left(p_0 - \frac{\Delta p}{2}, p_0 + \frac{\Delta p}{2}\right)$ if

$$\int_{I_p} |\hat{\psi}(p/\hbar)|^2 \, dp \geq (1 - \varepsilon_p) \int_{-\infty}^{\infty} |\hat{\psi}(p/\hbar)|^2 \, dp$$

where $\hat{\psi}$ denotes the Fourier transform of ψ. If both these estimates are satisfied with $\varepsilon_x, \varepsilon_p < \varepsilon_2$ we shall say that the system is localized to the phase-space region $\Omega = I_x \times I_p$ with the confidence level $1 - \varepsilon_2$. ε_2 is to be a very small positive number. As e.g. the example of Gaussian wave-packet shows, we can obtain such a localization with very small ε_2 if $\Delta x \cdot \Delta p$ is a few times larger than $\hbar$.

We cannot obtain this localization with both ε_x and ε_p zero. It can be shown that there is a positive lower limit of $\varepsilon_x + \varepsilon_p$ depending on $\Delta x \cdot \Delta p$.

The theorems of (1) Chapter I are of fundamental importance for the present theory. In the application to macroscopic events, $\Delta x \cdot \Delta p \gg \hbar$ and many wave-functions $\psi(x)$ will satisfy the phase-space localization condition above. By the results of (1) Chapter I, the set of all such wave-functions essentially constitutes a subspace N of the Hilbert space with finite dimension

$$\dim(N) \approx \frac{\Delta x \cdot \Delta p}{\hbar \cdot 2\pi}$$

Macroscopic phase-space localizations can thus be represented by finite-dimensional subspaces N of the Hilbert space. Our concept of deterministic process can then be generalized to quantum theory by replacing the M:s (sets of orbits) by corresponding subsets N of Heisenberg picture vectors and the logical implication $M_1 \subset M_2$ by

$$N_1 \subset_{\sqrt{\varepsilon_2}} N_2 \tag{1}$$

as defined in (1) Chapter I. We have thus introduced a quantum-mechanical confidence level $1 - \varepsilon_2$ beside our previous confidence level $1 - \varepsilon$ associated with exceptional states. The generalization of (1) to ε-processes, analogous to the definition of 4.2.9, is straightforward.

The introduction of subspaces is very convenient from a technical point of view. It means that we can build the whole theory on subspaces and the corresponding projection operators.

The direct connection between phase-space volume and the dimension of the corresponding subspace means that we can instead use the expression

$$S = k \log \dim(N)$$

in the definition of entropy in the quantum-mechanical theory. This reduces the dimensional anomaly of taking the logarithm of a quantity with dimension (action)^{3n} for an n-particle system, and gives a well-defined zero point of the entropy. It shows the deep connection between thermodynamics and quantum theory.

In conventional discussions of quantum-mechanical measurement processes, considering interaction with the macroscopic measuring apparatus, this macroscopic apparatus way of ultimately transforming a superposition to a macroscopic fait accompli of a wave function collapse is often referred to as an irreversible macroscopic process handled by irreversible statistical mechanics. See 3.3 and 3.5 above.

In the present theory, the collapse and the irreversible process is handled in a unified way by the same concept of ε-equiangular sequence of projections (see below) of which (1) above is a special case. This also shows a deep connection between thermodynamics and quantum theory.

Chapter 5

Heuristic Derivation of the Confidence Theory

5.1 The Problematic Nature of the Concept of State

There exists two difficulties, which make it generally problematic to ascribe a state to a system.The first difficulty is the general impossibility to define states of subsystems, due to the "indivisible unity" of quantum systems. The second difficulty is connected with the collapses of the wave functions.

One of the most characteristic non-classical features of quantum physics is the indivisible unity of quantum systems (Bohm (2)). This means roughly that a system cannot in general be considered as divided into its constituent parts. The "constituent parts" are in general bound together into a unit in such a way that the constituents have no longer any separate meaning and individuality. E.g. an atom is not composed in a classical sense of a nucleus and electrons, but these particles are bound together into an indivisible unit. If we try to identify one of its constituents, then we destroy the original atom, and we are left with two or more other systems (e.g. an electron and an ion).

Expressed in terms of wave functions, the indivisible unity of a system "composed" of a set of particles means that the wave function of the composite system cannot be factorized into separate wave functions for some subsystems. However, one can express the indivisible unity in a more general way by using the concept of state. Thus, the indivisible unity of a system means that it cannot be divided into some subsystems in such a way that the state of the whole system is uniquely determined by the states of its parts.

The indivisible units are not confined to atomic systems but can take macroscopic dimensions in space. A molecule is not composed in a classical sense of atoms and a molecule can be very large. Even if the molecules in a piece of matter are small, they are bound together by intermolecular forces and this composite system is in principle an indivisible unit. Strictly speaking, the whole world is an indivisible unit, and the occurrence of separate subsystems having well-defined states is always connected to some degree of approximation.

On the other hand, even if indivisible units cover large space regions, we can ascribe properties to much smaller regions in space. E.g. we can define a mass

density (approximate but with a high degree of accuracy) within a large crystal. Thus "property" is a more generally applicable concept than "state".

Another difficulty in the definition of states arises from the occurrence of collapses of wave functions.Even if the wave function may be well-defined before and after the scattering of a particle, there is no reasonable way of defining an intermediate wave function at an instant of time during the process of scattering. In the description of a single scattering experiment we can of course refrain from this, but if we intend to develop a general theory describing general complex systems with many more or less simultaneous transitions, this is obviously a problem. Thus, for such a system, we cannot in general ascribe an instantaneous state (wave function) to the whole system. On the other hand, the indivisible unity, previously discussed, makes also the definition of states of parts difficult.

5.2 The Irreducibility of Collapses

The occurrence of stochastic transitions in quantum physics has thrown in a completely new concept, which does not fit into the general customary conceptual frameworks of classical physical theories described by differential equations. This has led to several attempts to explain and reduce these transitions to the dynamics.

One way is to try to describe the transition as caused by some disturbance in the system due to its interaction with the environment. However, the treatment by quantum mechanics of a larger system composed of our first system under concern and its environment ultimately leads to the same problems of explaining the cause of the transition for this larger system. The stochastic element in the transition cannot e.g. be ascribed to a "stochastic molecular chaos" in the macroscopic environment as one could perhaps be inclined to believe before a thorough analysis of the problem.

Another way is the search for "hidden parameters". Although some interesting ideas are proposed by D. Bohm which shows in a sense an equivalence with a "submicroscopic" deterministic process, this branch of research has not so far given any reasonable, consistent and convincing alternative to quantum mechanics.

The resulting, overwhelming impression left is that the stochastic quantum transitions cannot in a reasonably simple and consistent way be reduced to dynamics. We therefore consider it as an irreducible, elementary concept of quantum physics.

5.3 The Collapse-Projection Connection

In the wave function description of a quantum transition, an initial Hilbert space vector evolves according to the time-dependent Schrödinger equation into a superposition

$$\Psi(t) = \sum_{i \in I} c_i(t) \Psi_i$$

of very different states at a later instant of time t. These different vectors correspond to very different eigenvalues a_i of some property (observable) A of the system, but by the circumstances, the actual observable must have taken a more or

less defined value at the time t. Thus only a small subset Ψ_i, $i \in I'$ of the different states fits this actual value taken of the observable, and by the statistical interpretation of the wave function, the probability for the obtained value of the observable is

$$p = \sum_{i \in I'} |c_i(t)|^2$$

(the Ψ:s above are supposed to be normalized).

However, this also means that the vector $\Psi(t)$ has collapsed into a new vector $\Psi'(t)$, fitting the value taken by the observable. The most natural assumption is that the "circumstances forcing the new value of the observable" simply cuts off the vectors Ψ_i not fitting the new situation and we are left with

$$\Psi'(t) = \sum_{i \in I'} c_i(t)\Psi_i.$$

This can be described in a simple way by means of a projection operator. The set of vectors Ψ_i, $i \in I'$ corresponds to a subspace M of the Hilbert space of wave functions and to it corresponds a projection operator P projecting on M. We can then write

$$\Psi' = P\Psi$$

and

$$p = \|P\Psi\|^2$$

Thus we are led to the following assumption which we shall call "the collapse-projection connection":

1) A stochastic quantum transition is always connected to and described by a projection operator.

2) In a transition described by the projection operator P, an initial wave function corresponding to the Hilbert space vector Ψ collapses into the new wave function corresponding to the Hilbert space vector $P\Psi$.

As was pointed out in section 1, there is in general a great ambiguity in the choice of wave functions. This is certainly one reason why the choice $P\Psi$ for the wave function after a collapse has never been accepted as a general basic principle in quantum-mechanical expositions. The choice of wave function after a collapse is to a high degree dependent on what kind of idealization we make in the division into object system and disturbing environment, and is therefore to some extent a matter of definition. The choice $P\Psi$ is however simple and natural and we shall use it in the heuristic motivation of the confidence theory.

Remark: If we compare the collapse-projection connection with the Copenhagen interpretation formalism described in chapter 2, then condition 1) follows from the coupling between transitions and measurements and from the (practical) restriction to yes-no questions. Condition 2) corresponds to the assumption of ideal preparatory measurements (see section 3.2 above).

5.4 The Conceptual Basis of a Physical Theory

We have pointed out in chapter 1 that a one-to-one relation between some set of basic quantities (or concepts) in the theory and some observable quantities (or concepts) in the physical reality is important for the objective interpretation of a theory. Let us now discuss what properties these basic quantities or "basic observables" must have in order to constitute the basis of a theory. We shall call them basic observables, using the term observable in a general sense. They need not necessarily coincide with the ordinary "quantum-mechanic observables" described by hermitian operators. It is important to point out the difference between what we shall call the "conceptual" (properly "physico-conceptual") or "kinematical" part of a theory, which gives the basic concepts having a direct physical interpretation, and the "dynamical" part of the theory which gives the complete mathematical description of the relations between the concepts of the theory.

If a certain set of concepts is to constitute the conceptual basis of a theory, then they must satisfy the following conditions:

1) It must be possible the describe the whole course of events in any real physical situation completely in terms of the basic observables. Other tangible physical quantities which are not among then basic observables must have well-defined relations to the basic observables so that they can be reduced to (i.e. defined in terms of) them.

2) All physical laws can ultimately be expressed in terms of the basic observables. The basic laws (dynamical equations etc.) can of course be expressed by means of auxiliary quantities not among the basic observables but must at least implicitly (by means of thought elimination) lead to a complete set of relations between the basic observables.

5.5 Initial and Boundary Conditions

Let us now discuss the interpretation of the basic observables of a theory.

The interpretation of some given set of observables ("object observables") describing a certain system under concern is depending on their relation to some other similar or different observables ("environment observables") describing some environment of the system. If the theory is to describe an objective physical world, then these relations between "object observables" and "environment observables" must be reduced to physical laws i.e. ultimately described by means of the laws connecting the basic observables. The environment here has to comprise a sufficiently large part of space-time in order to contain a sufficiently exhaustive description of the initial and boundary conditions.

More precisely, that an "object observable" has an objective interpretation means that it should in principle be possible to determine its value from the values of some other "environment observables". If the theory is to describe an objective physical reality, then every "basic observable" must admit of such an interpretation. We can draw two important conclusions from this.

First, it is important that not only the object system itself but also its environment, describing initial and boundary conditions and other systems, with which our object interacts, are described by the same set of basic observables.

Second, the theory has to be "dynamically closed" in the sense that it contains sufficiently many relations between the basic observables so that any of them, when describing an object system, could be determined from values of basic observables in the environment.

This does not mean that every basic event that occurs in reality is always measured or observed by its relation to other events in the environment, as one would be inclined to believe from the conventional quantum theory, which (at least in principle) only deals with measurements. (Here we have another interesting instance of the "arbitrariness of collapses", see 3.5 above: What about all these events that have occurred in the real world but have never been observed or measured?)

The statement must reasonably be interpreted in a weaker sense, replacing really measured by measurability in principle for the specific event. The specific event could be measured if it would occur in a (another) suitable environment.

Much of the interpretational problems in statistical mechanics and quantum theory, (especially the arbitrariness problem), are connected with the fact that these theories are formulated as "dynamically open" theories, where the environment, (especially the initial and boundary conditions), are not sufficiently formalized in the theory and are not handled by the dynamics of the theory in a unified way. We have pointed out in chapter 4, that the use of mean-values in statistical mechanics is difficult to motivate by physical arguments. This can be reformulated by saying that the mathematical structure given by these mean-values constitutes a dynamically open theory, which is not easily extended into the structure of a complete, dynamically closed theory. Such a theory should give a complete ("closed") set of relations connecting these mean-values and eventually some other quantities with each other, derived from the dynamics of the theory.

5.6 Goal and General Line of Direction

We shall now give a sequence of arguments leading to the confidence theory.

We shall try to find a suitable set of "basic observables" satisfying the general criteria given in the preceding two sections. We can then expect to obtain a theory, which is equally unproblematic with respect to the interpretation as classical theories are.

The usual quantum-mechanical theory works well (i.e. gives correct results) in several practical cases. Instead of looking for a completely new, different theory, we shall make the "continuity assumption" that the ordinary theory lies very near to a correct theory. In view of this "continuity assumption", we can then expect that wave functions will still play an important role in the dynamical part of the theory as "probability amplitudes" i.e. "auxiliary" quantities by means of which probabilities can be calculated, but they are not suitable for direct interpretation (i.e. they are not belonging to the "kinematical" part of the theory).

Thus, assuming that the ordinary mathematical machinery of wave functions satisfying the Schrödinger equation and transitions connected with certain projection operators essentially gives the correct dynamics, we shall direct our attention to the "kinematics".

5.7 A Theory Describing Successive Transitions

It is not only the time-dependent Schrödinger equation, but also the collapses of the wave functions which describes the really occurring changes in the physical reality. As we have pointed out in section 2 above, the discontinuous collapses cannot be reduced to the dynamics of the Schrödinger equation. The occurrence of stochastic quantum transitions is a fundamental physical phenomenon, which must be taken into account if we want to construct a general theory describing arbitrary courses of events. If we want to establish a one-to-one correspondence between events in the real world and some corresponding concepts in the theory, then also these collapses must be built into the theory.

Thus, our theory must be a "probabilistic" theory. It must describe successive transitions i.e. complete sequences of transitions in a unified way.

5.8 Basing the Theory on Observables

We have pointed out in chapter 1 that the state-observable dualism makes it difficult to find some concepts ("basic observables"), which stand in a one-to-one relation to reality. However, "the problematic nature of states" described in section 1 above makes it natural to reject the states altogether. Let us therefore try to avoid the "state-observable dualism" and the "problematic nature of states" by simply avoiding consequently the concept of state and base the whole theory solely on the concept of "observable". This means that the "role played by the states" must be taken over by the observables. How this can be done will be discussed in section 11 below.

In view of the "continuity assumption" above, it is natural in first hand to try to find the "basic observables" among the "quantum-mechanical observables", represented by hermitian operators. In view of the arbitrariness problem, it is also natural to expect that some special class of such observables shall play a basic role in the theory.

We shall assume that the basic observables are given by a special class of projection operators, namely those corresponding to space localization of particles. In the next two sections we shall give arguments for using projections. Arguments for the restriction to space localizations will be given later in section 14.

5.9 First Argument for Using Projections

The first problem encountered when describing a physical quantity by means of hermitian operators is that a wave function is not in general an eigenstate of the operator. This means that the quantity in general does not have a well-defined value. In situations where this spreading of values is "too large" as e.g. in the

outgoing wave in a scattering process, this problem is solved by the introduction of the statistical interpretation of the wave function and collapses of wave functions.

However, even if the collapse has reduced the wave function to an eigenstate of some operator, this eigenstate cannot in general correspond to a completely exact value of the observable for practical reasons. This is especially clear when one considers operators as e.g. those corresponding to position or momentum of a particle, which have continuous eigenvalue spectrum. Thus even after the collapse we are left with a value which is in general not exactly well-defined. This presents a problem when one tries to construct a one-to-one relation between the observables in the theory and the observables in the physical reality.

The common proposal for the description of rough (not exact) values of a physical quantity is to introduce some probability distribution for it. The "value" of the quantity is then given by its mean-value (expectation value) with respect to this distribution, and the accuracy of the (approximate) quantity, often denoted by its "uncertainty" or "mean error" is given by the usual "standard deviation" expression.

Although the wave function suggests a probability distribution for the quantity (that of interpretation 1 described in chapter 1), the interpretation of this probability distribution is not very clear since, as was just stated, the quantity has only a physical meaning as an approximate quantity. In chapter 4 we have given a critique against the use of mean-values to describe rough physical quantities. This critique can in its essential points also be used against the use of mean-values in quantum mechanics. The present task of describing a rough physical quantity in quantum mechanics and the task of chapter 4 of describing rough macroscopic quantities are in fact, from our general unified point of view, the same problem. Thus according to the critique given in chapter 4, supplemented by the arguments at the end of section 5 of the present chapter, we reject the use of mean-values when looking for a set of basic quantities suitable for a one-to-one correspondence with real observables.

However, there exists another way of describing a rough real-valued quantity, namely by means of an interval, the length of which giving a measure of the accuracy of the quantity. This is seldom used (outside pure mathematics) despite its simplicity. It is however very suitable in quantum mechanics since it corresponds to a very simple kind of hermitian operator namely a projection operator. That a quantity is localized to an interval of the real line means quantum-mechanically that the wave function vanishes when the corresponding variable lies outside the interval. The set of all such wave functions constitute a subspace of the Hilbert space and to it corresponds a projection operator. Localization to the interval then corresponds to the eigenvalue 1 of this projection operator.

The use of intervals, and more generally regions in an n-dimensional euclidean space, is in complete accordance with the theory of macroscopic systems, developed in 4.2, based on phase-space regions, and its coupling to subspaces and projection operators in quantum mechanics was already anticipated in section 4.2.11. As in chapter 4, we shall denote a localization statement (localization to a region) by the term "event".

Perhaps the most important advantage of using such events of localization (and the corresponding projection operators) is that they can be used directly to describe

macroscopic distributions of matter and are therefore suitable in a general unified theory of macroscopic and atomic phenomena.

5.10 Second Argument for Using Projections

Another argument for using projections is provided by the "collapse-projection connection". It is a great unification if both our characterization of physical properties, observables, and the description of the transitions (probabilities and wave functions after transitions) are provided by means of the same mathematical concept.

5.11 Initial Conditions and Equiangularity

If we reject the use of the concept of state, then, the role ordinarily played by the "states" has to be taken over by the basic observables corresponding to a special class of projection operators. Now, it is especially in the role of "initial state" that the wave functions insist on a "state interpretation". (After a collapse described by a projection operator corresponding to a basic observable, the system is naturally described by this observable, e.g. the rough position of a particle). Thus we have to describe also the "initial state" or properly the "initial conditions" by means of our basic observables (events). This is also in concordance with the arguments in section 5 above. The essential role of the wave function is that of a probability amplitude. From a practical (experimental) description of the initial conditions we have to associate to them, in some way, an initial state wave function, which is then used to determine transition probabilities. We have to replace this procedure by a way of determining the transition probabilities when the initial conditions are described by our basic observables (events). This could easily be done if we have a prescription of how to associate a wave function (now considered only as an auxiliary mathematical quantity without any "state" interpretation) to our sequence of events describing the initial conditions.

In the following it is convenient to use the Heisenberg picture (see section 7.6 below).

A subspace of wave functions having the property of "fitting the initial conditions" can be defined in a natural way. Let our initial conditions be described by a sequence of events $e_1, e_2, \ldots, e_m$ at times $t_1 \leq t_2 \leq \ldots \leq t_m$ represented by the projection operators $P_1, P_2, \ldots, P_m$ taken in the Heisenberg picture representation. This sequence of projections then describes a sequence of collapses which occur during the initial process which successively confirms the initial conditions (one would ordinarily say "prepares the initial state"). Let M_1 denote the subspace on which P_1 projects. Then any vector Ψ_1 in M_1 is of the form $\Psi_1 = P_1\Psi$ for some suitable Ψ (e.g. $\Psi = \Psi_1$). By the second part of the collapse-projection connection, it is then natural to express this by saying that after the collapse P_1, then every vector in M_1 fits the initial conditions (so far described only by P_1). By the second part of the collapse-projection connection, it then also follows that a vector Ψ_1 in M_1 collapses by P_2 into $P_2\Psi_1$. The set of all vectors of the form $P_2\Psi_1$ with Ψ_1 in M_1 is denoted by P_2M_1. It is thus natural to say that after the collapse

P_2, then any vector in $M_2 = P_2 M_1$ fits the initial conditions (so far described by P_1 and P_2. Repeating the argument, we come to the definition that the subspace $M_m = P_m P_{m-1} \ldots P_2 M_1$ is the set of vectors which fit the initial conditions $P_1, P_2, \ldots, P_m$.

A problem that now arises when one tries to determine the probabilities of subsequent transitions is that the preceding procedure for obtaining a wave function fitting the initial conditions does not give a single wave function. Even if we have a large number of initial conditions $P_1, \ldots, P_m$, we will in general end up with a subspace M_m of infinite dimension. A statistical distribution on M_m (ensemble) would of course give well-defined probabilities for subsequent transitions, but, according to the critique given in chapter 4, we reject this statistical-mechanical approach.

Now, the transition probability is a physically measurable quantity. This means that it must be uniquely determined by the initial conditions provided that we have a sufficiently complete description of them. It is reasonable to assume that a sufficiently complete description of these initial conditions can be obtained by means of our basic observables (localization events). The initial conditions are ultimately described by means of macroscopic observable concepts and this macroscopic description can be obtained by means of a sufficiently large number of our localization events taken over a sufficiently large period of time.

On the other hand, if quantum mechanics is correct, then this physical transition probability should be given by a wave function fitting the initial conditions. It is therefore reasonable to assume, in the case when several wave functions fit the initial conditions, that any of them can be used to calculate the probability, i.e. they all give the same probability, for a subsequent transition. The idea that a set of wave functions all have the same property is parallel to the idea of 4.2 above that a set of states all have the same macroscopic properties.

The central idea of the present theory is the study of sequences of events $e_1, e_2, \ldots, e_m, e'$ such that every wave function fitting the initial conditions $e_1, \ldots, e_m$ gives the same transition probability for a subsequent transition (event e'). Such sequences thus have the property, that the transition probability can be calculated from a knowledge of the initial conditions only, making no other assumption than the basic statistical interpretation of the wave function.

If M is a non-empty subspace and P' is a projection operator, we shall say that P' is "equiangular" with respect to M if

$$\|P'\Psi\|^2/\|\Psi\|^2$$

has the same value for all nonzero vectors Ψ in M, and this common value is denoted by $p(M|P')$. The term "equiangular" is motivated from the fact that $p(M|P') = \|P'\Psi\|^2/\|\Psi\|^2 = \cos^2\alpha$, where α is the angle between a vector Ψ in M and its projection $P'\Psi$ on the subspace M' on which P' projects. This angle is then equal for all vectors in M. We shall also say in this case that the $(1,1)$-tuple $(M|P')$ is equiangular and alternatively that the $(1,1)$-tuple $(P|P')$ is equiangular, where P is the (orthogonal) projection on M.

More generally, we shall say that the $(m,1)$-tuple

$$(P_1,\dots,P_m|P')$$

is equiangular if P' equiangular with respect to $M_m = P_m P_{m-1}\dots P_2 M_1$ where M_1 is the subspace on which P_1 projects and we define $p(P_1,\dots,P_m|P') = p(M_m|P')$.

We can now express the above property of sequences of events $e_1,\dots,e_m,e'$, by saying that a $(m,1)$-tuple $e_1,\dots,e_m,e'$ of events, where $e_1,\dots,e_m$ describes the "initial conditions" and e' describes the "outcome", corresponds to an equiangular sequence

$$(P_1,\dots,P_m|P')$$

of projections, where $P_1,\dots,P_m,P'$ are the projections corresponding to the events $e_1,\dots,e_m,e'$.

$p(P_1,\dots,P_m|P')$ is then the probability for the transition P' (event e'), under the assumption that the transitions $P_1,\dots,P_m$ (events $e_1,\dots,e_m$) have occurred and we shall alternatively denote this probability by $p(e_1,\dots,e_m|e')$.

An $(m,1)$-tuple $(e_1,\dots,e_m|e')$ of events with this property will be called a (possible) "process".

In the next section we shall generalize this to (m,n)–tuples $e_1,\dots,e_m|e'_1,\dots,e'_n$ describing a sequence $e'_1,\dots,e'_n$ of transitions under given initial conditions $e_1,\dots,e_m$.

5.12 Successive Transitions

Let us now consider successive transitions. Let

$$(P_1,\dots,P_m|P'_1,P'_2)$$

be a process with two successive transitions P'_1 and P'_2, under the given initial conditions $P_1,\dots,P_m$. The subspace M of wave functions fitting these initial conditions is given by

$$M = P_m P_{m-1}\dots P_2 M_1$$

where M_1 is the subspace on which P_1 projects. We want to define what is to be meant with the statement that the pair (P'_1,P'_2) is equiangular with respect to M.

If $(P_1,\dots,P_m|P'_1,P'_2)$ is a process, then the first step of it $(P_1,\dots,P_m|P'_1)$must also be a possible process and thus be equiangular.

We shall denote by $P^c = 1-P$ the projection which projects on the orthogonal complement of the subspace on which P projects.

The transition P'_1 has a certain probability

$$p_1 = \|P'_1\Psi\|^2$$

where Ψ is any unit vector in M, and the opposite transition P'^c_1 has the probability $p^c_1 = \|P'^c_1\Psi\|^2 = 1-p_1$ which corresponds to the possible (equiangular) process

$$(P_1,\dots,P_m|P'^c_1)$$

(Note that the equiangularity of $(P_1, \ldots, P_m | P_1'^c)$ follows from the equiangularity of $(P_1, \ldots, P_m | P_1')$.

After the transition of P_1', the sequence $P_1, \ldots, P_m, P_1'$ must be the initial conditions for the subsequent (possible) process $(P_1, \ldots, P_m, P_1' | P_2')$ and the corresponding subset fitting the initial conditions is $P_1' P_m \ldots P_2 M_1 = P_1' M$. Thus P_2' must be equiangular to $P_1' M$.

If $(P_1, \ldots, P_m | P_1', P_2')$ is a possible process, it is then natural to assume that also $(P_1, \ldots, P_m | P_1'^c, P_2')$ is a possible process and that consequently also the process $(P_1, \ldots, P_m, P_1'^c | P_2')$ is possible. Thus P_2' is also equiangular with respect to $P_1'^c M$.

If $(P_1, \ldots, P_m | P_1', P_2')$ is possible, then it is reasonable that also $(P_1, \ldots, P_m | P_2')$ is a possible process. It just means that we have chosen not to consider P_1' for the moment and study the occurrence of P_2' irrespectively of whether P_1 or P_1' has occurred. The equiangularity of P_2' with respect to M does not follow from the preceding statements on successive equiangularity. However, it follows if we assume that the four subspaces

$$P_2' P_1' M,\ P_2' P_1'^c M,\ P_2'^c P_1' M,\ P_2'^c P_1'^c M$$

are orthogonal.

This orthogonality condition is natural in view of the fact that the four "composite events" corresponding to

$$(P_1', P_2'),\ (P_1'^c, P_2'),\ (P_1', P_2'^c),\ (P_1'^c, P_2'^c)$$

are mutually exclusive. The orthogonality of the corresponding "fitting subspaces" is then natural, since these four cases should be distinguishable, e.g. in a subsequent detection. The possibility of such a detection (registration of the transitions by means of their coupling to some other events in the environment) was stated in section 5 above as an important condition for the possibility of an objective interpretation of the theory.

This makes it natural to make the following

Definition: $(M | P_1', P_2')$ is said to be equiangular if

1) $(M | P_1')$ is equiangular
2) $(P_1' M | P_2')$ is equiangular
3) $(P_1'^c M | P_2')$ is equiangular
4) The four subspaces $P_2' P_1' M,\ P_2' P_1'^c M,\ P_2'^c P_1' M,\ P_2'^c P_1'^c M$ are mutually orthogonal.

With $M = P_m P_{m-1} \ldots P_2 M_1$ this gives a definition of the equiangularity of $(P_1, \ldots, P_m | P_1', P_2')$. If $(M | P_1', P_2')$ is equiangular, then the quantity

$$\| P_2' P_1' \Psi \|^2$$

where Ψ is a unit vector in M, is independent of Ψ. We denote it by $p(M|P'_1, P'_2)$ and $p(P_1, \ldots, P_m|P'_1, P'_2)$ respectively.

We have the following

Theorem: If $(P_1, \ldots, P_m|P'_1, P'_2)$ is equiangular, then

$$p(P_1, \ldots, P_m|P'_1, P'_2) = p(P_1, \ldots, P_m|P'_1) \cdot p(P_1, \ldots, P_m, P'_1|P'_2) \tag{1}$$

and

$$p(P_1, \ldots, P_m|P'_2) = p(P_1, \ldots, P_m|P'_1, P'_2) + p(P_1, \ldots, P_m|P'^c_1, P'_2) \tag{2}$$

The generalization to an arbitrary number of events in the outcome is straightforward.

5.13 A Principle of Equiangularity

The arguments in section 12 make it natural to assume that the transition probabilities in any real course of events can be calculated from a wave function even if it is not uniquely determined by the initial conditions. The systematic study of this independence of the exact form of the wave function lead to the concept of equiangular sequences of projections. We have also given arguments for the assumption that any course of events can be described by (a sufficiently number of) our basic events, and that thus the description by means of equiangular sequences of projection is generally applicable.

On the other hand, the condition of equiangularity puts strong restrictions on the possible choice of sequences of projections. Thus, if we build our theory on equiangular sequences, we automatically get a strong restriction on which courses of events are possible. Thus a course of events $e'_1, \ldots, e'_n$ is only possible if it is contained in a process $(e_1, \ldots, e_m|e'_1, \ldots, e'_n)$ corresponding to an equiangular sequence of projections. Similar conditions can in turn be prescribed for $e_1, \ldots, e_m$ and for the complete sequence $e_1, \ldots, e_m, e'_1, \ldots, e'_n$.

If we now introduce a "principle of equiangularity" stating that the only possible courses of events are those corresponding to equiangular sequences of projections, this principle will act in a way similar to that of an equation of motion in classical theories, whose principal purpose also is to restrict the possible courses of events. A restriction on the possible sequences of events is suitable in view of the arbitrariness problem discussed in chapter 1 and 3.

A simple example of (one-step) equiangularity is given by a careful discussion of scattering by considering wave packets (instead of plane waves).

In an incoming stream of particles a single particle can fit many wave functions and they *all* give the *same probability* for the particle, as scattered, to come out in a special direction (differential cross section). In the language of the present theory this is characterized by the term equiangularity.

5.14 Space Localizations as Basic Observables

Although we have proposed a strong restriction on the possible sequences of events by means of the principle of equiangularity, there still remains an arbitrariness of observables if they can be chosen as arbitrary projections.

We use the Heisenberg picture to represent our projection operators. However, e.g. a space localization event is always associated to an instant of time and it is at that time the corresponding operator has the simple representation as the projection on the set of all wave functions vanishing outside the actual space region. In the Heisenberg picture this operator is represented by its transformation according to the time-dependent Schrödinger equation to some other fixed time. This transformed operator has a very complicated not explicitly expressible form. It cannot be given any interpretation as a physical event at that time. Thus, what fixes the time of an event represented by a Heisenberg picture operator is that it has a special very simple form when transformed back to this time.

On the other hand, if we have no prescribed restriction on allowed operators, then the above mentioned transformed operator could in principle be used to represent an event at that time. This obviously leads to a complete arbitrariness on at what time an event has occurred. It is exactly this arbitrariness which has lead to the Schrödinger cat paradox. There, the initial state, which at the initial time has an obvious physical interpretation, is transformed to a later time where it corresponds to a physically meaningless state. (The projection operator projecting on this state is a physically meaningless observable). What ultimately solves the problem is the restriction of the operators describing observables to a very special simple class of operators having an obvious physical interpretation. Thus we must determine such a class.

The simplest possible operators are the projections corresponding to space localization. If any operators have a simple obvious interpretation, the space localizations must be among them. On the other hand, the arguments in section 11 has made it plausible that, due to the ultimate macroscopic character of any description of a physical system, they are sufficient if we use sufficiently many of them at several instants of time. E.g. the momentum and velocity of a particle can be determined by space localizations at two different times.

We propose the space localizations to be the basic observables of our theory. They are the simplest ones possible and they are sufficient for the description of arbitrary courses of events.

It is an advantage to have the basic formulation of a theory as simple as possible. The restriction to space localizations means that projection operators corresponding to simultaneous basic events always commute. This gives an important simplification in the handling of sequences of events, some of which may be simultaneous. The order between simultaneous events is then irrelevant.

The restriction to space localizations might seem rather impractical. Its main purpose is however to establish a general, simple and consistent theory. With this general theory as basis, we can define other suitable auxiliary quantities and derive relations between them. It is the first task of the development of the theory to introduce definitions of such important physical quantities as energy, linear mo-

mentum, angular momentum, etc. However, it is important that these auxiliary quantities are introduced by means of definitions and not as extra assumptions.

5.15 Confidence Estimates

There remains a problem to be solved before we have a consistent theory. As was pointed out in the introduction, a wave function initially localized to a finite space region, will generally spread out in space in such a way that it cannot be localized exactly to any finite region at another instant of time. If we restrict our projections describing basic observables to space localizations, we cannot expect the condition of equiangularity to hold exactly. We can however expect it to hold approximately with very high accuracy.

We shall say that a projection operator P is ε-equiangular with respect to the subspace M, if

$$\left| \|P\Psi'\|^2 - \|P\Psi''\|^2 \right| \leq \varepsilon \tag{1}$$

for any unit vectors Ψ', Ψ'' in M. We denote by $p(M|P)$ any of the numbers $\|P\Psi\|^2$ where Ψ is a unit vector in M. Thus $p(M|P)$ is an approximate quantity. It is only defined to the accuracy ε. If ε is very small, however, it has a practically well-defined value. This definition of ε-equiangularity can easily be generalized to arbitrary sequences $(P_1, \ldots, P_m|P'_1, \ldots, P'_n)$.

Events with probabilities greater than $1-\varepsilon$ can be considered as practically certain and events with probabilities less than ε can be considered as practically impossible if ε is very small. The use of ε means essentially that we replace exact localization statements by confidence estimates. If e.g. $(M|P)$ is ε-equiangular and $p(M|P) \geq 1-\varepsilon$, then

$$\|P\Psi\|^2 \geq 1 - 2\varepsilon$$

for any unit vector Ψ in M which is a confidence estimate for the localization of Ψ to the space region R associated to P if P is a basic event of space localization.

Thus, the considering of confidence estimates constitutes the mathematical technique, which finally makes the simple direct theory using successive space localizations possible.

5.16 Conclusion

We have thus obtained a theory in which the basic concepts are events (expressing space localizations) and probabilities. It describes general courses of events in the (objective) physical reality. It describes courses of events and their initial conditions in a unified way, (i.e. by the same kind of events). The events establish an immediate one-to-one correspondence between the theory and the reality. The probabilities also have an obvious interpretation (the ordinary frequency interpretation).

$p(e_1, \ldots, e_m|e'_1, \ldots, e'_n)$ is the probability for the course of events $e'_1, \ldots, e'_n$ under the given initial conditions $e_1, \ldots, e_m$.

Beside this conceptual, kinematical, structure we have a dynamical structure which establishes the mathematical representation of the events and probabilities and their relations. Thus, the events are connected to projection operators in the Heisenberg picture (here the underlying time-dependent Schrödinger equation comes in) and the probabilities $p(e_1, \ldots, e_m | e'_1, \ldots, e'_n)$ are only defined for a certain class of possible processes

$$(e_1, \ldots, e_m | e'_1, \ldots, e'_n)$$

namely those which correspond to ε-equiangular sequences of projections.

Part II

Axiomatic Formulation

Introduction to Part II

In the preceding chapter we have given a "derivation" of the confidence theory. It was based on

1) The assumption that ordinary quantum mechanics, consisting of the Schrödinger equation together with the primitive original statistical interpretation ("interpretation 1" of chapter 1), is essentially correct.

2) Some general "physico-logical" principles expressed in sections 5.4 and 5.5 (and reformulated in section 6.2 below).

Since the confidence theory is heuristically derived from the ordinary theory, the two theories should be essentially equivalent with respect to problems and questions which can be formulated and treated in both theories.

The derivation can only be heuristic since the starting points are not fully clear and are not completely formalized. One purpose of the confidence theory is to propose a completely formalized alternative to the ordinary theory. We shall give a brief description of an axiomatic formulation of the confidence theory in chapters 6 and 7, referring to (1) for technical details. For reference purpose, definitions and theorems which are cited form (1) are given the same numbering as in (1).

Having established an axiomatic formulation of the theory, the further development of the theory is a deductive and in principle a pure mathematical task. A first step on this way is taken in (1).

One central concept of the confidence theory (summarized in section 5.16 above) is that of equiangular sequences of projections. From this concept it is convenient to extract another concept, which we shall call stochastic event structure, expressing in a way suitable for our purposes the general rules of probabilities. This concept constitutes in a sense the concept of a general indeterministic physical theory. It can be motivated from very general arguments. It can also be considered as a general axiomatization of the concept of "causality" (See section 6.4, comment on the condition 3 e) of definition of stochastic event structure.)

Another central idea of the theory, that of confidence estimates and confidence levels, implies a limitation of (the accuracy of) the theory expressed within the theory itself. This shows the importance of the concept of approximation and leads to the general concept of what we shall call a self-limiting theory.

Thus one can extract a general part of the theory which is independent of the Hilbert space formalism of quantum mechanics. This general part is formulated in chapter 6, leaving the concepts and postulates specific for "Hilbert space" quantum theory to chapter 7.

Chapter 6

General Principles of Deductive Theoretical Physics

6.1 The Concept of a Deductive Physical Theory

We shall use the common term "mathematical structure" for a system of sets, mappings, etc. constituting the basis of a mathematical theory. The stipulation of a mathematical structure means a set theoretical way of formulating an axiomatic theory, which is convenient also for (axiomatic) physical theories for two reasons. Firstly, it is in concordance with notations and methods in pure mathematics and is thus the natural one to use if we want to apply the mathematical method to physics and, secondly, it emphasizes and makes definite the conceptual part of the theory.

We state the following general

Definition: A (deductive) physical theory is a mathematical theory built on a mathematical structure $\mathscr{T}$, together with a correspondence between a set $\mathscr{O}_{\mathscr{T}}$ within the structure $\mathscr{T}$ and a set $\mathscr{O}_{\mathscr{M}}$ of "physical observables" describing a certain well-defined part of the physical reality. This part of the physical reality, which is described by the theory, is called the scope of the theory.

The mathematical part of the theory consists (as any mathematical theory) of a set of definitions and proved theorems concerning the structure $\mathscr{T}$. That the theorems of the theory are rigorously proved is synonymous to saying that we use the "deductive method".

The physical part of the theory consists merely of the correspondence or "identification" of elements in the sets $\mathscr{O}_{\mathscr{T}}$ and $\mathscr{O}_{\mathscr{M}}$. In the next section we shall give some "principles" which all have the purpose of making this identification (the "interpretation" of the theory) unproblematic.

6.2 Principles of Interpretation

The following principles of interpretation are stated here in order to stipulate some general attitudes taken to (deductive) physical theories and are intended to serve as a general ground for the application of the theory.

1) **"Postulate of Objective Reality"**: The observables in $\mathcal{O}_{\mathcal{M}}$ describes in an objective way an objective physical reality. Thus we postulate the existence of an objective physical reality. We also restrict the theory to describe only this part of the reality. Subjective phenomena are outside the scope of the theory.

2) **"Principle of One-to-one Representation"**: The set $\mathcal{O}_{\mathcal{M}}$ of observables describing the physical reality is mapped in a one-to-one way on the set $\mathcal{O}_{\mathcal{T}}$ of "observables" in the theory $\mathcal{T}$.

3) **"Principle of Trivial Interpretation"**: The observables in $\mathcal{O}_{\mathcal{M}}$ shall have a completely obvious, indisputable meaning in the physical reality.

4) **"Principle of Complete Conceptual Axiomatization"**: The set $\mathcal{O}_{\mathcal{M}}$ shall provide a complete description of any possible physical situation within the scope of the theory. Any other physical quantities of interest should be defined in terms of the elements of $\mathcal{O}_{\mathcal{M}}$ and the corresponding elements of $\mathcal{O}_{\mathcal{T}}$ by means of mathematical definitions in the theory built on $\mathcal{T}$. Any description of a physical situation is thus ultimately reduced to and described by elements in $\mathcal{O}_{\mathcal{M}}$ ($\mathcal{O}_{\mathcal{T}}$).

We shall call the elements in $\mathcal{O}_{\mathcal{M}}$ and $\mathcal{O}_{\mathcal{T}}$ "basic observables". Other concepts, which are defined in the theory and which have the character of observables will be called "derived observables" or "derived quantities".

Beside the set $\mathcal{O}_{\mathcal{T}}$ in the structure $\mathcal{T}$ we have in $\mathcal{T}$ a set of relations between the elements in $\mathcal{O}_{\mathcal{T}}$ and ordinarily also some other auxiliary concepts in $\mathcal{T}$. These relations constitute the "basic physical laws" of the theory. We shall generally refer to them as the "dynamics" of the theory. These relations are very important, not only by giving the properties of a system considered, but also in connection with the interpretation of the theory. It is through them that the properties of the basic observables are manifested and it is thus ultimately through these relations that the basic observables gets their meaning. We state this in the following principle.

5) **"Principle of Complete Relational Axiomatization" or "Principle of Self-explanation"**: The structure $\mathcal{T}$ shall contain sufficiently many relations (laws) to completely give the meaning to the basic observables.

In any application of the theory we have to make some assumptions on the environment of the limited system under concern (the object system). These assumptions generally appear in the form of initial and boundary conditions, see section 6.4. For discussion of boundary conditions we refer to (1). Two important consequences of the preceding principles are the following.

1. Object system and environment are (ultimately) described by the same set of basic observables.

2. The relations between object and environment are given by the physical laws in $\mathcal{T}$.

A theory satisfying all the preceding principles will be called a "closed theory". One is often interested in studying a separate part of a theory which, when

formulated separately, does not satisfy principle 5) above. An example of this is the study of the functional relation between a finite number of physical quantities for a given system such as e.g. the volume, pressure and temperature of a gas. Such "open theories" can of course be handled deductively but should ultimately be embedded in a closed theory.

6.3 Events as Basic Observables

In the following we shall chose as our set of basic observables $\mathcal{O}_{\mathcal{S}}$ a set E_o of events. By an event we shall mean a formal statement about the physical reality of such a character that it can be either true (occur) or false (not occur) or undefined/irrelevant depending on the real situation. This means that if e is an event, then "not e", which we shall denote by $-e$, is also an event such that if e has occurred, then $-e$ has not occurred and vice versa. If e is irrelevant then $-e$ is also irrelevant.

6.4 Stochastic Event Structures

To every element ("event") e in the set E_o below, there is defined another element $-e$ in E_o (with the intended interpretation "not e"). We introduce the notation $\underline{e}$ to mean any of the two elements e or $-e$.

Definition II.1:1. By a stochastic event structure we shall mean a structure $\mathcal{S} = (E_o, -, T, p)$ such that

1) E_o is a set, whose elements will be called events. To every element e in E_o is associated another distinct element $-e$ in E_o such that $-(-e) = e$.

2) T is a set of ordered (m,n)-tuples $(e_1, \dots, e_m|e'_1, \dots, e'_n)$, $m, n = 1, 2, 3, \dots$ of events in E_o such that if $(e_1, \dots, e_m|e'_1, \dots, e'_n)$ is in T, then so is $(e_1, \dots, e_m|\underline{e}''_1, \dots, \underline{e}''_p)$ where $e''_1, \dots, e''_p$ is any subsequence of $e'_1, \dots, e'_n$ and $\underline{e}''_j$ equals e'_j or $-e'_j$ for every $j = 1, \dots, p$. The elements in T are called processes.

3) p is a function from T to the closed interval [0, 1] such that

   a) if $(e_1, \dots, e_m|e'_1, \dots, e'_n)$ is in T, then $\sum p(e_1, \dots, e_n|\underline{e}'_1, \dots, \underline{e}'_n) = 1$ where the sum is over all 2^n choices $\underline{e}'_j = e'_j$ or $-e'_j$, $j = 1, \dots, n$. (We write $p(e_1, \dots, e_m|e'_1, \dots, e'_n)$ instead of $p((e_1, \dots, e_m|e'_1, \dots, e'_n))$ for the value of the function p for the argument $(e_1, \dots, e_m|e'_1, \dots, e'_n)$ in T).

   b) if $(e_1, \dots, e_m|e'_1, \dots, e'_n)$ is in T, $i_1, \dots, i_p$ a subsequence of $1, \dots, n$ and $e''_1, \dots, e''_p$ the corresponding subsequence of $e'_1, \dots, e'_n$, then $p(e_1, \dots, e_m|e''_1, \dots, e''_p) = \sum p(e_1, \dots, e_m|e'_1, \dots, e'_n)$ where the sum is over all $\underline{e}'_1, \dots, \underline{e}'_n$ such that $\underline{e}'_k = e''_j$ if $k = i_j$, $j = 1, \dots, p$.

   c) $p(e_1, \dots, e_m|e'_1, \dots, e'_m) = 0$ if $e'_i = -e'_j$ for some i, j.

   d) if $(e_1, \dots, e_m|e'_1, \dots, e'_n, e''_1, \dots, e''_k)$ is in T and $p(e_1, \dots, e_m|e'_1, \dots, e'_n)$ nonzero, then

$(e_1,\dots,e_m,e'_1\dots,e'_n|e''_1,\dots,e''_k)$ is in T and
$p(e_1,\dots,e_m|e'_1,\dots,e'_n,e''_1,\dots,e''_k) =$
$= p(e_1,\dots,e_m|e'_1,\dots,e'_n)\cdot p(e_1,\dots,e_m,e'_1,\dots,e'_n|e''_1,\dots,e''_k)$.

e) if $(e_1,\dots,e_m|e'_1,\dots,e'_n)$ and $(e''_1,\dots,e''_k,e_1,\dots,e_m|e'_1,\dots,e'_n)$ are in T, then, $p(e''_1,\dots,e''_k,e_1,\dots,e_m|e'_1,\dots,e'_n) = p(e_1,\dots,e_m|e'_1,\dots,e'_n)$. (See (1).)

In a tuple of the form $(e_1,\dots,e_m|e'_1,\dots,e'_n)$, $(e_1,\dots,e_m)$ will be called the "premise", $(e'_1,\dots,e'_n)$ will be called the "outcome" and $p(e_1,\dots,e_m|e'_1,\dots,e'_n)$ will be called the "conditional probability" or shorter "p-value". It is sometimes necessary to consider tuples of this form, without assuming them to be elements in T. We shall call such tuples "formal processes". An ordered tuple $(e_1,\dots,e_n)$ of events will be called a "course of events". For a process $(e_1,\dots,e_m|e'_1,\dots,e'_n)$, the premise $(e_1,\dots,e_m)$ will, in certain discussions, alternatively be called the "initial conditions" of the course of events $(e'_1,\dots,e'_n)$. We shall in the following use the notation $E, E', E_1,\dots$ for ordered tuples (sequences) of events in E_o in order to get shorter notations. Thus, e.g. if $E = (e_1,\dots,e_m)$, $E' = (e'_1,\dots,e'_n)$, $E'' = (e''_1,\dots,e''_n)$ and e is in E_o, we shall use the notation $(E,e,E'|E'')$ instead of $(e_1,\dots,e_m,e,e'_1,\dots,e'_n|e''_1,\dots,e''_n)$ and so on. We write $E_1 \subset E_2$ if E_1 is a subsequence of E_2 . We denote by $\underline{E}$ any of the 2^n sequences $\underline{E} = (\underline{e}_1,\dots,\underline{e}_n)$ where $E = (e_1,\dots,e_n)$.

Except for condition 3e), the stochastic event structure expresses ordinary rules for conditional probabilities. Condition 3e) has another character. It expresses a kind of completeness of the premise. The probability $p(E|E')$ is not changed if we augment the premise with further events E'' previous to E i.e. $p(E'',E|E') = p(E|E')$. (E'' in (E'',E) is said to be "previous" to E since, as we shall see in the next section, the ordering of events in processes is to be connected with a time-ordering). This completeness gives a "character of physical causality" to the relation between premises and outcomes and to the corresponding probabilities $p(E|E')$ of a stochastic event structure. (Note that 3e) is the only condition in the definition of a stochastic event structure with direct reference to the ordering of the events.) Thus a stochastic event structure is, beside being an axiomatization of probability, also an axiomatization of the concept of "causality", generalizing the ordinary "deterministic" causality to what we shall call "stochastic causality".

The stochastic event structure constitutes a general framework for making predictions or inferences (both deterministic and stochastic) on events at certain instants of time from other events at preceding instants of time (initial conditions) and from other simultaneous events (boundary conditions). (Boundary conditions are discussed in (1)).

That the mathematical structure of our physical theory has the form of a stochastic event structure expresses the fact that the basic laws of nature (as considered by the present theory) has an irreducible element of indeterminism.

The ordering of events in a process $(e_1,\dots,e_m|e'_1,\dots,e'_n)$, will be connected with a time ordering (see next section). Thus there is an asymmetry between forward and backward time direction already in the basic concepts of the theory. This gives to irreversibility a status of fundamentality in the present formalism.

6.5 Time-ordering and Time

Definition II.3:24. By a time-ordered stochastic event structure we shall mean a structure
$\mathcal{S} = (E_o, -, T, p, |)$ where $(E_o, -, T, p)$ is a stochastic event structure and $|$ is a relation $e_1|e_2$ between elements e_1, e_2 in E_o (not supposed to be transitive!) such that

1) If $e_1|e_2$ then $\underline{e}_1|\underline{e}_2$

2) If $(e_1, \dots, e_m|e'_1, \dots, e'_n)$ is in T, then

   a) $e_i|e_j$ if $i < j$, $i, j = 1, \dots, m$

   b) $e'_i|e'_j$ if $i < j$, $i, j = 1, \dots, n$

   c) $e_i|e'_j$ for all $i = 1, \dots, m$, $j = 1, \dots, n$

3) If $(E_1|E_2)$ and $(E'_1|E'_2)$ are processes in T and E'_1 is a permutation of E_1 (contains the same events) and E'_2 is a permutation of E_2, then $p(E_1|E_2) = p(E'_1|E'_2)$ (See (1).)

A sequence $E = (e_1, \dots, e_m)$ of elements in E_o is called time-ordered if $e_i|e_j$ for every $i, j = 1, \dots, n$ with $i < j$.

For "instantaneous events" we have associated to the events e a value $t(e)$ of the time variable. $e_1|e_2$ then simply means $t(e_1) \leq t(e_2)$. A more general time ordering, suitable for relativistic theories, is defined in the next section.

6.6 Space-time Localization of Events

Definition II.4:31. By space-time we shall mean $\mathbb{R}^4$ considered as $\mathbb{R}^3 \times \mathbb{R}^1$. For a point $X = (x, t) = ((x_1, x_2, x_3), t) = (x_1, x_2, x_3, t)$ in space-time, $x = (x_1, x_2, x_3)$ will be called the space coordinates or components and t will be called the time component or simply the time. We define a time-ordering relation $|$ on $\mathbb{R}^4$ in two different ways.

a) "pure time-ordering": for $X' = (x', t')$ and $X'' = (x'', t'')$ in $\mathbb{R}^4$ we define $X'|X''$ to mean $t' \leq t''$

b) "relativistic time-ordering": for $X' = (x'_1, x'_2, x'_3, t')$ and $X'' = (x''_1, x''_2, x''_3, t'')$ in $\mathbb{R}^4$ we define $X'|X''$ to mean that not both
$(x'_1 - x''_1)^2 + (x'_2 - x''_2)^2 + (x'_3 - x''_3)^2 - (t' - t'')^2 \leq 0$
and $t' > t''$.

In both cases a) and b) we define $R'|R''$, where R' and R'' are subsets of $\mathbb{R}^4$, to mean that $X'|X''$ for any points X' in R' and X'' in R''. (See (1).)

Definition II.4:32. By a space-time localized stochastic event structure we shall man a structure
$(E_o, -, T, p, |, \subset)$ such that

1) $(E_o, -, T, p, |)$ is a time-ordered stochastic event structure

2) $\subset$ is a relation between elements e in E_o and regions R in $\mathbb{R}^4 = \mathbb{R}^3 \times \mathbb{R}^1$ such that

   a) $e \subset R$ implies $-e \subset R$

   b) $e \subset R \subset R'$ implies $e \subset R'$

   c) $e \subset R$ and $e \subset R'$ implies $e \subset R \cap R'$

3) $e_1|e_2$ iff $e_1 \subset R_1$ and $e_2 \subset R_2$ for some subsets R_1 and R_2 of $\mathbb{R}^4$ with $R_1|R_2$, where the latter |-relation is given by definition II.4:2 1 or 2. (See (1).)

There should be no risk of confusion in using the same symbol | for the relation $e_1|e_2$ and $R_1|R_2$. (By a "region" we shall generally mean a measurable set whose boundary has measure zero).

6.7 Localization of Particles as Basic Events

Let $X = \mathbb{R}^3$ denote the 3-dimensional configuration space. For a system of n distinguishable particles we define our basic events as pairs (Ω, t) where t is an instant of time and Ω is a region in the $3n$-dimensional configuration space X^n. The event $e = (\Omega, t)$ means that at time t, the n-particle system is localized to Ω. We define the operation – ("not e") by $-e = (X^n - \Omega, t)$ if $e = (\Omega, t)$. Thus $-e$ means localization to the complement region $X^n - \Omega$.

For quantum-mechanical particles, some of the n-particles may be identical and thus indistinguishable. We shall then restrict the region Ω to be symmetric in the corresponding coordinates i.e. if particles i and j are identical and $x = (\dots, x_i, \dots, x_j, \dots)$ is a point in Ω, then also $x' = (\dots, x_j, \dots, x_i, \dots)$ is in Ω.

For classical particles these events have a quite obvious meaning. The n-particle system has at every time t a well-defined configuration $x(t) = (x_1(t), \dots, x_n(t)) \in X^n$ where $x_i(t)$ is the position (in $\mathbb{R}^3$) of the i:th particle at time t. (We shall call $x(t)$ as a function of t the "orbit" of the system). $e = (\Omega, t)$ then means that $x(t) \in \Omega$.

For quantum-mechanical particles this point-interpretation is meaningless. Instead we consider the localization to space-regions as a fundamental, axiomatic, irreducible property of "quantum-mechanical particles". For a quantum-mechanical n-particle system, the localization to symmetric region Ω in X^n is a fundamental, irreducible property of the n-particle system. It cannot be reduced to one-particle statements.

This irreducibility is an expression for the "non-classical" properties of quantum systems and an expression for "the indivisible unity of quantum systems" (see Bohm (2)). Having once freed ourselves from the classical point ideas, the meaning of these localization statements should be quite obvious and immediate. The localization of a physical entity to a region in space is perhaps the most basic of our everyday experiences and its generalization to localization to regions in X^n is plain. As an example showing the immediate meaning of many-particle localizations we consider an electron (particle 1) and a proton (particle 2) confined to the 3-dimensional region Ω. Then the region $\Omega_2 \subset X^2 = X \times X$ defined by

$\Omega_2 = \{(x_1, x_2);\, x_1 \in \Omega,\, x_2 \in \Omega,\, |x_1 - x_2| \leq r\}$ expresses that the particles are confined to the region $\Omega \subset X$ and that the electron is bound to the proton (in some way!) to form an "atom" of radius less than r.

The preceding definition of events presupposes a given total number of particles. In order to make statements on the mass-distribution in a region Ω' of X irrespectively of the situation outside Ω' we shall generalize the formalism to variable number of particles. Consider for simplicity the case of only one kind of quantum-mechanical particle. (The generalization to several kinds of particles is straightforward). A basic event e is then defined as

$$e = (\Omega, t)$$

where

$$\Omega = \{\Omega_n\}_{n=0}^{\infty}$$

is a sequence of regions

$$\Omega_0,\, \Omega_1,\, \Omega_2, \ldots$$

and where Ω_n is a symmetric region in X^n. It has the meaning: "for every n: if the total number of particles is n, then these n particles are localized to Ω_n". We make the convention that $\Omega_n = \emptyset$ means that the total number of particles is not n. One can then express statements like "at time t, the total number of particles in the 3-dimensional region Ω' is m" or lies in the interval $[m_1, m_2]$ or "at time t, the center of mass of the particles in Ω' lies in Ω'', (Ω', Ω'' 3-dimensional regions) etc.

The two mentioned events are examples of events e localized to the region Ω' in X and we shall write $e \subset \Omega'$ in this case. Thus our structure for localizations of many-particle systems implies a structure with momentaneous space localization of events (to $X = \mathbb{R}^3$). In a relativistic theory we shall define E_o as the set of all events e which are of the form (Ω, t) in some system of reference (inertial system), which can be different for different e. The relation of instantaneous space localization of events then defines in an obvious way a relation of localization of events to space-time regions and a relativistic time-ordering of events. An event $e = (\Omega, t)$, $\Omega = \{\Omega_n\}_{n=0}^{\infty}$, localized to the 3-dimensional space region Ω' i.e. $e \subset \Omega'$, is then localized to the space-time region Ω'' in $\mathbb{R}^4$ if $\Omega' \times \{t\} \subset \Omega''$. We shall write $e \subset \Omega''$ in this case. (There should be no risk for confusion in using the same relation notation $\subset$ for $e \subset \Omega'$ with $\Omega' \subset \mathbb{R}^3$ and $e \subset \Omega''$ with Ω'' in space-time $\mathbb{R}^4$

6.8 Derived Events

We shall now discuss how the set E_o of basic events in a stochastic event structure can be extended to a set E_o' of "derived events". We are sometimes interested to consider not a single course of basic events $E = (e_1, \ldots, e_n)$ but some more general aspect of the physical situation satisfied by a set $\{E_1, \ldots, E_n\}$ of such sequences. This corresponds to a coarse statement about the physical situation. We shall call

such a statement a "derived event". It can be expressed as a logical disjunction

$$e = E_1 \vee E_2 \ldots \vee E_n = \bigvee_{j=1}^{n} E_j$$

of the sequences $E_1, E_2, \ldots$ of basic events in the set, stating that E_1 or E_2 or … or E_n has occurred.

When using such disjunctions we have to take into account the following "logical tautology". (e_1, e_2) means that e_1 and e_2 has occurred (i.e. a logical conjunction $e_1 \wedge e_2$). Then obviously the statement "(e_1, e_2) or $(-e_1, e_2)$" must be equivalent to just stating e_2. More generally, if $E' = (e'_1, \ldots, e'_p)$ is a subsequence of $E = (e_1, \ldots, e_n)$ then a set of sequences of the form $\underline{E}'$ is equivalent in this way to some set of sequences of the form $\underline{E}$.

Let $(E_o, -, T, p)$ be a given stochastic event structure. If E is the outcome of some process in T, then every set of sequences of the form $\underline{E}$ defines a derived event. We define the set E'_o as the set of all such derived events with the above mentioned equivalence taken into account. If $(E'|E)$ is a process in T and $E_1 = \{\underline{E}\}' \in E'_o$ is a derived event, defined by means of the events in E, then the tuple $(E'|E_1)$ will be called a derived process and we define

$$p(E'|E_1) = \sum p(E'|\underline{E})$$

where the sum is over the $\underline{E}$:s in E_1. We shall denote by T' the set of all such derived processes.

This logic of events is completely in concordance with ordinary "classical" logic, which is thus the general and the same basic logic for both classsical and quantum physics. (All considerations of "quantum logic" are left outside.)

We call the events in E'_o above "logically derived events". Some of these logical extensions of the set of events can be used to define what we shall call "dynamically derived events". For instance, linear momentum observables can, by using the dynamics and considering the propagation of wave packets, ultimately be reduced to successive space localizations. All interesting dynamical quantities, i.e. linear momentum, angular momentum, energy, electromagnetic fields, etc., can thus ultimately be reduced to an underlying stochastic event structure of events of space localizations.

The restriction to processes fulfilling this condition of ultimate reducibility to stochastic event structures of space localizations is the basic principle of the present theory. The basic motivation of this ultimate reducibility is that any, arbitrary complicated, experiment and observation of quantum phenomena is always coupled to phenomena observable in our everyday macroscopic world. Thus quantum physics is considered as a generalization of the dynamical rules of classical physics described ultimately by the same observables of space localizations – not as "another kind of physics".

6.9 Dynamics

Having identified the set $\mathcal{O}_\mathcal{F}$ of basic observables with the set E_o of events of a stochastic event structure $(E_o, -, T, p)$, we shall now discuss the dynamics of the theory. Before we do this, we shall make a comment about the interpretation of probabilities. One might think that also the probabilities, beside the events, should be considered as basic observables with a well-defined, obvious interpretation namely the frequency interpretation. It is of course quite possible to do so. However, another attitude is possible. The frequency interpretation can be reduced to the identification of events with probabilities greater than a given level of confidence $1-\varepsilon$ as certain events. Then our probabilities and the corresponding "statistical inferences" are reduced to "deterministic inferences". The probabilities, still being physical observables and still having the frequency interpretation are then considered as derived observables (defined by means of the p:s in the stochastic event structure). Thus, despite the simplicity of the frequency interpretation, we prefer, by an argument of "economy of assumptions", this latter attitude.

It is important to note that this presupposes that we have equipped with the theory a given level of confidence $1-\varepsilon$. This will be further discussed in the next section.

As a consequence of this attitude we consider the p-values in first hand, i.e. when formulating the basic axiomatics of the theory, formally as auxiliary dynamical quantities.

We now state the basic task of dynamics as follows.

1) To determine which courses of events $e_1, \dots, e_n$ are possible i.e. can occur in reality.

2) To determine which deterministic inferences are true i.e. in which cases we can infer that $e_1, \dots, e_n$ (or more generally a derived event $E' \in E'_o$) is certain to occur provided that we know that some other course of events has occurred.

If our theory is given in the form of a stochastic event structure, then question 1) will not be answered in the "absolute" sense but only relative to a given premise. This is an expression for the fact that the theory is a local rather than a global, cosmological theory. In any application of the theory, we have to make certain assumptions on the initial conditions. We then ultimately have to assume that the sequence of events describing these initial conditions is possible, without trying to prove this from the theory.

Question 1) is then answered by the statement that $e_1, \dots, e_n$ is possible relative to the premise $e'_1, \dots, e'_m$, if it is outcome in the process $(e'_1, \dots, e'_m | e_1, \dots, e_n)$ in T with significantly positive p-value i.e. with $p(e'_1, \dots, e'_m | e_1, \dots, e_n) \gg \varepsilon$.

Question 2) is in first hand answered by assigning those processes $e'_1, \dots, e'_m | e_1, \dots, e_n$ in T and more generally derived processes $(e'_1, \dots, e'_m | E)$ in T' $(E \in E'_o)$, which have p-value greater than the confidence level $1-\varepsilon$.

It is important to note, however, that we can generally infer $(e_1, \dots, e_n)$ from $(e'_1, \dots, e'_m)$ when $p(e'_1, \dots, e'_m | e_1, \dots, e_n) > 1-\varepsilon$ only provided we know or assume that $e_1, \dots, e_n$ is a relevant sequence of events for the actual situation. Within

a general stochastic event structure we cannot generally infer from the fact that $p(e'_1,\dots,e'_m|e_1,\dots,e_n) > 1-\varepsilon$ that $e_1,\dots,e_n$ is a relevant sequence of events under the assumption . There might be other possible sequences $(e''_1,\dots,e''_p)$ which can occur as outcomes of the premise $(e'_1,\dots,e'_m)$ but which are incompatible with $(e_1,\dots,e_n)$. We shall say that a set of events is compatible if it is contained in the outcome of some process. $(e_1,\dots,e_n)$ and $(e''_1,\dots,e''_p)$ are said to be compatible if they are subsequences of a common sequence which is outcome in some process. Thus our deterministic, and generally also our stochastic, inferences generally has to be studied within some given set of compatible events. We refer to (1) for further discussion of compatibility.

There is, however, another stronger kind of deterministic inference. The dynamics restricted by the set T generally puts very strong restrictions on which sequences of events that are possible as outcomes of given initial and boundary conditions. These restrictions may then be so strong that the above mentioned ambiguity in choice between mutually incompatible outcomes more er less disappears. In this case, the initial conditions determines which outcome events are relevant, and in the case of p-value $1-\varepsilon$, we can unconditionally infer the outcome.

The fact that not all events are compatible under a given premise means that only a limited set of events are relevant for the actual situation. This means that all other events are neither true nor false but just irrelevant and undefined. This is a characteristic property of quantum theory which does not occur in classical theories.

That not all events are compatible – simultaneous relevant – might seem to be an apparent quantum characteristic when seen on the microscopic level. There it is connected to what is characterized by such notions as complementarity, the uncertainty relation etc. which stands in a manifest contrast to the classical concept of event as point in (more general subset of) a given state space of a given system.

However, compatibility/incompatibility is a general, basic notion in the theory of stochastic event structures. It is, in fact, also a common experience in our everyday macroscopic world. This also shows the deep connection between "macroscopic physics" and "quantum phenomena" as proposed by the present theory.

Having established the probabilities as physical quantities, we can add as the third basic task of dynamics the following.

3) To determine the probabilities for those possible courses of events for which deterministic inference is not possible.

The T and p of a stochastic event structure constitute what we shall call general or abstract dynamics. Several general concepts and questions can be handled within a general stochastic event structure without referring to the detailed specific dynamics of the theory. To get a complete theory we must of course add to this basic structure extra rules which determine the set T and the values of p. Let us mention briefly four examples of such theories.

Example 1: Classical (Newtonian) n-particle mechanics

The set E_o is defined as the set of events of the form (Ω,t), $\Omega \subset X^n$, as defined in the beginning of section 6.7 above.

The dynamics is given in the form of equations of motion determining which orbits $x(t)$ are possible. An event $e = (\Omega, t_1)$ has occurred for a given orbit $x(t)$ if $x(t_1) \in \Omega$. We define the set T_1 as the set of all (r,s)-tuples $(e_1, \ldots, e_r | e'_1, \ldots, e'_s)$ of events $e_i = (\Omega_i, t_i)$, $i = 1, \ldots, r$, $e'_i = (\Omega'_i, t'_i)$, $i = 1, \ldots, s$ such that

1) $t_1 \leq \ldots \leq t_r \leq t'_1 \leq \ldots \leq t'_s$

2) All $e'_1, \ldots, e'_s$ has occurred for every orbit for which all $e_1, \ldots, e_r$ has occurred and define $p(e_1, \ldots, e_r | e'_1, \ldots, e'_s) = 1$ for the tuples in T_1. We then define T as the set of all tuples $(e_1, \ldots, e_r | e'_1, \ldots, e'_s)$ with $(e_1, \ldots, e_r | e'_1, \ldots, e'_s)$ in T_1 and define $p(e_1, \ldots, e_r | e'_1, \ldots, e'_s) = 0$ for the tuples in T not in T_1. Then $(E_o, -, T, p)$, is a "deterministic" stochastic event structure or what we call a "deterministic event structure". Obviously a knowledge of the structure $(E_o, -, T, p)$ means a complete knowledge of the dynamics of the system.

Example 2: A classical field theory described by a hyperbolic system of differential equations

Let $f(x,t)$ be a scalar or vector-valued field amplitude which as function of the space variable $x \in \mathbb{R}^3$ and time t satisfies some linear hyperbolic field equation such as e.g. the acoustic wave equation or Maxwell's equations. If we suppose f to be a continuous function of x and t, then the field in a finite space region $\Omega \subset \mathbb{R}^3$ at time t can be described approximately with practically sufficient accuracy by using a finite sequence of events e_i, $i = 1, \ldots, n$ defined by statements of the form $|f(x_i, t) - a_i| < \delta_i$ where $x_i \in \Omega$. Now the values of the field at one time t is uniquely determined by the field and its time derivatives at another previous time $t' < t$ within a space region Ω' which covers the backward light cone or influence cone projection of Ω at t'. Instead of time derivatives we can use the field itself at two different times t' and t''. The field at time t' and t'' can then be described by a similar sequence of events $e'_1, \ldots, e'_m$. We define $(e'_1, \ldots, e'_m | e_1, \ldots, e_n) \in T_1$ and $p(e'_1, \ldots, e'_m | e_1, \ldots, e_n) = 1$ if $e_1, \ldots, e_n$ are determined from $e'_1, \ldots, e'_n$ by the field equations. We can then extend T_1 to a T in a way similar to the previous example to get a "deterministic" stochastic event structure.

The approximative approach on evolution equations in example 1 and 2 is natural both from a conceptual and technical point of view since it emphasizes the stability properties of the equation of motion.

Example 3: Non-relativistic quantum mechanics with electrostatic interaction

Example 4: Relativistic quantum electrodynamics

The set E_o for these examples was defined in section 6.7. The dynamics of these theories will be discussed in the next chapter.

6.10 Self-limiting Theories

Theoretical physics comprises a lot of different theories. Some of them are of a very general character such as examples 1–4 of the preceding section, but most of

them are very special theories concerning very special systems or phenomena. The ultimate goal of deductive theoretical physics is not only to describe and handle the special theories in a deductive way but also to use rigorous deductive methods in the discussion of the relations between theories and in the derivation of one theory from another as an approximation.

If we want to derive rigorously a theory $\mathscr{T}_2$ from another theory $\mathscr{T}_1$, we have to derive, not only the theory $\mathscr{T}_2$ itself, but also rigorous estimates on the accuracy of the quantities and relations in the theory $\mathscr{T}_2$ and rigorous conditions for the validity of the theory $\mathscr{T}_2$.

Deriving e.g. classical mechanics rigorously from quantum mechanics, we obtain together with the classical differential equations of motion in some form also certain limits on the accuracy due to the uncertainty relations. Using the confidence theory formalism based on sequences of events, this derivation leads in a natural way to the formulation of the classical equations in the form of what we shall call approximative differential equations, which are differential inequalities rather than exact differential equations. This gives a formulation of classical mechanics in the form of a theory containing within itself, in closed form, its own limitations due to quantum mechanics.

This leads us to consider theories which contain formally within themselves, at least in some respects, their own limits. We shall call such theories "self-limiting theories".

In the confidence theory formalism, this formulation of classical mechanics as a self-limiting theory comes out as special instances of the general courses of events determined by a quantum-mechanical stochastic event structure. Thus classical mechanics in this form is a part of (or a special case of) the more general quantum-mechanical theory. This way of looking means a considerable conceptual unification of theories.

The occurrence of a given confidence level $1 - \varepsilon$ with $\varepsilon > 0$ in the basic formulation of the theory implies a general limitation of the theory. Thus quantum mechanics as a whole is, in our formulation, a self-limiting theory.

We have seen how classical mechanics as a self-limiting theory comes out as a special case of the more general quantum-mechanical theory. This more general theory, however, does not reduce this self-limiting of the theory. We still have in quantum mechanics the same limits concerning particle motion caused by the uncertainty relations. Thus, in this case, the self-limiting of the theory has an absolute character.

It is not easy to see how the self-limiting quantum-mechanical theory in turn could come out as a special case of some other more general theory. In any case this is not necessary. We can and shall use the confidence theory as it stands as an "absolute" self-limiting theory irrespectively of its relation to any other more general theory. Thus, with a suitable small ε we can, with sufficient accuracy, handle a large number of physical situations and the set of all these situations constitutes the scope of the present theory.

Chapter 7

Hilbert Space Quantum Dynamics

7.1 Quantum-Mechanical Stochastic Event Structures

We shall now show how one can express quantum mechanics as a dynamics on a stochastic event structure. We shall then complement the general structure $(E_o, -, T, p)$ of a stochastic event structure by extra mathematical structural elements and rules having principally no other purpose than to specify the set T and the values of p. The conceptual or "kinematical" part of our quantum-mechanical theory is already given by specifying E_o to be the set of space localization statements for quantum-mechanical particles defined in section 6.7 above.

The specific dynamics of quantum mechanics is determined by the ordinary mathematical formalism of wave functions satisfying a Schrödinger equation (a Schrödinger-Schwinger-Tomonaga equation in the relativistic case). The connection between this formalism and the stochastic event structure will be made in three steps. First we establish a representation of a stochastic event structure by means of (orthogonal) projections in an abstract Hilbert space. Then we introduce the concept of configuration space wave functions and a corresponding representation in terms of subspaces of such wave functions of the abstract events of space localization defined in section 6.7. Finally, we connect the instantaneous events at different times by means of the Heisenberg picture defined by the Schrödinger equation. In the following we assume $\hbar = 1$.

7.2 Equiangular Sequences of Projections

In the following we shall denote by $P, P_1, P_2, \ldots$ orthogonal projections in a given Hilbert space $\mathscr{H}$. $M, M_1, \ldots$ will denote closed subspaces. Subspaces, not necessarily closed, are denoted by $N, N_1, \ldots$. $P^c = 1 - P$ denotes the projection on the orthogonal complement of the subspace on which P projects. We introduce the notation $\underline{P}$ to mean any of P or P^c. $R(P)$, the range of P, is the (closed) subspace on which P projects.

Definition III.1:1. Let P be a projection and N be a nonzero (not necessarily closed) subspace. P is called equiangular with respect to N (or N-equiangular) if

$$\frac{\|Pu\|}{\|u\|} = \frac{\|Pu'\|}{\|u'\|}$$

for any nonzero u, u' in N, and the common value of $\|Pu\|^2/\|u\|^2$ is denoted by $p(N|P)$. If $N = 0$, any P is N-equiangular and we define $p(N|P) = 0$. (See (1).)

Definition III.2:15. We say that the ordered sequence $(P_1, \ldots, P_n)$ is N-equiangular if

1) For every integer $i = 1, \ldots, n$, P_i is equiangular with respect to all the 2^{i-1} subspaces $\underline{P}_{i-1}\underline{P}_{i-2} \cdots \underline{P}_1 N$ (where $\underline{P}_j$ is either P_j or P_j^c). If $i = 1$, this shall mean that P_1 is N-equiangular.

2) All the 2^n subspaces $\underline{P}_n\underline{P}_{n-1} \cdots \underline{P}_1 N$ are orthogonal.

If $N = R(P)$ we shall also say that $(P_1, \ldots, P_n)$ is P-equiangular. (See (1).)

Theorem III.2:17. Suppose that $(P_1, \ldots, P_n)$ is N-equiangular, N nonzero, and let $(\underline{P}_1, \ldots, \underline{P}_n)$ be any sequence with $\underline{P}_j = P_j$ or P_j^c, $j = 1, \ldots, n$. Then, for any integer $i = 1, \ldots, n$, the quantity $\|\underline{P}_i\underline{P}_{i-1} \ldots \underline{P}_1 u\|/\|u\|$, u nonzero vector in N, is independent of u. (See (1).)

Definition III.2:18. If $P_1, \ldots, P_n$ is N-equiangular and u a nonzero vector in N, we denote the number $\|P_n P_{n-1} \ldots P_1 u\|^2/\|u\|^2$, by $p(N|P_1, \ldots, P_n)$. If $N = 0$, we define $p(N|P_1, \ldots, P_n) = 0$. If $N = R(P)$ we shall also denote this number by $p(P|P_1, \ldots, P_n)$. (See (1).)

Definition III.3:28. We shall say that $(P_1', \ldots, P_n')$ is equiangular with respect to $(P_1, \ldots, P_m)$ (or $(P_1, \ldots, P_m)$-equiangular) if $(P_1', \ldots, P_n')$ is equiangular with respect to $N = P_m P_{m-1} \ldots P_2 R(P_1)$. In this case we then also say that the (m, n)-tuple $(P_1, \ldots, P_m|P_1', \ldots, P_n')$ is equiangular and define
$p(P_1, \ldots, P_m|P_1', \ldots, P_n') = p(N|P_1', \ldots, P_n')$. (See (1).)

7.3 Stochastic Event Structures of Projections

Theorem III.4:32. Suppose E is a set of projections, closed under the operation of orthogonal complement (i.e. P in E implies P^c in E). Let T be a set of (m, n)-tuples $(P_1, \ldots, P_m|P_1', \ldots, P_n')$, $m, n = 1, 2, \ldots$ such that

a) all $P_1, \ldots, P_m, P_1', \ldots, P_n'$ are in E

b) $(P_1, \ldots, P_m|P_1', \ldots, P_n')$ is equiangular

c) $P_m P_{m-1} \ldots P_2 R(P_1)$ is nonzero.

Suppose also that the set T has the properties

d) if $(P_1, \ldots, P_m|P_1', \ldots, P_n')$ is in T, then so is $(P_1, \ldots, P_m|\underline{P}_1'', \ldots, \underline{P}_r'')$ where $P_1'', \ldots, P_r''$ is any subsequence of $P_1', \ldots, P_n'$.

e) if $(P_1, \ldots, P_m | P'_1, \ldots, P'_n, P''_1, \ldots, P''_k)$ is in T and $p(P_1, \ldots, P_m)|(P'_1, \ldots, P'_n)$ is nonzero, then $(P_1, \ldots, P_m, P'_1, \ldots, P'_n | P''_1, \ldots, P''_k)$ is in T.

Let $p : T \mapsto [0,1]$ be defined as in definition III.3:28 and let $-P$ stand for the operation of orthogonal complement $-P = P^c$. Then $(E, -, T, p)$ is a stochastic event structure. (See (1).)

Theorem III.4:33. Let E be a set of projections, closed under the operation of orthogonal complement. Let $|$ be a relation on E such that

1) if $P_1|P_2$, then $\underline{P}_1|\underline{P}_2$

2) if $P_1|P_2$ and $P_2|P_1$, then P_1 and P_2 commute.

Let T be the set of all (m,n)-tuples of projections in E satisfying conditions a) b) and c) of theorem III.4:1 and the condition

3) if $(P_1, \ldots, P_m | P'_1, \ldots, P'_n)$ is in T, then

   a) $P_i|P_j$ if $i < j$, $i, j = 1, \ldots, m$

   b) $P'_i|P'_j$ if $i < j$, $i, j = 1, \ldots, n$

   c) $P_i|P'_j$ for all $i = 1, \ldots, m$, $j = 1, \ldots, n$.

Define $-$ and p as in theorem III.4:32. Then $(E, -, T, p)$ is a time-ordered stochastic event structure. (See (1).)

7.4 Confidence Levels and approximate equiangularity

The special projections representing events to be defined in the following sections cannot in general be supposed to constitute sequences satisfying the condition of equiangularity exactly. However, we can assume that they build sequences satisfying the equiangularity conditions approximately with a high level of accuracy. Thus we have to generalize the results of the two preceding sections to the concept of approximate equiangularity (ε-equiangularity) associated to a level of confidence $1 - \varepsilon$.

Definition III.6:39. P_2 is called ε-equiangular with respect to N_1 if

$$\left| \|P_2 u'\|^2 - \|P_2 u''\|^2 \right| \leq \varepsilon$$

for any u', u'' in N_1 with $\|u'\| = \|u''\| = 1$.
If P_2 is ε-equiangular with respect to N_1, we denote by $p(N_1|P_2)$ any of the values of $\|P_2 u\|^2$, u in N_1, $\|u\| = 1$.
If $N_1 = 0$ we define any P_2 to be ε-equiangular with respect to N_1 and we define $p(N_1|P_2) = 0$. If $N_1 = R(P_1)$ we also say that P_2 is ε-equiangular with respect to P_1 and that $(P_1|P_2)$ is ε-equiangular and write $p(P_1|P_2)$ for $p(N_1|P_2)$. (See (1))

Note that $p(N_1|P_2) = p(P_1|P_2)$ is here an approximately defined quantity. It is only defined to the accuracy ε. For $\varepsilon = 0$, the concepts of ε-equiangularity and equiangularity (0-equiangularity) (definition III.1:1) coincide.

For the generalization to (m,n)-tuples $(P_1,\dots,P_m|P'_1,\dots,P'_n)$ and for further discussion we refer to (1.)

7.5 Wave Functions and Space Localization

As in section 6.7 above we first consider a fixed number n of particles. We also assume for simplicity that all particles are of the same kind, having mass m and spin s, the generalization to several kinds of particles being straightforward.

A space localization wave function is always defined relative to a certain inertial system. By a "n-particle (configuration space), Schrödinger picture, space localization wave function" we shall mean a complex valued amplitude

$$\psi(x_1,\alpha_1,x_2,\alpha_2,\dots,x_n,\alpha_n,t) \tag{1}$$

where t is the time variable (in the given inertial system in the relativistic case). $x_i \in X = \mathbb{R}^3$ is the position variable for the i:th particle and α_i is a spin index for the i:th particle. ψ is supposed to be symmetric or anti-symmetric with respect to interchange of particles depending on if the particles are bosons or fermions.

For fixed t, ψ, as a function of the x_i:s and α_i:s, is a vector $\Psi(t)$ in a Hilbert space $\mathscr{H}_n$ with scalar product defined by

$$\langle \Psi_1(t),\Psi_2(t)\rangle = \sum_{\alpha_i} \int \psi_1(x_1,\alpha_1,\dots,t)\overline{\psi_2(x_1,\alpha_1,\dots,t)}\, d^3x_1 \dots d^3x_n$$

$\Psi(t)$ is usually assumed to be normalized so that

$$\|\Psi(t)\|^2 = \langle \Psi(t),\Psi(t)\rangle = 1.$$

Note that these "spin" indexes are introduced here quite formally as auxiliary dynamical quantities. The introduction of spin as a derived physical observable then has to be done according to the principles of sections 6.2 and 6.8.

To every symmetric region Ω_n in X^n there corresponds a closed subspace $M(\Omega_n)$ of $\mathscr{H}_n$ defined by those functions $u(x_1,\alpha_1,\dots,x_n,\alpha_n)$ in $\mathscr{H}_n$ which vanish (almost everywhere) outside Ω_n i.e. when $x=(x_1,\dots,x_n)$ is outside Ω_n. To the closed subspace $M(\Omega_n)$ corresponds a projection operator which we shall denote by $P(\Omega_n)$.

This can be generalized to variable number of particles as in 6.7. To every sequence

$$\Omega = \{\Omega_n\}_{n=0}^{\infty}$$

of symmetric regions Ω_n in X^n, $n=0,1,2,\dots$ there corresponds a closed subspace

$$M(\Omega) = \bigoplus_{n=0}^{\infty} M(\Omega_n)$$

in the Fock space

$$\mathscr{H}_\infty = \bigoplus_{n=0}^{\infty} \mathscr{H}_n$$

where $\bigoplus$ denotes direct sum of Hilbert spaces. The corresponding projection operator in $\mathscr{H}_\infty$ is denoted by $P(\Omega)$. For $n = 0$, $\mathscr{H}_n$ is supposed to be a one-dimensional Hilbert space and $X^0 = \{x_0\}$ is a set containing one "dummy" element x_0. The "regions" $\{x_0\}$ and $\varnothing$ in X^0 then correspond to the two subspaces $\mathscr{H}_0$ and $\mathscr{H}_\infty - \mathscr{H}_0$ respectively.

7.6 Non-relativistic Quantum Dynamics Determined by the Schrödinger Equation

For non-relativistic electrostatic interactions, a connection between the instantaneous n-particle wave functions $\Psi(t)$ at different times t, is given by the Schrödinger equation

$$i\frac{d\Psi_n(t)}{dt} = H_n\Psi_n(t) \tag{1}$$

where H_n is the ordinary n-particle Schrödinger operator (with electrostatic potential between the charges q_i

$$H_n = -\sum_i \frac{1}{2m}\Delta_i + \sum_{i<j}\frac{q_i q_j}{|x_i - x_j|}$$

and where Δ_i is the laplacian operator acting on the i:th particle variables x_i. The Hamiltonian operator H_n defines a unitary operator

$$U_n(t_1, t_2) = e^{-i(t_2-t_1)H_n} \tag{2}$$

in $\mathscr{H}_n$ such that

$$\Psi_n(t_2) = U_n(t_1, t_2)\Psi_n(t_1) \tag{3}$$

The set of all equations (1) for $n = 0, 1, 2, \ldots$ (with the definition $H_0 = 0$) then generates a corresponding Schrödinger equation

$$i\frac{d\Psi(t)}{dt} = H\Psi(t) \tag{4}$$

for the time-dependent vector $\Psi(t)$ in the Fock space $\mathscr{H}_\infty$ and a corresponding unitary operator

$$U(t_1, t_2) = e^{-(t_2-t_1)H}$$

in $\mathscr{H}_\infty$, satisfying

$$\Psi(t_2) = U(t_1, t_2)\Psi(t_1) \tag{5}$$

for every $\Psi(t)$ satisfying (4).

The existence of the unitary mapping U with the property (5) shows that the set of all time-dependent vectors $\Psi(t)$ in $\mathscr{H}_\infty$ satisfying the Schrödinger equation (4) can be considered as a Hilbert space $\mathscr{H}'_\infty$ with scalar product defined by

$$\langle \Psi_1, \Psi_2 \rangle = \langle \Psi_1(t_1), \Psi_2(t_1) \rangle \quad \Psi_1, \Psi_2 \in \mathscr{H}'_\infty$$

where the right member (which is a scalar product in $\mathscr{H}_\infty$) is independent of the choice of t_1. The vectors Ψ in $\mathscr{H}'_\infty$ can be represented by the values of $\Psi(t)$ ($\in \mathscr{H}_\infty$) for some fixed value of t e.g. $t = 0$. This makes $\mathscr{H}'_\infty$ isomorphic to $\mathscr{H}_\infty$. Operators A in $\mathscr{H}_\infty$ operating on $\Psi(t)$ at time t can be represented by operators A' in $\mathscr{H}'_\infty$ by

$$A' = U(0,t)^{-1} A U(0,t)$$

acting on $\Psi(0)$, since $U(0,t)^{-1}(A\Psi(t)) = A'\Psi(0)$. This is the usual Heisenberg picture.

We can now define a representation R_0 of the set E_0, defined in section 6.7, of particle localization statements as a set $\mathscr{P}$ of projection operators in $\mathscr{H}'_\infty$ i.e. a mapping

$$R_0 : E_0 \mapsto \mathscr{P}.$$

Every element e in E_0 is of the form (Ω, t) with $\Omega = \{\Omega_n\}_{n=0}^{\infty}$, and we define

$$R_0(e) = U(0,t)^{-1} P(\Omega) U(0,t) \tag{6}$$

where $P(\Omega)$ was defined in the preceding section.

Putting the constructions in sections 7.2 – 7.5 together we can now finally express our quantum-mechanical theory as the structure

$$\mathscr{T} = (E_0, -, T, p, |, R_0, \varepsilon) \tag{7}$$

which is a time-ordered stochastic event structure complemented with the representation R_0 of events by projection operators, together with the condition that this shall constitute a structure of ε-approximate equiangular sequences ("principle of equiangularity").

7.7 Covariant Wave Functions

Since the proposed theory is based on space localizations of particles, a comment on how this is treated in the case of relativistic dynamics is necessary. Space localization is stepmotherly left undiscussed, in texts on quantum field theory where the treatment is limited to scattering of plane waves. This might seem sufficient for calculating the S-matrix in high energy particle physics. The proposed theory aims at a general theory of classical and quantum phenomena in finite space and time and, in due cases, take relativistic effects into consideration. A proper discussion of finite space localizations needs a discussion of wave-packets and their approximate localization (confidence estimates) to finite space regions. The instantaneous space localizations and the on them ultimately based descriptions of the actual phenomena is only defined relative a special inertial system (the laboratory system). This does not however, exclude the use of relativistic dynamics.

In a relativistic theory we have to consider two kinds of wave functions which we shall call "space localization wave functions" and "covariant wave functions" respectively. The two are connected with a one-to-one transformation and are thus only different representations of one and the same vector in the Hilbert space. The first wave function (amplitude) is connected with the space localization of the

particles. The second, which has simpler relativistic transformation properties, is convenient for the formulation of the relativistic interaction equations. (In a non-relativistic theory we only need the space localization wave functions).

We shall restrict the discussion here to massive particles with spin zero. The generalization to particles with other spin is more complicated and will not be discussed here.

A relativistic wave function for a spin zero particle with mass m satisfying the Klein-Gordon equation is given by

$$\Phi_c(x) = \int \frac{d^3\vec{k}}{k^0} e^{-ik\cdot x} \cdot a(\vec{k}) \tag{1}$$

where $x = (\vec{x}, t)$ is the space-time variable, $\vec{k}$ is a 3-dimensional momentum space variable, $k = (\vec{k}, k^0)$ and $k^0 = \sqrt{m^2 + \vec{k}^2}$. We denote this amplitude by Φ_c since it is covariant, i.e. transforms as a scalar under Lorentz transformations

However, Φ_c as function of $\vec{x}$ is not the correct probability amplitude for finding the particle in a given space region. This amplitude is instead given by

$$\Phi_s(x) = \int \frac{d^3\vec{k}}{k^0} \cdot k_0^{1/2} \cdot e^{-ik\cdot x} \cdot a(\vec{k}) \tag{2}$$

The statement that the particle is localized to the region $\Omega \subset \mathbb{R}^3$ at time t thus means that $\Phi_s(\vec{x}, t)$ vanishes when $\vec{x} \notin \Omega$. Thus our space localization events and their corresponding projection operators P_Ω has to be defined by means of the amplitude Φ_s.

On the other hand, Φ_c is needed for the formulation of the relativistic interaction equation. This can be written in the interaction picture as a Schwinger-Tomonaga equation. The Hilbert space is then a Fock space built on many particle amplitudes Φ_c, see Schweber (8) Part Two, section 7. With the usual expression of the hamiltonian in terms of annihilation and creation operators, the interaction equation results in a system of coupled wave functions Φ_c for different particle numbers.

7.8 Relativistic Quantum Electrodynamics as a Self-limiting Theory

One way of handling the self-energy problems of quantum electrodynamics is to introduce a high-momentum cutoff in the interaction equations corresponding to a non-locality in the configuration space variables. The annihilation and creation operators occurring in the interaction hamiltonian density then simply means an inner (scalar) product and an outer (direct) product, respectively, with the cutoff function. This gives a simple formulation of the relativistic theory directly in terms of wave functions and refutes the general misconception that relativistic quantum mechanics must be based on quantized fields. With a cutoff it is possible to formulate a mathematically well-defined time-dependent Schrödinger equation for the interaction in terms of the covariant wave functions. This leads to finite mass and

charge renormizations, but the calculated (renormalized) results generally depend on the cutoff.

However, if the characteristic length of the cutoff is small in comparison with the characteristic length of our problem under concern (e.g. atomic dimensions) then the results will be insensitive to the cutoff. The small dependency on the cutoff can then be reinterpreted by saying that the calculated quantities are only approximately defined, with an accuracy given by the cutoff and that this is an expression for the limits of the theory. We can then accept the theory as a self-limiting theory.This should reasonably be sufficient for a "low-energy quantum electrodynamics" constituting a general dynamical basis for physics, except high energy particle physics and cosmological gravitation.

A certain formal elegance of this theory is obtained if we interpret the characteristic length of the cutoff (formally) as a fundamental length. Then this fundamental length has a function of limiting the accuracy of certain quantities, similar to that of Planck's constant.

Part III

Mathematical Theory

Preface to Part III

By a confidence estimate we shall mean an estimate of the form

$$\int_R |u(x)|^2\, dx \geq (1-\epsilon)\|u\|^2$$

where u is a function in $L^2(\mathbb{R}^n)$ and R is some region in $\mathbb{R}^n$. ϵ is to be thought of as a very small number. A reason for using confidence estimates rather than exact localization statements comes from the fact that a "one-particle wave function" $u_1(t,x)$, with time variable t and space variable x in $\mathbb{R}^3$, satisfying the Schrödinger equation and localized exactly at time $t = t_1$ to a space region R_1 (i.e. vanishing outside R_1), will, according to the Schrödinger equation, generally spread out in space so that it can not be localized exactly to any finite region R_2 at another time $t = t_2$. On the other hand we can under certain assumptions obtain that

$$\int_{R_2} |u_1(t_2,x)|^2\, d^3x \geq (1-\epsilon)\|u_1(t_2)\|^2$$

($\|\ \|$ on the right side denotes the L^2-norm) for suitable finite R_2 and very small ϵ.

The following is a mathematical theory which describes solutions to the many-particle Schrödinger equation by means of confidence estimates rather than mean values, used in conventional quantum mechanics and quantum statistical mechanics.

Quantum mechanics is not just the question of finding a solution to the Schrödinger equation. A physical course of events may in general contain stochastic quantum transitions. Such a transition corresponds to a "collapse" of the wave function i.e. a transition from one wave function to another. Therefore a general course of events must be described by a sequence of solutions to the Schrödinger equation.

We shall study sequences of interrelated closed subspaces of solutions. The confidence estimates correspond to certain projection operators and the restriction to certain sequences of such projections, called (approximately) equiangular sequences, permits a unified description both of solutions and the corresponding initial and boundary conditions by means of these projection operators. The concept of equiangular sequences of projections can be considered as a generalization of the asymptotic concept of S-matrix (scattering matrix) and its factorizations/subdivisions in subprocesses to processes in finite regions in space and time.

Technically, this is an elementary approximation theory of subspaces of L^2-functions in connection with Fourier transforms and partial differential equations. Although the theory has an obvious physical content, it will be treated formally as a pure mathematical theory, which will meet common standards with respect to precise definitions and rigorous proofs.

Chapter I

Confidence Approximation Theory

In chapter I we have collected concepts and tools of the present theory which are of a more general mathematical nature than those in the following chapters which are directly connected to a study of the Schrödinger equation.

I.1 Approximate inclusions in Hilbert space

In the following we denote by M, M_1, M', … closed subspaces and by P, P_1, P', … projections in a given complex separable Hilbert space $\mathscr{H}$. With projections (or projection operators) we always mean orthogonal projections.

$R(P)$ denotes the closed subspace on which P projects (the *range* of P). M^c denotes the orthogonal complement of M and $P^c = I - P$ denotes the projection on the orthogonal complement of $R(P)$.

Subspaces which are not necessarily closed will be denoted by N, N_1, N', …. $\dim N$ denotes the dimension (which may be infinite) of N. If X is a subset of $\mathscr{H}$, $P(X)$ denotes the set of all Px with x in X. $M_1 \oplus M_2$ denotes the direct sum of M_1 and M_2. If M_1 is a subspace of M_2, $M_2 - M_1$ denotes the complement of M_1 in M_2, i.e. the set of all vectors in M_2, which are orthogonal to M_1.

For vectors u, v in $\mathscr{H}$, $\langle u, v\rangle$ denotes the scalar product, linear in the first and antilinear in the second argument and $\|u\| = |\langle u, u\rangle|^{1/2}$ denotes the norm of u. If a is a complex number $|a|$ denotes the absolute value. If A is an operator, $\|A\|$ denotes the operator norm. The type of argument of $\|\ \|$ will always make clear what is meant.

A^* denotes the adjoint of the operator A.

Let I be a finite or denumerable index set. We denote by $l^2(I)$ the Hilbert space of sequences a_i, i in I, of complex numbers a_i such that $\sum_{i\in I} |a_i|^2 < \infty$. Linear combination and scalar product of two elements a_i, $i \in I$ and b_i, $i \in I$ are defined by $ca_i + db_i$, $i \in I$, c, d complex numbers, and $\sum_{i\in I} a_i\overline{b_i}$ respectively.

If $I = 1, 2, \ldots, n$, we shall identify $l^2(I)$ with the Hilbert space $\mathbb{C}^n$ of complex n-tuples $(a_1, a_2, \ldots, a_n)$.

I.1.1 Distance From one Subspace to Another

Definition I.1:1. For any nonzero vectors u and v in the Hilbert space $\mathscr{H}$, we define

$$\operatorname{dist}(u,v)) = \left(1 - |\langle u_1, v_1\rangle|^2\right)^{1/2}$$

where $u_1 = u/\|u\|$ and $v_1 = v/\|v\|$.

Definition I.1:2. If u is a nonzero vector and N is a nonzero (not necessarily closed) subspace we define

$$\operatorname{dist}(u,N) = \inf_{v\neq 0 \text{ in } N} \operatorname{dist}(u,v).$$

If u is nonzero and $N = 0$ we define $\operatorname{dist}(u,N) = 1$.

Definition I.1:3. If N_1 and N_2 are nonzero subspaces, we define

$$\operatorname{dist}(N_1,N_2) = \sup_{u\neq 0 \text{ in } N_1} \inf_{v\neq 0 \text{ in } N_2} \operatorname{dist}(u,v).$$

If $N_1 = 0$ we define, for any N_2, $\operatorname{dist}(N_1,N_2) = 0$. If N_1 is nonzero and $N_2 = 0$ we define $\operatorname{dist}(N_1,N_2) = 1$.

Obviously $0 \leq \operatorname{dist}(u,v) \leq 1$, $0 \leq \operatorname{dist}(u,N) \leq 1$ and $0 \leq \operatorname{dist}(N_1,N_2) \leq 1$. If M is closed then $\operatorname{dist}(N,M) = 0$ iff $N \subset M$.

Remark 1. If M_u and M_v are one-dimensional subspaces spanned by the vectors u and v respectively, we have $\operatorname{dist}(M_u,M_v) = \operatorname{dist}(u,M_v) = \operatorname{dist}(u,v)$. If N_1 is nonzero we have $\operatorname{dist}(N_1,N_2) = \sup_{v\neq 0 \text{ in } N_1} \operatorname{dist}(v,N_2)$.

Remark 2. Note that $\operatorname{dist}(u,v) = \operatorname{dist}(v,u)$ for any (nonzero) vectors, but $\operatorname{dist}(M_1,M_2)$ need not be equal to $\operatorname{dist}(M_2,M_1)$. Although one gets a metric on the set of closed subspaces by defining

$$\mathrm{d}(M_1,M_2) = \max\left(\operatorname{dist}(M_1,M_2), \operatorname{dist}(M_2,M_1)\right),$$

(see Kato (1) p. 198, $\mathrm{d}(M_1,M_2) = \|P_1 - P_2\|$ by theorems I.1:7 and I.1:13 below, where P_1 and P_2 are the projections on M_1 and M_2 respectively) the "single-directed" distance $\operatorname{dist}(M_1,M_2)$ will be important in the present theory.

Lemma I.1:4. Let u be a nonzero vector, M a nonzero closed subspace and P the projection on M. Then

$$\operatorname{dist}(u,M) = \operatorname{dist}(u,aPu)$$

where a is any complex nonzero number. If $\|u\| = 1$, then

$$\operatorname{dist}(u,M) = \|u - Pu\|.$$

For any vector v in M which is not in the form aPu, a complex and nonzero, we have

$$\operatorname{dist}(u,v) > \operatorname{dist}(u,M).$$

Proof. Let a be a nonzero complex number and v be any nonzero vector in M. Since by definitions I.1:1 and I.1:2, $\operatorname{dist}(u, v)$ and $\operatorname{dist}(u, M)$ does not change if we multiply by a nonzero complex number, it is no limitation to assume that $\|u\| = 1$. From definition I.1:1 then follows that $\operatorname{dist}(u, aPu) = \|u - Pu\|$. We can choose a complex number b such that $v_1 = bv$ is equal to the projection of u on the one-dimensional subspace spanned by v. Then v_1 and $u - v_1$ are orthogonal and $\operatorname{dist}(u, v) = \|u - v_1\|$. We have

$$u - v_1 = \big(u - Pu\big) + \big(Pu - v_1\big)$$

where $\big(u - Pu\big)$ and $\big(Pu - v_1\big)$ are orthogonal ($Pu - v_1$ is in M and $u - Pu$ is orthogonal to M). Thus

$$\|u - v_1\|^2 = \|u - Pu\|^2 + \|Pu - v_1\|^2$$

and

$$\operatorname{dist}(u, v) = \|u - v_1\| \geq \|u - Pu\| = \operatorname{dist}(u, Pu).$$

This shows that

$$\begin{aligned} \operatorname{dist}(u, M) &= \inf_{v \neq 0 \text{ in } M} \operatorname{dist}(u, v) \\ &= \operatorname{dist}(u, Pu) = \operatorname{dist}(u, aPu) = \|u - Pu\|. \end{aligned}$$

If v is not of the form $v = cPu$, c complex and nonzero, then $\|Pu - v_1\| > 0$ and

$$\operatorname{dist}(u, v) = \|u - v_1\| > \|u - Pu\| = \operatorname{dist}(u, Pu).$$

□

Lemma I.1:5. For any nonzero u, v, w in $\mathscr{H}$, we have

$$\operatorname{dist}(u, w) \leq \operatorname{dist}(u, v) + \operatorname{dist}(v, w).$$

Proof. Choose $u_1 = au$, $v_1 = bv$, $w_1 = cw$, a, b, c complex numbers such that $\|u_1\| = \|v_1\| = \|w_1\| = 1$, $\langle u_1, v_1 \rangle \geq 0$ and $\langle v_1, w_1 \rangle \geq 0$. Set

$$\begin{aligned} u_1 &= \cos A \cdot v_1 + \sin A \cdot u_1', \quad \cos A \ = \langle u_1, v_1 \rangle, \\ w_1 &= \cos B \cdot v_1 + \sin B \cdot w_1', \quad \cos B = \langle w_1, v_1 \rangle, \\ \langle u_1', v_1 \rangle &= \langle w_1', v_1 \rangle = 0, \quad 0 \leq A, B \leq \frac{\pi}{2}. \end{aligned}$$

If $A + B \leq \pi/2$

$$\begin{aligned} &\langle u_1, w_1 \rangle = \cos A \cos B + \sin A \sin B \cdot d, \quad d = \langle u_1', w_1' \rangle, \quad |d| \leq 1 \\ &|\langle u_1, w_1 \rangle| \geq \cos A \cos B - \sin A \sin B = \cos(A + B) \geq 0 \end{aligned}$$

and

$$\operatorname{dist}(u,w)=\operatorname{dist}(u_1,w_1)=\left[1-|\langle u_1,w_1\rangle|^2\right]^{1/2}\le\sin(A+B)$$
$$\le\sin A+\sin B=\operatorname{dist}(u,v)+\operatorname{dist}(v,w).$$

If $A+B\ge\pi/2$,

$$\operatorname{dist}(u,v)+\operatorname{dist}(v,w)=\sin A+\sin B\ge 1\ge\operatorname{dist}(u,w).$$

□

Theorem I.1:6. For any subspaces N_1, N_2, N_3, we have

$$\operatorname{dist}(N_1,N_3)\le\operatorname{dist}(N_1,N_2)+\operatorname{dist}(N_2,N_3).$$

Proof. If any of N_1, N_2, N_3 is zero, the result follows from definition I.1:3. Suppose N_1, N_2, N_3 nonzero. Let ε be an arbitrary positive number. For arbitrary vector u_1 in N_1 there exists, by definition, a vector u_2 in N_2 with $\operatorname{dist}(u_1,u_2)\le\operatorname{dist}(N_1,N_2)+\varepsilon$ and a vector u_3 in N_3 with $\operatorname{dist}(u_2,u_3)\le\operatorname{dist}(N_2,N_3)+\varepsilon$. Then by lemma I.1:5,

$$\operatorname{dist}(u_1,u_3)\le\operatorname{dist}(u_1,u_2)+\operatorname{dist}(u_2,u_3)$$
$$\le\operatorname{dist}(N_1,N_2)+\operatorname{dist}(N_2,N_3)+2\varepsilon.$$

Since u_1 is arbitrary in N_1 we have, by definition, that

$$\operatorname{dist}(N_1,N_3)\le\operatorname{dist}(N_1,N_2)+\operatorname{dist}(N_2,N_3)+2\varepsilon,$$

and, since ε is arbitrary, the result follows. □

Theorem I.1:7. Let P_1 and P_2 be the projections on M_1 and M_2 respectively. Then

$$\operatorname{dist}(M_1,M_2)=\|(I-P_2)P_1\|.$$

If M_1 is nonzero then

$$\operatorname{dist}(M_1,M_2)=\sup_{u\in M_1,\,\|u\|=1}\|(I-P_2)u\|$$

Proof. The result is obvious if $M_1=0$. Suppose M_1 is nonzero. By remark 1 and lemma I.1:4

$$\operatorname{dist}(M_1,M_2)=\sup_{u\ne 0\in M_1}\operatorname{dist}(u,M_2)\;=\sup_{\substack{u\in M_1\\ \|u\|=1}}\operatorname{dist}(u,M_2)$$
$$=\sup_{\substack{u\in M_1\\ \|u\|=1}}\|u-P_2u\|\quad=\sup_{\substack{u\in M_1\\ \|u\|=1}}\|(I-P_2)u\|.$$

For v in $\mathscr{H}$ with $\|v\| = 1$ we have that $u = P_1 v / \|P_1 v\|$ is in M_1, with $\|u\| = 1$ and

$$\|(I - P_2)P_1 v\| = \|(I - P_2)u\| \cdot \|P_1 v\| \leq \|(I - P_2)u\|$$

with equality if v is in M_1. Thus

$$\|(I - P_2)P_1\| = \sup_{\|v\|=1,\ v\in\mathscr{H}} \|(I - P_2)P_1 v\| = \sup_{\|u\|=1,\ u\in M_1} \|(I - P_2)P_1 u\|.$$

□

Theorem I.1:8. For any M_1 and M_2 we have

$$\operatorname{dist}(M_1, M_2) = \operatorname{dist}(M_2^c, M_1^c).$$

Proof. By theorem I.1:7

$$\begin{aligned}\operatorname{dist}(M_1, M_2) &= \|(I - P_2)P_1\| = \|((I - P_2)P_1)^*\| \\ &= \|P_1^*(I - P_2)^*\| = \|P_1(I - P_2)\| = \|(I - P_1^c)P_2^c\| \\ &= \operatorname{dist}(M_2^c, M_1^c).\end{aligned}$$

□

Remark 3. It follows immediately from definition I.1:3 that, if M_1 and M_2 are the closures of N_1 and N_2 respectively,

$$\operatorname{dist}(N_1, N_2) = \operatorname{dist}(N_1, M_2) = \operatorname{dist}(M_1, N_2) = \operatorname{dist}(M_1, M_2).$$

Thus, the preceding theorem holds also for nonclosed subspaces N_1, N_2 instead of M_1, M_2.

Lemma I.1:9. If $\dim N_1 > \dim N_2$ then there exists a nonzero u in N_1 orthogonal to N_2.

Proof. If $\dim N_1 > \dim N_2$ then N_2 must be finite-dimensional. Let N_1' be any $n+1$ dimensional subspace of N_1, $n = \dim N_2$ ($N_1' = N_1$ if $\dim N_1 = n+1$). Choose $n+1$ linearly independent vectors $x_1, x_2, \ldots, x_{n+1}$ in N_1' and let P_2 be the projection on N_2. Then the $n+1$ vectors $P_2 x_1, \ldots, P_2 x_{n+1}$ in N_2 are linearly dependent. Thus there exist numbers $c_1, c_2, \ldots, c_{n+1}$, not all zero, such that $c_1 P_2 x_1 + \cdots + c_{n+1} P_2 x_{n+1} = 0$. Then $u = c_1 x_1 + \cdots + c_{n+1} x_{n+1}$ is a nonzero vector in N_1', with $P_2 u = 0$, thus orthogonal to N_2. □

Theorem I.1:10. If $\operatorname{dist}(N_1, N_2) < 1$ then $\dim N_1 \leq \dim N_2$.

Proof. Suppose $\dim N_1 > \dim N_2$. Then, by lemma I.1:9, there exists a nonzero vector u_1 in N_1, orthogonal to N_2. Then $\operatorname{dist}(u_1, u_2) = 1$ for every u_2 in N_2 and $\operatorname{dist}(N_1, N_2) = 1$ by definition I.1:3. □

Lemma I.1:11. If $\operatorname{dist}(M_1, M_2) < 1$ and P_2 is the projection on M_2 then P_2 maps M_1 onto $P_2 M_1$ one-to-one and bicontinuously (both map and inverse are continuous).

Proof. The case $M_1 = 0$ is trivial. Suppose M_1 nonzero. For arbitrary u in M_1 with $\|u\| = 1$

$$\Big|\|P_2u\| - 1\Big| = \Big|\|P_2u\| - \|u\|\Big| \le \|u - P_2u\|$$
$$\le \sup_{\substack{u' \in M_1 \\ \|u'\|=1}} \|u' - P_2u'\| = \operatorname{dist}(M_1, M_2) = d < 1$$

according to theorem I.1:7, thus $\|P_2u\| \ge 1 - d > 0$. Then, for arbitrary u in M_1

$$\|P_2u\| \ge (1 - d)\|u\|$$

which shows that the map is one-to-one with continuous inverse. □

Theorem I.1:12. If $\operatorname{dist}(M_1, M_2) < 1$ and P_2 is the projection on M_2, then P_2M_1 is closed.

Proof. Recalling that our M:s are closed subspaces in the complete metric space $\mathscr{H}$, the theorem is an immediate consequence of lemma I.1:11. □

Remark 4. P_2M_1 need not be closed if $\operatorname{dist}(M_1, M_2) = 1$.

Cosely related to the present theory is the following theorem by Kato.

Theorem I.1:13. Let P_1 and P_2 be projections, projecting on M_1 and M_2 respectively and suppose that $\|(I - P_2)P_1\| = d < 1$. Then either

1) P_2 maps M_1 onto M_2 one-to-one and bicontinuously and

$$\|P_1 - P_2\| = \|(I - P_1)P_2\| = \|(I - P_2)P_1\| = d$$

or

2) P_2 maps M_1 onto a proper subspace M_2' of M_2 one-to-one and bicontinuously and if P_2' is the projection on M_2', then

$$\|P_1 - P_2'\| = \|(I - P_1)P_2'\| = \|(I - P_2')P_1\| = \|(I - P_2)P_1\| = d$$

and

$$\|P_1 - P_2\| = \|(I - P_1)P_2\| = 1.$$

We shall not reproduce the proof here but refer to Kato (1) p. 56-58.

I.1.2 Approximate Inclusions

Definition I.1:14. We say that N_1 is an ε-approximate ($\varepsilon \ge 0$) subspace to N_2 and write $N_1 \subset_\varepsilon N_2$ if $\operatorname{dist}(N_1, N_2) \le \varepsilon$.

Remark 5. For $\varepsilon = 0$, the relation $N_1 \subset_0 M_2$ coincides with the subspace relation $N_1 \subset M_2$ (M_2 closed!)

If M_1 is the one-dimensional subspace spanned by the vector u, we shall also write $u \subset_\varepsilon N_2$ for $M_1 \subset_\varepsilon N_2$.

Theorem I.1:15. $N_1 \subset_\varepsilon M_2$ iff

$$\|P_2 u\|^2 \geq (1-\varepsilon^2)\|u\|^2$$

for every u in N_1, where P_2 is the projection on M_2

Proof. The relation

$$\|P_2 u\|^2 \geq (1-\varepsilon^2)\|u\|^2$$

is obviously equivalent to $\operatorname{dist}(u, P_2 u) \leq \varepsilon$. The theorem then follows from lemma I.1:4 and remark 1 after definition I.1:3. □

Theorem I.1:16. If $N_1 \subset_{\varepsilon'} N_2$ and $N_2 \subset_{\varepsilon''} N_3$ then $N_1 \subset_{\varepsilon'+\varepsilon''} N_3$

Proof. A direct reformulation of theorem I.1:6 by means of definition I.1:14 □

Theorem I.1:17. If $M_1 \subset_{\varepsilon'} M_2$ and $M_2 \subset_{\varepsilon''} M_1$ with $\varepsilon', \varepsilon'' < 1$, then

$$\operatorname{dist}(M_1, M_2) = \operatorname{dist}(M_2, M_1) = \|P_1 - P_2\| \leq \min(\varepsilon', \varepsilon'')$$

where P_1 and P_2 are the projections on M_1 and M_2 respectively.

Proof. By theorem I.1:7

$$\begin{aligned} &\operatorname{dist}(M_1, M_2) = \|(I-P_2)P_1\| \leq \varepsilon' < 1 \quad \text{and} \\ &\operatorname{dist}(M_2, M_1) = \|(I-P_1)P_2\| \leq \varepsilon'' < 1. \end{aligned}$$

Then, by theorem I.1:13

$$\|P_1 - P_2\| = \|(I-P_1)P_2\| = \|(I-P_2)P_1\| \leq \min(\varepsilon', \varepsilon'').$$

□

Theorem I.1:18. If $M_1 \subset_\varepsilon M_2$, then $M_2^c \subset_\varepsilon M_1^c$.

Proof. A reformulation of theorem I.1:8 by means of definition I.1:14. □

Theorem I.1:19. If $N_1 \subset_\varepsilon N_2$ with $\varepsilon < 1$, then $\dim N_1 \leq \dim N_2$.

Proof. A reformulation of theorem I.1:10 by means of definition I.1:14. □

Definition I.1:20. N_1 and N_2 are called ε.orthogonal if $|\langle u_1, u_2\rangle| \leq \varepsilon$ for every u_1 in N_1 and u_2 in N_2 with $\|u_1\| = \|u_2\| = 1$.

Theorem I.1:21. N_1 and N_2 are ε-orthogonal iff $N_1 \subset_\varepsilon N_2^c$

Proof. Let u_1 and u_2 be arbitrary vectors in N_1 and N_2 respectively with $\|u_1\| = \|u_2\| = 1$, and set $u_1 = u_1' + u_1''$ with u_1' in N_2^c and u_1'' in the closure of N_2. Then $\langle u_1, u_2\rangle = \langle u_1'', u_2\rangle$ and the result follows since $N_1 \subset_\varepsilon N_2^c$ iff $\|u_1''\| \leq \varepsilon$. □

Lemma I.1:22. If $N_1, \ldots, N_m, N_1', \ldots, N_n'$ are $m+n$ ε-orthogonal subspaces, then $N = \bigoplus_{i=1}^m N_i$ and $N' = \bigoplus_{i=i}^n N_i'$ are ε_1-orthogonal with $\varepsilon_1 = 2mn\varepsilon$.

Proof. Let $u = \sum_{i=i}^{m} u_i$, u_i in N_i and $v = \sum_{i=1}^{n} v_i$, v_i in N_i'. Then

$$\|u\|^2 = \sum_{i,j} \langle u_i, u_j \rangle = \sum \|u_i\|^2 + \varepsilon' \quad \text{where} \quad \varepsilon' = \sum_{i \neq j} \langle u_i, u_j \rangle$$

$$|\varepsilon'| \leq \varepsilon \sum_{i \neq j} \|u_i\| \cdot \|u_j\| = \varepsilon \left[\Big(\sum \|u_i\| \Big) \Big(\sum \|u_j\| \Big) - \sum \|u_i\|^2 \right]$$

$$\leq \varepsilon (m-1) \sum \|u_i\|^2.$$

Thus

$$\|u\|^2 \geq \Big[1 - (m-1)\varepsilon \Big] \sum \|u_i\|^2 \geq \frac{1}{2} \sum \|u_i\|^2$$

if $2m\varepsilon < 1$ and similarly for v. Then

$$|\langle u, v \rangle| = \Big| \sum_i \sum_j \langle u_i, v_j \rangle \Big| \leq \varepsilon \sum_i \sum_j \|u_i\| \cdot \|v_j\| = \varepsilon \Big(\sum \|u_i\| \Big) \Big(\sum \|v_j\| \Big)$$

$$\leq \varepsilon \cdot m \cdot n \Big[\sum \|u_i\|^2 \Big]^{1/2} \cdot \Big[\sum \|v_j\|^2 \Big]^{1/2} \leq 2mn\varepsilon \cdot \|u\| \cdot \|v\|.$$

and thus u and v are $2mn\varepsilon$-orthogonal. If $2m\varepsilon \geq 1$ or $2n\varepsilon \geq 1$, then $2mn\varepsilon \geq 1$ and the statement is trivially satisfied. □

I.1.3 Some Auxiliary Theorems on Approxmation of Subspaces

Lemma I.1:23. $N_1 \subset_\varepsilon N_2$ if for every vector u_1 in N_1 there exists a u_2 in N_2 with

$$\|u_1 - u_2\| \leq \varepsilon \cdot \|u_1\|$$

Proof. A direct consequence of lemma I.1:4 and remark 1 after definition I.1.1:3 if N_1 and N_2 are closed. If N_1 or N_2 are not closed, let M_1 and M_2 be the closures of N_1 and N_2. The statement for N_1 and N_2 then follows from the statement for M_1 and M_2 by using remark 3 after theorem I.1:8. □

Lemma I.1:24. $N_1 \subset_\varepsilon M_2$ iff for every u in N_1

$$\|u - P_2 u\| \leq \varepsilon \|u\|,$$

where P_2 is the projection on M_2

Proof. A direct consequence of lemma I.1:4 and remark 1 after definition I.1:3 □

Lemma I.1:25. If $N_1 \subset_\varepsilon M_2$ and P_2 is the projection on M_2 then

$$N_1 \subset_\varepsilon P_2 N_1 \quad \text{and}$$
$$P_2 N_1 \subset_\varepsilon N_1.$$

Proof. The first statement follows directly from lemma I.1:24.

Let u be an arbitrary nonzero vector in $P_2 N_1$. Then u is of the form $u = P_2 v$ with v in N_1 and lemma I.1:4 gives $\operatorname{dist}(u, v) = \operatorname{dist}(v, u) = \operatorname{dist}(v, P_2 v) = \operatorname{dist}(v, M_2) \leq$

ε. Thus for arbitrary nonzero u in P_2N_1 there is a v in N_1 with $\operatorname{dist}(u,v) \le \varepsilon$. Hence $\operatorname{dist}(P_2N_1, N_1) \le \varepsilon$ and $P_2N_1 \subset_\varepsilon N_1$. □

Lemma I.1:26. Suppose that M_1 and M_2 are orthogonal and that $M_1 \subset_{\varepsilon_1} M_1'$ and $M_2 \subset_{\varepsilon_2} M_2'$. Then $M_1 \oplus M_2 \subset_{\varepsilon_1+\varepsilon_2} M_1' \oplus M_2'$.

Proof. Let u be an arbitrary vector in $M_1 \oplus M_2$. Then u is of the form $u = u_1 + u_2$ with u_1 and u_2 in M_1 and M_2 respectively. By lemma I.1:23 there exists u_1' and u_2' in M_1' and M_2' respectively with $\|u_1 - u_1'\| \le \varepsilon_1\|u_1\|$ and $\|u_2 - u_2'\| \le \varepsilon_2\|u_2\|$. Hence there exists an $u' = u_1' + u_2'$ in $M_1' \oplus M_2'$ with

$$\begin{aligned}\|u - u'\| &\le \|u_1 - u_1'\| + \|u_2 - u_2'\| \le \varepsilon_1\|u_1\| + \varepsilon_2\|u_2\| \\ &\le \left(\varepsilon_1^2 + \varepsilon_2^2\right)^{1/2}\left(\|u_1\|^2 + \|u_2\|^2\right)^{1/2} = \left(\varepsilon_1^2 + \varepsilon_2^2\right)^{1/2}\|u\| \\ &\le \left(\varepsilon_1 + \varepsilon_2\right)\|u\|\end{aligned}$$

since u_1 and u_2 are orthogonal and thus $M_1 \oplus M_2 \subset_{\varepsilon_1+\varepsilon_2} M_1' \oplus M_2'$ □

Lemma I.1:27. Suppose $M \subset_{\varepsilon'} M'$ and $M \subset_{\varepsilon''} M''$. Then $M \subset_{\varepsilon'+\varepsilon''} P''M'$ where P'' is the projection on M''.

Proof. Let u be an arbitrary vector in M and let P' be the projection on M'. Then by lemma I.1:24 $\|P'u - u\| \le \varepsilon'\|u\|$, $\|P''u - u\| \le \varepsilon''\|u\|$ and

$$\begin{aligned}\|P''P'u - u\| &\le \|P''P'u - P''u\| + \|P''u - u\| \le \|P''(P'u - u)\| + \varepsilon''\|u\| \\ &\le \|P'u - u\| + \varepsilon''\|u\| \le \left(\varepsilon' + \varepsilon''\right)\|u\|.\end{aligned}$$

Sinse $P''P'u$ is in $P''M'$, lemma I.1:23 then gives that $M \subset_{\varepsilon'+\varepsilon''} P''M'$. □

Lemma I.1:28. Suppose $M \subset_{\varepsilon'} M'$ and $M \subset_{\varepsilon''} M''$. Then $P''M \subset_{\varepsilon'+\varepsilon''} M'$ where P'' is the projection on M''.

Proof. Let u be a vector in M. Then we have that $\operatorname{dist}(P''u, u) = \operatorname{dist}(u, P''u) \le \varepsilon''$, $\operatorname{dist}(u, P'u) \le \varepsilon'$ and by lemma I.1:5

$$\operatorname{dist}(P''u, P'u) \le \operatorname{dist}(P''u, u) + \operatorname{dist}(u, P'u) \le \varepsilon' + \varepsilon''.$$

Since $P'u$ is in M', this shows that $\operatorname{dist}(P''u, M') \le \varepsilon' + \varepsilon''$ and since an arbitrary vector in $P''M$ is of the form $P''u$ with u in M, this also shows that $\operatorname{dist}(P''M, M') \le \varepsilon' + \varepsilon''$ and it follows that $P''M \subset_{\varepsilon'+\varepsilon''} M'$. □

Lemma I.1:29. Suppose that $M_1 = R(P_1)$, $M_2 = R(P_2)$, P commutes with P_1 and P_2 and $M_1 \subset_\varepsilon M_2$. Then $PM_1 \subset_\varepsilon PM_2$.

Proof. For arbitrary u in PM_1, we have, since $PM_1 \subset M_1$, by lemma I.1:24 that $\|u - P_2u\| \le \varepsilon$. But $P_2u = P_2Pu = PP_2u$ is in PM_2 and the result follows from lemma I.1:23. □

From the preceding theorems and lemmas one can derive more composite results. The following will be used in I.2.1.

Lemma I.1:30. Suppose that

1) $M_0 \subset M_1 \oplus M_2$, M_1 and M_2 orthogonal,

2) $M_1' \oplus M_2' = \mathscr{H}$, M_1' and M_2' orthogonal,

3) $M_0 \subset_{\varepsilon'} M_1'$,

4) $M_2 \subset_{\varepsilon''} M_2'$,

Then $M_0 \subset_{\varepsilon'+\varepsilon''} M_1$

Proof. Let P_2 be the projection on M_2. By 2), 4) and theorem I.1:18 $M_1' = M_2'^c \subset_{\varepsilon''} M_2^c$. 3) and theorem I.1:16 then gives $M_0 \subset_{\varepsilon'+\varepsilon''} M_2^c$. But $M_0 \subset_0 M_1 \oplus M_2$ according to 1) and lemma I.1:27 then gives $M_0 \subset_{\varepsilon'+\varepsilon''} P_2^c(M_1 \oplus M_2) = M_1$. □

I.1.4 Approximation in Product Spaces

In this section we shall consider several complex, separable Hilbert spaces, $\mathscr{H}$, $\mathscr{H}_1$, $\mathscr{H}_2$, … simultaneously. Although we shall use the same notation $\langle\ ,\ \rangle$ for scalar product, ^{c} for orthogonal complement etc. in different spaces, there should be no risk for confusion since the arguments in the expressions will always make clear what is meant.

Definition I.1:31. Let $\mathscr{H}$, $\mathscr{H}_1$,…, $\mathscr{H}_n$, be Hilbert spaces. A mapping T of $\mathscr{H}_1 \times \ldots \times \mathscr{H}_n$ into $\mathscr{H}$ is called a tensor product mapping and we write

$$T :_{\otimes} \mathscr{H}_1 \times \ldots \times \mathscr{H}_n \to \mathscr{H}$$

if

1° $\langle T(u_1, \ldots, u_n), T(u_1', \ldots, u_n')\rangle = \langle u_1, u_1'\rangle \cdots \langle u_n, u_n'\rangle$ for any u_i, u_i' in $\mathscr{H}_i$, $i = 1, \ldots, n$

2° $\mathscr{H}$ is the closed linear hull of $T(\mathscr{H}_1 \times \ldots \times \mathscr{H}_n)$.

If M_i, $i = 1, \ldots, n$ are closed linear subspaces in $\mathscr{H}_i$, $i = 1, \ldots, n$ respectively, the closed linear hull of $T(M_1 \times \ldots \times M_n)$ will be denoted by $\bigotimes_{i=1}^{n}{}_T M_i$, or alternatively by $M_1 \otimes_T M_2 \otimes_T \ldots \otimes_T M_n$, thus especially $\mathscr{H} = \bigotimes_{i=1}^{n}{}_T \mathscr{H}_i$. If some of the factors M_i are one dimensional spanned by vectors u_i we shall also write $M_1 \otimes_T \ldots \otimes_T u_i \otimes_T \ldots$ for their product, thus especially $T(u_1, \ldots, u_n) = \bigotimes_{i=1}^{n}{}_T u_i = u_1 \otimes_T \ldots \otimes_T u_n$.

If $A_1, \ldots, A_n$ are bounded linear operators in $\mathscr{H}_1$,…, $\mathscr{H}_n$ respectively we denote by $\bigotimes_{1}^{n}{}_T A_i = A_1 \otimes_T \cdots \otimes_T A_n$ the bounded linear operator, which is the extension to $\mathscr{H}$ of the bounded map A_0 in $T(\mathscr{H}_1 \times \ldots \times \mathscr{H}_n)$ defined by

$$A_0\Big(\bigotimes_{i=1}^{n}{}_T u_i\Big) = \bigotimes_{i=1}^{n}{}_T A_i u_i, \quad u_i \text{ in } \mathscr{H}_i,\ i = 1, \ldots, n$$

If $T :_{\otimes} \mathscr{H}_1 \times \ldots \times \mathscr{H}_n \to \mathscr{H}$, we shall also say that T is a tensor product decomposition of $\mathscr{H}$ into $\mathscr{H}_1 \times \ldots \times \mathscr{H}_n$.

Lemma I.1:32. Suppose that

$$T :_{\otimes} \mathscr{H}_1 \times \ldots \times \mathscr{H}_n \to \mathscr{H}$$

and $\{u_{i,j}\}$, $j \in J_i$ is an orthonormal basis for $\mathscr{H}_i$, $i = 1, \ldots, n$. Then $\{u_k\}$, $k \in K = J_1 \times \ldots \times J_n$ where $u_k = u_{1,j_1} \otimes \ldots u_{n,j_n}$, $k = (j_1, \ldots, j_n)$, is an orthonormal basis for $\mathscr{H}$.

Proof. A simple consequence of definition I.1:31. □

Lemma I.1:33. Let $\{u_{i,j}\}$, $i \in I$, $j \in J$ be a double-indexed orthonormal basis for $\mathscr{H}$ and set $\mathscr{H}_1 = l^2(I)$, $\mathscr{H}_2 = l^2(J)$ and

$$T(u, v) = \sum_{\substack{i \in I \\ j \in J}} a_i b_j u_{i,j}$$

for $u = \{a_i\}$, $i \in I$, $v = \{b_j\}$, $j \in J$. Then

$$T :_{\otimes} \mathscr{H}_1 \times \mathscr{H}_2 \to \mathscr{H}.$$

Proof. Also a simple consequence of definition I.1:31. □

The following variant of the preceding lemma will be used in III.5.

Lemma I.1:34. Suppose that

$$\mathscr{H} = \bigoplus_{i \in I} M_i$$

is a decomposition of $\mathscr{H}$ into orthogonal closed subspaces M_i, $i \in I$, all with the same dimension and

$$U_i : \mathscr{H}_2 \to M_i$$

are isometric onto mappings and set

$$T(u, v) = \sum_{i \in I} a_i U_i v$$

for $u = \{a_i\}$, $i \in I$ in $l^2(I)$ and v in $\mathscr{H}_2$. Then $T :_{\otimes} l^2(I) \times \mathscr{H}_2 \to \mathscr{H}$.

Proof. Also a simple consequence of definition I.1:31. □

Theorem I.1:35. Suppose that $T :_{\otimes} \mathscr{H}_1 \times \mathscr{H}_2 \to \mathscr{H}$, M_1 and M_2 are closed subspaces in $\mathscr{H}_1$ and $M_1' = M_1 \otimes_T \mathscr{H}_2$, $M_2' = M_2 \otimes_T \mathscr{H}_1$. Then $M_1 \subset_\varepsilon M_2$ iff $M_1' \subset_\varepsilon M_2'$.

Proof. Let $\{u_i\}$, $i \in I = I_1 \cup I_2$ be an orthonormal basis for $\mathscr{H}_1$ such that $\{u_i\}$, $i \in I_1$ is a basis for M_2 and $\{u_i\}$, $i \in I_2$ is a basis for M_2^c. Let $\{v_j\}$, $j \in J$ be an orthonormal basis of the space $\mathscr{H}_2$. Then, by lemma I.1:32, $\{u_i \otimes_T v_j\}$, $i \in I$, $j \in J$ is an orthonormal basis for $\mathscr{H}$ and $\{u_i \otimes_T v_j\}$, $i \in I_k$, $j \in J$ is an orthonormal basis for $M_2 \otimes_T \mathscr{H}_2$ if $k = 1$ and for $M_2^c \otimes_T \mathscr{H}_2$ if $k = 2$.

Suppose $M_1 \subset_\varepsilon M_2$. Let u be an arbitrary element in M_1' with $\|u\| = 1$. Then $u = \sum_{\substack{i\in I\\ j\in J}} a_{ij}u_i \otimes_T v_j$ where $\sum_{j\in J} |a_{ij}|^2 = 1$. For every $j \in J$, $\sum_{i\in I} a_{ij}u_i$ is in M_1. Thus $\sum_{i\in I_2} |a_{ij}|^2 \le \varepsilon^2 \sum_{i\in I} |a_{ij}|^2$ and $\sum_{\substack{i\in I_2\\ j\in J}} |a_{ij}|^2 \le \varepsilon^2$, which means that $u \subset_\varepsilon M_2 \otimes \mathscr{H}_2 = M_2'$. Conversely, suppose $M_1' \subset_\varepsilon M_2'$. Let u be an arbitrary element in M_1 with $\|u\| = 1$ and let P_2 and P_2' be the projections on M_2 and M_2' respectively. If v is any vector in $\mathscr{H}_2$ with $\|v\| = 1$ we have $u \otimes_T v \in M_1'$ and

$$\|P_2'^c(u \otimes_T v)\| \le \varepsilon \|u \otimes_T v\| = \varepsilon \|u\| \cdot \|v\| = \varepsilon.$$

But $P_2'^c(u \otimes_T v) = \left(P_2^c u\right) \otimes_T v$ and $\|P_2'^c(u \otimes_T v)\| = \|P_2^c u\|$. Thus $\|P_2^c u\| \le \varepsilon$ and since u is arbitrary in M_1, we have $M_1 \subset_\varepsilon M_2$. □

Theorem I.1:36. Suppose that $T :_\otimes \mathscr{H}_1 \times \mathscr{H}_2 \to \mathscr{H}$, M_1, M_1' are closed subspaces in $\mathscr{H}_1$ and M_2, M_2' are closed subspaces in $\mathscr{H}_2$. If $M_1 \subset_{\varepsilon_1} M_1'$ and $M_2 \subset_{\varepsilon_2} M_2'$ then

$$M_1 \otimes_T M_2 \subset_{\varepsilon_1+\varepsilon_2} M_1' \otimes_T M_2'.$$

Proof. Suppose $M_1 \subset_{\varepsilon_1} M_1'$ and $M_2 \subset_{\varepsilon_2} M_2'$. Then, by theorem I.1:35

$$M_1 \otimes_T M_2 \subset M_1 \otimes_T \mathscr{H}_2 \subset_{\varepsilon_1} M_1' \otimes_T \mathscr{H}_2 \text{ and}$$
$$M_1 \otimes_T M_2 \subset \mathscr{H}_1 \otimes_T M_2 \subset_{\varepsilon_2} \mathscr{H}_1 \otimes_T M_2'.$$

If P is the projection on $\mathscr{H}_1 \otimes_T M_2'$ we have that

$$P\left(M_1' \otimes_T \mathscr{H}_2\right) = M_1' \otimes_T M_2'$$

and the rest follows from lemma I.1:27. □

I.2 Confidence Estimates and Fourier Transforms

We use the standard notations $L^1(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$ for the set of measurable functions on $\mathbb{R}^n$ for which

$$\int |f(x)|\,dx < \infty \text{ or } \int |f(x)|^2\,dx < \infty$$

respectively. $L^2(\mathbb{R}^n)$ is a separable Hilbert space with scalar product $\langle f, g\rangle = \int f(x)\overline{g(x)}\,dx$. (Functions differing only on a set of measure zero are in this context considered as identical.) We denote by $\mathrm{supp}(f)$ the support of f (smallest closed set outside which f vanishes almost everywhere).

Let R be a measurable subset of $\mathbb{R}^n$ whose boundary $R^{\mathrm{clos}} - R^{\mathrm{int}}$ has measure zero. (R^{clos} denotes the closure and R^{int} the interior of R.) We have $\mathrm{supp}(f) \subset R^{\mathrm{clos}}$ iff $f = \chi_R f$ (almost everywhere), where χ_R denotes the characteristic function of R. For such a region R we denote by $L^2(R)$ the set of L^2-functions with support in

R^{clos}. For every such set R, $L^2(\mathbb{R}^n)$ decomposes into a direct sum

$$L^2(\mathbb{R}^n) = L^2(R) \bigoplus L^2(\mathbb{R}^n - R)$$

of the two closed, orthogonal subspaces $L^2(R)$ and $L^2(\mathbb{R}^n - R)$.

We define the Fourier transform $\hat{f}$ of a function f in $L^1(\mathbb{R}^n)$ by the formula

$$\hat{f}(y) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \int e^{-i\langle x,y\rangle} f(x)\, dx$$

where $\langle x, y\rangle = x_1 y_1 + \cdots + x_n y_n$, $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$. The factor $1/\left(\sqrt{2\pi}\right)^n$ is chosen since it gives simultaneous normalization of f and $\hat{f}$ (see (2) below). The Fourier transform can be extended to a unitary map of $L^2(\mathbb{R}^n)$ onto itself and for later reference we recapitulate the following formulas, valid for any f, g in $L^2(\mathbb{R}^n)$.

$$\text{If } g = \hat{f} \text{ then } f(x) = \hat{g}(-x) \tag{1}$$

$$\|\hat{f}\| = \|f\| \tag{2}$$

$$\langle \hat{f}, \hat{g}\rangle = \langle f, g\rangle \tag{3}$$

$$g(x) = e^{i\langle b,x\rangle} f(x),\ b \in \mathbb{R}^n \text{ iff } \hat{g}(y) = \hat{f}(y - b) \tag{4}$$

$$\text{If } g(x) = f(ax)\, a \in \mathbb{R}^1,\ a \neq 0 \text{ then } \hat{g}(y) = \frac{1}{a^n}\hat{f}(y/a) \tag{5}$$

$$g = -i\frac{\partial f}{\partial x_j} \text{ iff } \hat{g}(y) = y_j \hat{f}(y) \tag{6}$$

$$\widehat{f \cdot g} = \left(\sqrt{2\pi}\right)^{-n} \hat{f} * \hat{g} \tag{7}$$

where $*$ denotes the convolution, i.e.

$$(f * g)(x) = \int f(x - x')g(x')\, dx'.$$

In (1), (4), (5) and (6) the relations are to be interpreted in the usual "almost everywhere" sense. The derivative in (6) is to be interpreted in the weak (distributional) sense. (If f and g are continuous, the weak derivative coincides with the strong (classical) derivative.)

For a subspace N of $L^2(\mathbb{R}^n)$ we denote by $\hat{N}$ the subspace of functions $\hat{f}$ with f in N.

I.2.1 Simultaneous Confidence Estimates on a Subspace of Functions and the Subspace of their Fourier Transforms

Definition I.2:37. Let f be a function in $L^2(\mathbb{R}^n)$ and R a region (measurable set whose boundary has measure zero) in $\mathbb{R}^n$. By a confidence estimate (on f) we

shall mean an estimate of the form

$$\int_R |f(x)|^2\,dx \geq (1-\varepsilon^2)\int_{\mathbb{R}^n} |f(x)|^2\,dx.$$

We shall denote this relation by $f \subset_\varepsilon R$.

Remark: Let $M = L^2(R)$ be the closed subspace of $L^2(\mathbb{R}^n)$ of functions with support in R. Then, with the notation of I.1.2, $f \subset_\varepsilon R$ iff $f \subset_\varepsilon M = L^2(R)$.

In the following we shall restrict ourselves to the case where R is an interval $I^n = (a_1, b_1) \times \ldots \times (a_n, b_n)$ in $\mathbb{R}^n$. We also restrict ourselves to the case $n = 1$, the generalization to arbitrary n being straightforward.

Definition I.2:38. Let $I_1 = (a_1, b_1)$ and $I_2 = (a_2, b_2)$ be intervals in $\mathbb{R}^1$. We shall denote by $L(I_1, I_2, \varepsilon_1, \varepsilon_2)$ the set of functions f in $L^2(\mathbb{R}^1)$ such that $f \subset_{\varepsilon_1} I_1$ and $\hat{f} \subset_{\varepsilon_2} I_2$.

We shall use the notation phase-space for the set $\mathbb{R}^2 = \mathbb{R}^1 \times \mathbb{R}^1$ considered as the set of all pairs (x, y) where x is associated with the "ordinary space" variable for f ($f(x)$) and y is the "fouriertransformed variable", i.e. a variable for $\hat{f}$ ($\hat{f}(y)$). With this notation a function f in $L(I_1, I_2, \varepsilon_1, \varepsilon_2)$ can in a sense be said to be approximately localized to the phase-space region $I_1 \times I_2$.

Suppose $f \subset_{\varepsilon_1} I_1$ and $\hat{f} \subset_{\varepsilon_2} I_2$. Let M_2 be the subspace with $\hat{M}_2 = L^2(I_2)$. Then $\hat{f} \subset_{\varepsilon_2} \hat{M}_2$ and $f \subset_{\varepsilon_2} M_2$. Let P_1 be the projection on $M_1 = L^2(I_1)$ and $f_1 = P_1 f$. Then $f_1 \subset_0 I_1$ and by lemma I.1:28 it follows that $f_1 \subset_{\varepsilon_1+\varepsilon_2} M_2$. Thus $f_1 \in L(I_1, I_2, 0, \varepsilon_1+\varepsilon_2)$. Thus there is no essential restriction if we take one of the ε_1 or ε_2 to be $= 0$ or if we take $\varepsilon_1 = \varepsilon_2$. In the following we put $\varepsilon_1 = \varepsilon_2$. In view of I.2 (4) there is also no essential restriction to consider intervals $I_1 = (-A, A)$ and $I_2 = (-B, B)$ centered at the origin.

Definition I.2:39. We define $L_0(A, B, \varepsilon) = L(I_1, I_2, \varepsilon, \varepsilon)$ where $I_1 = (-A, A)$ and $I_2 = (-B, B)$.

The set $L_0(A, B, \varepsilon)$ of functions f, with simultaneous confidence estimates on f itself and on its Fourier transform $\hat{f}$, is not a linear space. To get a linear subspace of $\mathscr{H}$ approximating $L_0(A, B, \varepsilon)$, we introduce the following concept.

Definition I.2:40. Let $I_1 = (-A, A)$, $I_2 = (-B, B)$ be intervals in $\mathbb{R}^1$ and d_1, d_2, ε positive numbers. A closed subspace M (or the corresponding projection P on M) is said to be an (d_1, d_2, ε)-localization to the phase-space rectangle $I_1 \times I_2$ or shortly an $(A, B, d_1, d_2, \varepsilon)$-phase-space localization if

1) $f \subset_\varepsilon I_1$ and $\hat{f} \subset_\varepsilon I_2$ implies $f \subset_{4\varepsilon} M$.
2a) $M \subset_\varepsilon (-A - d_1, A + d_1)$
2b) $\hat{M} \subset_\varepsilon (-B - d_2, B + d_2)$

The factor 4 in 1) multiplying ε is chosen for convenience. It cannot however be chosen arbitrarily small.

1) can be expressed as

$$L_0(A, B, \varepsilon) \subset_{4\varepsilon} M$$

in the sense that $f \subset_{4\varepsilon} M$ for every f in $L_0(A,B,\varepsilon)$. Thus M approximates $L_0(I_1,I_2,\varepsilon)$ from above. On the other hand 2) means that

$$M \subset L_0(A+d_1, B+d_2, \varepsilon)$$

so M also approximates from below the $L_0(A',B',\varepsilon)$ corresponding to the larger intervals $(-A',A')$, $(-B',B')$ with $A' = A+d_1$, $B' = B+d_2$.

The approximation of $L_0(A,B,\varepsilon)$ by M:s are "single directed" in the sense that it is based on the nonsymmetric relation $\subset_\varepsilon$. A given $L_0(A,B,\varepsilon)$ can be approximated from above with (or covered by) an M but has to be approximated from below by another "smaller" M'.

Definition I.2:41. Let $M(A,B,d_1,d_2)$ and $G(A,B,d_1,d_2)$ be two functions of the real positive variables A, B, d_1, d_2.The values of M are closed subspaces and the values of G are positive numbers. The pair (M,G) is called a family of phase-space localizations if for every A, B, d_1, d_2 the subspace $M(A,B,d_1,d_2)$ is an $(A,B,d_1,d_2,\varepsilon)$-phase-space localization with $\varepsilon = G(A,B,d_1,d_2)$.

We shall now give a constructive example (M,G) of a family of phase-space localizations.

In the following the square root function $\mathrm{sqrt}(z) = z^{1/2}$ for a complex argument is always to be interpreted by its principal branch value. It will only be used for values of z in the open right halfplane $\mathrm{Re}\, z > 0$.

Lemma I.2:42. If $|\mathrm{Im}\, z| < b$ then
1° $\mathrm{Re}\,\mathrm{sqrt}(b^2+z^2) > |\mathrm{Re}\, z|$ and
2° $|\mathrm{Im}\,\mathrm{sqrt}(b^2+z^2)| < |\mathrm{Im}\, z|$.

Proof. Quite elementary. □

Definition I.2:43. For A, $a > 0$ we define the function

$$F_{A,a}(z) = \mathrm{sqrt}\int_{-A}^{A} \exp\left[-\mathrm{sqrt}\left(\frac{\pi^2}{4} + \frac{(z-u)^2}{a^2}\right)\right] du$$

of the complex argument $z = x + iy$.

Lemma I.2:44. $F_{A,a}(z)$ is uniquely defined and analytic for $|\mathrm{Im}\, z| < \pi a/2$.

Proof. Since $\mathrm{Re}\left[\pi^2/4 + (x+iy)^2/a^2\right] > 0$ for $|y| < \pi a/2$, the square root in the exponential and thus also in the integral is uniquely defined and analytic for $|\mathrm{Im}\, z| < \pi a/2$.

For $|y| < \pi a/2$ we have that

$$\left|\mathrm{Im}\,\mathrm{sqrt}\left[\pi^2/4 + (x+iy-u)^2/a^2\right]\right| < |y/a| < \pi/2$$

which shows that the real part of the integral is positive. Thus the square root of the integral is uniquely defined and analytic for $|\mathrm{Im}\, z| < \pi a/2$. □

Lemma I.2:45. $F_{A,a}$ satisfies

1° $|F_{A,a}(z)| < \mathrm{sqrt}(2a)$ if $|\mathrm{Im}\, z| < \pi a/2$.
2° $|F_{A,a}(z)| < \mathrm{sqrt}(a) \cdot \exp\big[-(|\mathrm{Re}\, z| - A)/2a\big]$ if $|\mathrm{Im}\, z| < \pi a/2$ and $|\mathrm{Re}\, z| > A$.

Proof. Suppose $|\mathrm{Im}\, z| < \pi a/2$. Then, by lemma I.2:42

$$\begin{aligned}|F_{A,a}(z)|^2 &= \Big|\int_{-A}^{A} \exp\big[-\mathrm{sqrt}[\pi^2/2 + (z-u)^2/a^2]\big]\, du\Big| \\ &< \int_{-A}^{A} \exp\big[-\mathrm{Re}\,\mathrm{sqrt}[\pi^2/4 + (z-u)^2/a^2]\big]\, du \\ &< \int_{-A}^{A} \exp(-|\mathrm{Re}\, z - u|/a)\, du < 2\int_0^{\infty} \exp(-u/a)\, du = 2a.\end{aligned}$$

If $|\mathrm{Re}\, z| > A$ we have the estimate

$$\begin{aligned}\int_{-A}^{A} \exp\big(-|\mathrm{Re}\, z - u|/a\big)\, du &< \int_{|\mathrm{Re}\, z|-A}^{\infty} \exp\big(-u/a\big)\, du \\ &= a \cdot \exp\big(-(|\mathrm{Re}\, z| - A)/a\big)\end{aligned}$$

and the result follows. □

Lemma I.2:46. The Fourier transform of $F_{A,a}$ satisfies

$$|\hat{F}_{A,a}(y)| < C \cdot \exp\big(-\pi a|y|/2\big)$$

where $C = \dfrac{2\sqrt{a}}{\sqrt{\pi}}\big(A + \sqrt{2}\cdot a\big)$.

Proof. According to lemmas I.2:44 and I.2:45 we can change the contour of integration to a line parallel to the real axis in the integral for $\hat{F}_{A,a}$. This gives

$$\begin{aligned}\hat{F}_{A,a}(y) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx} F_{A,a}(x)\, dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iy(x+iy_1)} F_{A,a}(x+iy_1)\, dx = \\ &= \frac{1}{\sqrt{2\pi}} e^{yy_1}\int_{-\infty}^{\infty} e^{-iyx} F_{A,a}(x+iy_1)\, dx\end{aligned}$$

for arbitrary y_1 with $|y_1| < \pi a/2$, thus

$$|\hat{F}_{A,a}(y)| \le \frac{1}{\sqrt{2\pi}} e^{-\pi a|y|/2}\int_{-\infty}^{\infty} |F_{A,a}(x+iy_1)|\, dx.$$

The last integral according to lemma I.2:45 can be estimated by

$$\int_{-A}^{A} \sqrt{2a}\, dx + 2\int_{A}^{\infty} \sqrt{a}\exp -(|x| - A)/2a\, dx < 2A\sqrt{2a} + 2\sqrt{a}\cdot 2a$$

and the result follows. □

Definition I.2:47. For every integer n we define the function

$$F_{A,a,n}(x) = e^{in\pi x/A} \cdot F_{A,a}(x).$$

We denote by $M_{A,a}$ the closed subspace of $L^2(\mathbb{R}^1)$ spanned by the functions $F_{A,a,n}$, $n = 0, \pm 1, \pm 2, \ldots$. If $I = [n_1, n_2]$ is an integer interval, we denote by $M_{A,a,I}$ the closed subspace spanned by all $F_{A,a,n}$ with $n_1 \le n \le n_2$.

Lemma I.2:48. The functions $F_{A,a,n}$, $n = 0, \pm 1, \pm 2, \ldots$ are orthogonal and their norms are given by

$$\|F_{A,a,n}\| = 2\sqrt{A \cdot a \cdot C_0}$$

where $C_0 = \int_0^\infty \exp\big(-\operatorname{sqrt}(\pi^2/4 + x^2)\big)\, dx$.

Proof. By substituting $x - u = y$ and interchanging the order of integration we obtain

$$\begin{aligned}\langle F_{A,a,m}, F_{A,a,n}\rangle &= \int_{-\infty}^{\infty} e^{i(m-n)\pi x/A} F_{A,a}(x)^2\, dx \\ &= \int_{-\infty}^{\infty} \left[e^{i(m-n)\pi x/A} \int_{-A}^{A} \exp\left[-\sqrt{\pi^2/4 + (x-u)^2/a^2}\right] du\right] dx \\ &= \int_{-\infty}^{\infty} e^{i(m-n)\pi y/A} \exp\left[-\sqrt{\pi^2/4 + y^2/a^2}\right] dy \cdot \int_{-A}^{A} e^{i(m-n)\pi u/A}\, du.\end{aligned}$$

The last integral vanishes when $m \neq n$. which proves the orthogonality.

For $m = n$ we obtain

$$\|F_{A,a,n}\|^2 = \int_{-\infty}^{\infty} \exp\Big(-\sqrt{\pi^2/4 + y^2/a^2}\Big)\, dy \cdot 2A = 4A \cdot a \cdot C_0$$

and $\|F_{A,a,n}\| = 2\sqrt{AaC_0}$. □

Theorem I.2:49. Suppose $d > 0$ and $0 < a < A$. Then

$$M_{A,a} \subset_\varepsilon L^2(-A - d, A + d)$$

with $\varepsilon = 3e^{-d/2a}$.

Proof. Let g be an arbitrary function in $M_{A,a}$. By definition I.2:47 and lemma I.2:48 g is of the form

$$g = \sum_n c_n F_{A,a,n} = F_{A,a} \cdot f_{\text{per}}$$

where $f_{\text{per}}(x) = \sum c_n e^{i\pi n x/A}$ is a periodic function with period $2A$ whose restriction f to $(-A, A)$ is an element in $L^2(-A, A)$. For the norms of g and f we get

$$\|g\|^2 = \sum_n |c_n|^2 \|F_{A,a,n}\|^2 = 4AaC_0 \sum |c_n|^2$$

C_0 as in lemma I.2:48 and

$$\|f\|^2 = \int_{-A}^{A} |f_{\text{per}}|^2\, dx = 2A \sum |c_n|^2,$$

thus

$$\|g\|^2 = 2aC_0\|f\|^2.$$

Using the periodicity of f_{per} and lemma I.2:45 we get

$$\begin{aligned}\int_{|x|>A+d} |g(x)|^2\,dx &= \sum_{k=1}^{\infty}\left\{\int_{d+(2k-1)A}^{d+(2k+1)A} |F_{A,a}(x)|^2\,|f_{\text{per}}(x)|^2\,dx\right.\\ &\quad\left.+\int_{-d-(2k+1)A}^{-d-(2k-1)A} |F_{A,a}(x)|^2\,|f_{\text{per}}(x)|^2\,dx\right\}\\ &\le 2a\sum_{k=1}^{\infty} e^{-\frac{d+2(k-1)A}{a}}\cdot\int_{-A}^{A}|f_{\text{per}}|^2\,dx = 2ae^{-d/a}\cdot\frac{1}{1-e^{-2A/a}}\cdot\|f\|^2\end{aligned}$$

and

$$\int_{|x|>A+d} |g(x)|^2\,dx\Big/\|g\|^2 \le \frac{1}{C_0}\cdot\frac{1}{1-e^{-2A/a}}\cdot e^{-d/a}.$$

The first factor can be estimated by

$$C_0 = \int_0^\infty e^{-\sqrt{\pi^2/4+x^2}} > \int_0^\infty e^{-(\pi/2+x)}\,dx = e^{-\pi/2}$$

and assuming $a < A$ we get

$$\frac{1}{C_0}\cdot\frac{1}{1-e^{-2A/a}} < e^{\pi/2}\cdot\frac{1}{1-e^{-2}} < 9.$$

Thus, for arbitrary g in $M_{A,a}$ we have

$$\operatorname{dist}(g, L^2(-A-d, A+d)) = \left\{\int_{|x|>A+d}|g|^2\,dx\right\}^{1/2}\Big/\|g\| < 3e^{-d/2a}$$

and the theorem is proved. □

Lemma I.2:50. Suppose $d > 0$, $0 < a < A$, $d < A$ and x real. Then

$$\frac{F_{A,a}(x)}{\sqrt{2aC_0}} < 1 \quad \text{for all } x$$

and

$$1 - \frac{F_{A,a}(x)}{\sqrt{2aC_0}} < \frac{1}{C_0}e^{-d/a} \quad \text{for } |x| < A-d,$$

with C_0 as in lemma I.2:48.

Proof. For any real x we have

$$F_{A,a}(x)^2 = \int_{-A}^{A} e^{-\sqrt{\frac{\pi^2}{4}+\frac{(x-u)^2}{a^2}}}\,du < \int_{-\infty}^{\infty} e^{-\sqrt{\frac{\pi^2}{4}+\frac{(x-u)^2}{a^2}}}\,du = a\int_{-\infty}^{\infty} e^{-\sqrt{\frac{\pi^2}{4}+y^2}}\,dy = 2aC_0$$

which proves the first statement.

For $|x| < A - d$ we have

$$\begin{aligned} F_{A,a}(x)^2 &= \int_{-\infty}^{\infty} e^{-\sqrt{\frac{\pi^2}{4}+\frac{(x-u)^2}{a^2}}}\,du - \int_{|u|>A} e^{-\sqrt{\frac{\pi^2}{4}+\frac{(x-u)^2}{a^2}}}\,du \\ &= 2aC_0 - \int_{|u|>A} e^{-\sqrt{\frac{\pi^2}{4}+\frac{(x-u)^2}{a^2}}}\,du \\ &> 2aC_0 - 2a\int_{d/a}^{\infty} e^{-\sqrt{\frac{\pi^2}{4}+y^2}}\,dy > 2aC_0 - 2a\int_{d/a}^{\infty} e^{-y}\,dy \\ &= 2aC_0 - 2ae^{-d/a}. \end{aligned}$$

Using the inequality $1-\varepsilon < \sqrt{1-\varepsilon}$ for $\varepsilon = \frac{1}{C_0}e^{-d/a} < 1$, the second inequality follows. (If $\varepsilon \geq 1$ the second inequality is trivial.) □

Theorem I.2:51. Suppose $0 < a, d' < A$. Then

$$L^2(-A+d',\, A-d') \subset_{\varepsilon'} M_{A,a}$$

with $\varepsilon' = 4e^{-d'/2a}$.

Proof. Let f be an arbitrary element in $L^2(-A+d',\, A-d')$ i.e. an element in $L^2(\mathbb{R}^1)$ which vanishes almost everywhere outside $(-A+d', A-d')$. Let f_1 be the restriction of f to $(-A, A)$ and f_{per} the periodic continuation with period $2A$ of f_1. Set $g = F_{A,a} \cdot f_{\text{per}}$ and $h = g/\sqrt{2aC_0}$, C_0 as in lemma I.2:48. From lemma I.2:50 and the proof of theorem I.2:49 follows that g and h are in $M_{A,a}$ and

$$\begin{aligned} \|f-h\|^2 &= \int_{|x|<A} |f-h|^2\,dx + \int_{|x|>A} |f-h|^2\,dx \\ &= \int_{|x|<A-d'} |f|^2 \left|1 - \frac{F_{A,a}(x)}{\sqrt{2aC_0}}\right|^2 dx + \int_{|x|>A+d'} |h|^2\,dx \\ &< \|f_1\|^2 \cdot \frac{1}{C_0}e^{-d'/a} + \frac{1}{C_0}\cdot\frac{1}{1-e^{-2}} \cdot e^{-d'/a}\|f_1\|^2 \\ &< e^{\pi/2}\left(1 + \frac{1}{1-e^{-2}}\right)\cdot e^{-d'/a}\|f_1\|^2 < 16e^{-d'/a}\cdot\|f_1\|^2 \\ &= 16e^{-d'/a}\|f\|^2. \end{aligned}$$

Thus for every f in $L^2(-A+d', A-d')$ there is an h in $M_{A,a}$ with $\|f-h\| < \varepsilon'\|f\|$, $\varepsilon' = 4e^{-d'/2a}$, and the theorem follows from lemma I.1:23. □

Theorem I.2:52. Let $I = [n_1, n_2]$ be an integer interval, $B_1 = \frac{\pi n_1}{A}$, $B_2 = \frac{\pi n_2}{A}$, $d'' > 0$ and $0 < a < A/10$. Then

$$\hat{M}_{A,a,I} \subset_{\varepsilon''} L^2(B_1 - d'', B_2 + d'')$$

with $\varepsilon'' = \frac{A}{a} e^{-\pi a d''/2}$.

Proof. Let f be an arbitrary function in $\hat{M}_{A,a,I}$. Then f is of the form

$$f(x) = \sum_{n=n_1}^{n_2} c_n \cdot \hat{F}_{A,a,n}(x) = \sum_{n=n_1}^{n_2} c_n \hat{F}_{A,a}\Big(x - \frac{n\pi}{A}\Big)$$

and

$$\|f\|^2 = 4aAC_0 \sum_{n=n_1}^{n_2} |c_n|^2$$

with C_0 as in lemma I.2:48.

Suppose that $x > B_2 + d''$. Then

$$|f(x)|^2 \le \sum_{n=n_1}^{n_2} |c_n|^2 \cdot \sum_{n=n_1}^{n_2} \Big|\hat{F}_{A,a}\Big(x - \frac{n\pi}{A}\Big)\Big|^2$$

and lemma I.2:46 gives

$$\begin{aligned} |f(x)|^2 &< \frac{\|f\|^2}{4aAC_0} \cdot C^2 \cdot \sum_{n=n_1}^{n_2} e^{-\pi a(x - n\pi/A)} \\ &\le \frac{\|f\|^2 C^2}{4aAC_0} \cdot e^{-\pi a(x - B_2)} \cdot \sum_{n=0}^{\infty} \Big(e^{-\pi^2 a/A}\Big)^n \\ &= \frac{\|f\|^2 C^2}{4aAC_0} \cdot e^{-\pi a(x - B_2)} \cdot \frac{1}{1 - e^{-\pi^2 a/A}} \end{aligned}$$

with C as in lemma I.2:46. Then

$$\int_{B_2 + d''}^{\infty} |f(x)|^2 \, dx / \|f\|^2 \le \frac{C^2}{4aAC_0} \cdot \frac{1}{\pi a} \cdot \frac{e^{-\pi a d''}}{1 - e^{-\pi^2 a/A}}.$$

Using the inequalities $\frac{1}{C_0} < e^{\pi/2}$ (see proof of theorem I.2:49), $a < A/10$ and

$\frac{1}{1-e^{-y}} < 2 \cdot \frac{1}{y}$ for $0 < y < 1$ with $y = \pi^2 a/A$, we get

$$C < \frac{2\sqrt{a}}{\sqrt{\pi}}\Big(A + \frac{\sqrt{2}A}{10}\Big),$$
$$\frac{1}{C_0} < 5,$$
$$\frac{C^2}{4aAC_0} \cdot \frac{1}{\pi a} \cdot \frac{1}{1-e^{-\pi^2 a/A}} < \frac{1}{4} \cdot \Big(\frac{A}{a}\Big)^2.$$

Thus, for any f in $\hat{M}_{A,a,I}$ we have

$$\big\{\operatorname{dist}\big(f, L^2(B_1 - d'', B_2 + d'')\big)\big\}^2$$
$$= \left\{ \int\limits_{x<B_1-d''} |f|^2\,dx + \int\limits_{x>B_2+d''} |f|^2\,dx \right\} / \|f\|^2 < 2 \cdot \frac{1}{4}\Big(\frac{A}{a}\Big)^2 e^{-\pi a d''} < \varepsilon''^2$$

with $\varepsilon'' = \frac{A}{a} e^{-\pi a d''}$ and the theorem is proved. □

For the cases $n_1 = -\infty$ or $n_2 = \infty$ we have the following slightly sharper result.

Theorem I.2:53. If $d'' > 0$, $0 < a < A/10$ and n is an integer then

$$\hat{M}_{A,a,(-\infty,n]} \subset_{\varepsilon''/2} L^2\Big(-\infty, \frac{\pi n}{A} + d''\Big)$$

and

$$\hat{M}_{A,a,[n,\infty)} \subset_{\varepsilon''/2} L^2\Big(\frac{\pi n}{A} - d'', \infty\Big)$$

with ε'' as in theorem I.2:52.

Proof. Follows from the proof of theorem I.2:52 □

Note that theorem I.2:52 follows from theorem I.2:53 by using lemma I.1:27.

Theorem I.2:54. If $d'' > 0$, $0 < a < A/10$ and n_1, n_2 are integers then

$$\hat{M}_{A,a,(-\infty,n_1]} \oplus \hat{M}_{A,a,[n_2,\infty)} \subset_{\varepsilon''} L^2\Big(-\infty, \frac{\pi n_1}{A} + d''\Big) \oplus L^2\Big(\frac{\pi n_2}{A} - d'', \infty\Big)$$

with ε'' as in theorem I.2:52.

Proof. The theorem is trivial if $n_1 \geq n_2 - 1$ and if $n_1 < n_2$ it follows from theorem I.2:53 by applying lemma I.1:26. □

Theorem I.2:55. Suppose $0 < a < A/10$ and $B, d', d'' > 0$. Let n be the smallest integer such that $n+1 \geq (A+d')(B+d'')/\pi$ and set

$$\varepsilon' = 4e^{-d'/2a} \quad \text{and}$$
$$\varepsilon'' = \frac{A+d'}{a} \cdot e^{-\pi a d''/2}.$$

Then, for any M and $\varepsilon_1, \varepsilon_2 > 0$ with $M \subset_{\varepsilon_1} L^2(-A, A)$ and $\hat{M} \subset_{\varepsilon_2} L^2(-B, B)$ we have

$$M \subset_{2\varepsilon_1+\varepsilon_2+2\varepsilon'+\varepsilon''} M_{A+d',a,[-n,n]}.$$

Proof. Suppose $M \subset_{\varepsilon_1} L^2(-A, A)$ and $\hat{M} \subset_{\varepsilon_2} L^2(-B, B)$. By theorem I.2:51 we see that $L^2(-A, A) \subset_{\varepsilon'} M_{A+d',a}$ and from theorem I.1:16 it then follows that $M \subset_{\varepsilon_1+\varepsilon'} M_{A+d',a}$ and $\hat{M} \subset_{\varepsilon_1+\varepsilon'} \hat{M}_{A+d',a}$. Let M_0 be the projection of $\hat{M}$ on $\hat{M}_{A+d',a}$. Then $\hat{M} \subset_{\varepsilon_1+\varepsilon'} M_0 \subset \hat{M}_{A+d',a}$ and $M_0 \subset_{\varepsilon_1+\varepsilon'} \hat{M}$ according to lemma I.1:25.

Set $M_1 = \hat{M}_{A+d',a,[-n,n]}$ and $M_2 = \hat{M}_{A+d',a,(-\infty,-n-1]} \oplus \hat{M}_{A+d',a,[n+1,\infty)}$. Then $M_1 \oplus M_2 = \hat{M}_{A+d',a}$ and

1°: $M_0 \subset M_1 \oplus M_2$, M_1 and M_2 orthogonal.

Set $M_1' = L^2(-B.B)$ and $M_2' = L^2(-\infty, -B) \oplus L^2(B, \infty)$. Then

2°: $M_1' \oplus M_2' = \mathscr{H} = L^2(\mathbb{R}^1)$, M_1' and M_2' orthogonal.

From $M_0 \subset_{\varepsilon_1+\varepsilon'} \hat{M} \subset_{\varepsilon_2} L^2(-B, B)$ we get

3°: $M_0 \subset_{\varepsilon_0} M_1'$, with $\varepsilon_0 = \varepsilon_1 + \varepsilon_2 + \varepsilon'$.

From theorem I.2:54 and the assumption on n we get since $\frac{\pi(n+1)}{A+d'} - d'' \geq B$ that

4°: $M_2 \subset_{\varepsilon''} M_2'$.

Applying lemma I.1:30 to 1°–4° we get

$$M_0 \subset_{\varepsilon_1+\varepsilon_2+\varepsilon'+\varepsilon''} M_1.$$

$\hat{M} \subset_{\varepsilon_1+\varepsilon'} M_0$ and theorem I.1:16 then gives

$$\hat{M} \subset_{2\varepsilon_1+\varepsilon_2+2\varepsilon'+\varepsilon''} \hat{M}_{A+d',a,[-n,n]}$$

and

$$M \subset_{2\varepsilon_1+\varepsilon_2+2\varepsilon'+\varepsilon''} M_{A+d',a,[-n,n]}.$$

□

Theorem I.2:56. Under the assumptions of theorem I.2:55, $M = M_{A+d',a,[-n,n]}$ is a (d_1, d_2, ε)localization to the phase-space rectangle $(-A, A)\times(-B, B)$ if we set $\varepsilon = 2\varepsilon' + \varepsilon''$, $d_1 = 2d'$ and $d_2 = 2d''$.

Proof. We have to prove 1), 2a) and 2b) of definition I.2:40. 1) and 2a) follow immediately from theorems I.2:55 and I.2:49 respectively.

By the definition of n we have $\frac{\pi n}{A+d'} < B + d''$ and theorem I.2:52 gives

$$\hat{M} \subset_{\varepsilon''} L^2\Big(-\frac{n\pi}{A+d'} - d'', \frac{n\pi}{A+d'} + d'' \Big) \subset L^2(-B-d_2, B+d_2)$$

which proves condition 2b). □

I.2.2 Finite Approximation Theory

Theorem I.2:56 can be considered as a confidence estimate formulation of the uncertainty principle. It says that only a finite number, approximately determined by the phase-space volume, of orthogonal wave functions can be localized to a given phase-space interval. It generalizes the "pure state" (one-dimensional) formulation to consider many-dimensional subspaces. And the confidence estimates are much sharper than the confidence estimates given by the Tchebycheff theorem applied to the standard deviations used in the Heisenberg formulation.

As a complement to this we observe that the sum $\varepsilon_1 + \varepsilon_2$ of ε_1 and ε_2 in theorem I.2:55 cannot be arbitrarily small. It is always greater than a positive, in the interesting cases very small, number depending only on the phase-space volume $A \cdot B$. This theorem can easily be proved in the case $\varepsilon_2 = 0$ by observing that, in this case, the minimum value of ε_1 can be obtained by the highest eigenvalue of a compact integral operator with a $\sin x/x$ kernel and then applying the Paley-Wiener theorem. The general case can be reduced to this case by applying the arguments following definition I.2:38. The details are left to the reader.

This result of a finite dimension coupled to a phase-space volume suggests a general confidence definition of entropy as (Boltzmans constant times) the logarithm of this dimension. The general increase of entropy can then be understood in terms of theorems I.1:19 and III.6:42 below.

The positive lower limits on the ε:s shows the importance of developing a systematic theory of finite approximations and approximately defined quantities which we shall call "finite approximation theory". It generalizes the ordinary concepts and techniques of ε:s and δ:s of classical analysis, where the ε:s are supposed to ultimately tend to zero in order to obtain exact and "infinite" continuous results.

In III.6 we shall give an example of approximate quantities – the p-values of an approximately equiangular sequence of projections. In II.7 we shall briefly indicate how a theory of approximate partial differential equations can be built on the basic concepts of the present theory.

In the approximation theory based on the concepts of the present theory we are generally not seeking the best but rather a sufficiently good approximation, where "sufficiently good" depends on the special problem under concern.Thus especially the confidence levels $1 - \varepsilon^2$ are merely to be considered as necessary technical means to give rigorous formulations of essential results. The ε:s themselves have no other interest than that they have to be sufficiently small to give sufficient accuracy to other essential quantities.

Chapter II

Causal Event Structures

The basic idea of the present theory is to build, on an underlying structure of wave-functions satisfying wave equations, a mathematical structure, which we shall call causal event structure. The intended interpretation of this structure should be obvious from the names ("event", "conditional probability", "compatible", "possible", "time-order", etc.) given to basic concepts in the structure and it is through this structure that a physical interpretation of the following mathematical theory can be established. From the mathematical point of view, these names should of course be considered formally as "working names" of concepts in an abstract deductive theory.

II.1 Stochastic Event Structures

To every element ("event") e in the set E_0 below, there is defined another element $-e$ in E_0 (with the intended interpretation "not e"). We introduce the notation $\underline{e}$ to mean any of the two elements e or $-e$. $-e$ is called the opposite event to e or alternatively the negation of e.

Definition II.1:1. By a stochastic event structure we shall mean a structure $S = (E_0, -, T, p)$ such that

1) E_0 is a set, whose elements will be called events. To every element e in E_0 is associated another distinct element $-e$ in E_0 such that $-(-e) = e$.

2) T is a set of ordered (m,n)-tuples $(e_1, \dots, e_m | e'_1, \dots, e'_n)$ $m, n = 1, 2, 3, \dots$ of events in E_0 such that if $(e_1, \dots, e_m | e'_1, \dots, e'_n)$ is in T, then so is $(e_1, \dots, e_m | \underline{e}''_1, \dots, \underline{e}''_p)$ where $(e''_1, \dots, e''_p)$ is any subsequence of $(e'_1, \dots, e'_n)$ and $\underline{e}''_j$ equals e'_j or $-e'_j$ for every $j = 1, 2 \dots, p$. The elements in T are called processes.

3) p is a function from T to the closed interval $[0, 1]$ such that

   a) if $(e_1, \dots, e_m | e'_1, \dots, e'_n)$ is in T, then

$$\sum p(e_1, \dots, e_m | \underline{e}'_1, \dots, \underline{e}'_n) = 1$$

where the sum is over all the 2^n choices $\underline{e}'_j = e'_j$ or $-e'_j$, $j = 1,2,\dots,n$. (We write $p(e_1,\dots,e_m|e'_1,\dots,e'_n)$ instead of $p((e_1,\dots,e_m|e'_1,\dots,e'_n))$ for the value of the function p for the argument $(e_1,\dots,e_m|e'_1,\dots,e'_n)$ in T.)

b) if $(e_1,\dots,e_m|e'_1,\dots,e'_n)$ is in T, $i_1,\dots,i_p$ is a subsequence of $1,\dots,n$, and $(e''_1,\dots,e''_p)$ the corresponding subsequence of $(e'_1,\dots,e'_n)$, then

$$p(e_1,\dots,e_m|e''_1,\dots,e''_p) = \sum(e_1,\dots,e_m|\underline{e}'_1,\dots,\underline{e}'_n)$$

where the sum is over all 2^{n-p} $\underline{e}'_1,\dots,\underline{e}'_n$ such that $\underline{e}'_{i_j} = e''_j$ for $j = 1,\dots,p$.

c) $p(e_1,\dots,e_m|e'_1,\dots,e'_n) = 0$ if $e'_i = -e'_j$ for some i, j.

d) if $(e_1,\dots,e_m|e'_1,\dots,e'_n,e''_1,\dots,e''_k)$ is in T and $p(e_1,\dots,e_m|e'_1,\dots,e'_n)$ nonzero, then $(e_1,\dots,e_m,e'_1,\dots,e'_n|e''_1,\dots,e''_k)$ is in T and

$$\begin{aligned}&p(e_1,\dots,e_m|e'_1,\dots,e'_n,e''_1,\dots,e''_k)\\&= p(e_1,\dots,e_m|e'_1,\dots,e'_n)\cdot p(e_1,\dots,e_m,e'_1,\dots,e'_n|e''_1,\dots,e''_k).\end{aligned}$$

e) if $(e_1,\dots,e_m|e'_1,\dots,e'_n)$ and $(e''_1,\dots,e''_k,e_1,\dots,e_m|e'_1,\dots,e'_n)$ are in T, then

$$p(e''_1,\dots,e''_k,e_1,\dots,e_m|e'_1,\dots,e'_n) = p(e_1,\dots,e_m|e'_1,\dots,e'_n).$$

A pair $(E_0,-)$ with E_0 and $-$ satisfying condition 1) of definition II.1:1 will generally be called an "event structure".

In a tuple of the form $(e_1,\dots,e_m|e'_1,\dots,e'_n)$ $(e_1,\dots,e_m)$ will be called the "premise", $(e'_1,\dots,e'_n)$ will be called the "outcome" and $p(e_1,\dots,e_m|e'_1,\dots,e'_n)$ will be called the "conditional probability" or shorter "p-value". It is sometimes necessary to consider tuples of this form, without assuming them to be elements in T. We shall call such tuples "formal processes".

An ordered tuple $(e_1,\dots,e_n)$ of events will sometimes be called a "course of events". For a process $(e_1,\dots,e_m|e'_1,\dots,e'_n)$ the premise $(e_1,\dots,e_m)$ will, in certain discussions, alternatively be called the "initial conditions" of the course of events $(e'_1,\dots,e'_n)$. We shall make no essential distinction between the n-tuple $(e_1,\dots,e_n)$ and the corresponding sequence $e_1,\dots,e_n$.

We shall in the following use the notation $E, E', E_1,\dots$ for ordered tuples (sequences) of events in E_0 in order to get shorter notations. Thus, e.g. if $E = (e_1,\dots,e_m)$, $E' = (e'_1,\dots,e'_n)$, $E'' = (e''_1,\dots,e''_r)$ and e is in E_0, we shall instead of $(e_1,\dots,e_m,e,e'_1,\dots,e'_n|e''_1,\dots,e''_r)$ use the notation $(E,e,E'|E'')$ and so on. We write $E_1 \subset E_2$ if E_1 is a subsequence of E_2.

We denote by $\underline{E}$ any of the 2^n sequences $\underline{E} = (\underline{e}_1,\dots,\underline{e}_n)$ where $E = (e_1,\dots,e_n)$.

An important feature of this structure is that initial conditions and outcomes are described by the same kind of concepts (events). The outcome of one process can therefore be used as initial conditions in another subsequent process. Similarly, the complete set of events in the initial conditions and outcome of one process can be considered as the outcome of another larger process. These considerations will be taken up in the next section.

The concept of stochastic event structure is more general than the common concept of "stochastic process", which considers times series of transitions in a given state space. Two mutually "incompatible" sequences $(e'_1, \ldots, e'_m)$ and $(e''_1, \ldots, e''_n)$ can each be a possible course of events. Thus we have not only the problem of studying probabilities of given events, but also the choice between mutually incompatible sequences of events is an important subject of consideration in the present theory. The definition of the concept of "compatibility" will be given in the next section.

Let $(E_0, -, T, p)$ be a given stochastic event structure.

Theorem II.1:2. If $(E|E')$ is in T and $E'' \subset E'$, then

$$p(E|E'') \geq p(E|E').$$

Proof. By condition 3b) of definition II.1:1 $p(E|E'')$ is a sum of nonnegative terms, one of which is $p(E|E')$. □

Theorem II.1:3. Multiple occurrences of the same event e in the outcome of a process does not change its p-value, i.e.

$$p(E|E_1, e, E_2, e, E_3) = p(E|E_1, e, E_2, E_3) = p(E|E_1, E_2, e, E_3)$$

Proof. By 3b) of definition II.1:1

$$p(E|E_1, e, E_2, E_3) = p(E|E_1, e, E_2, e, E_3) + p(E|E_1, e, E_2, -e, E_3)$$

and the last term equals zero according to 3c) of definition II.1:1. □

Theorem II.1:4. Let $S = (E_0, -, T, p)$ be a stochastic event structure and E'_0 a subset of E_0 which is closed under the operation $-$. Let T' be the set of all processes $(E|E')$ in T built from sequences E, E' of events in E'_0 and let $-'$ and p' be the restrictions of $-$ and p to E'_0 and T' respectively. Then $S' = (E'_0, -', T', p')$ is a stochastic event structure which will be called the restriction of S to E'_0 and will be denoted by $S|_{E'_0}$.

Proof. Conditions 1), 2), 3a)-e) of definition II.1:1 for S' follows immediately from the corresponding conditions for S. □

II.2 Compatibility of Events

We shall now consider sequences of events and processes as embedded in the outcome $(e'_1, \ldots, e'_n)$ of a larger process $(e_1, \ldots, e_m|e'_1, \ldots, e'_n)$ in a given stochastic event structure $(E_0, -, T, p)$.

The following theorem establishes the consistency in considering a process as embedded in the outcome of another process and the consistency in interpreting $p(e_1, \ldots, e_m|e'_1, \ldots, e'_n)$ as a "conditional probability".

Theorem II.2:5. Suppose that $(e_1,\dots,e_m|e'_1,\dots,e'_n)$ and $(e''_1,\dots,e''_s|e'''_1,\dots,e'''_t)$ are processes in T such that $(e_1,\dots,e_m,e'_1,\dots,e'_n)$ is a subsequence of $(e'''_1,\dots,e'''_t)$. Then

$$p(e''_1,\dots,e''_s|e_1,\dots,e_m,e'_1,\dots,e'_n) = p(e''_1,\dots,e''_s|e_1,\dots,e_m)\cdot p(e_1,\dots,e_m|e'_1,\dots,e'_n).$$

Proof. Suppose that $p(e''_1,\dots,e''_s|e_1,\dots,e_m)$ is nonzero. Then, by condition 3d) of definition II.1:1 $(e''_1,\dots,e''_s,e_1,\dots,e_m|e'_1,\dots,e'_n)$ is in T and condition 3e) of definition II.1:1 gives that

$$p(e''_1,\dots,e''_s,e_1,\dots,e_m|e'_1,\dots,e'_n) = p(e_1,\dots,e_m|e'_1,\dots,e'_n).$$

The result then follows from the second statement of condition 3d) of definition II.1:1.

If $p(e''_1,\dots,e''_s|e_1,\dots,e_m) = 0$, then condition 3b) of definition II.1:1 gives that also
$p(e''_1,\dots,e''_s|e_1,\dots,e_m,e'_1,\dots,e'_n) = 0$ so the result is true also in this case. □

It should be noted that property 3e) of definition II.1:1 is crucial for this result.

The p-value (conditional probability) of a given outcome $(e'_1,\dots,e'_n)$ is always defined relative to some premise $(e_1,\dots,e_m)$. With another premise, especially a larger one containing $(e_1,\dots,e_m)$, the p-value generally has a different value (except for the important case covered by 3e) of definition II.1:1). However, for the special case $p = 1$, the premise can be given a more absolute meaning according to the following theorem.

Theorem II.2:6. Suppose

1) $E' \subset E''$
2) There exists an E''' such that
   a) $(E'''|E'',E)$ is in T
   b) $p(E'''|E'')$ nonzero
3) $(E'|E)$ is in T and $p(E'|E) = 1$
4) $(E''|E)$ is in T

Then $p(E''|E) = 1$.

Proof. By assumptions 1), 2a) and condition 2) of definition II.1:1 $(E'''|E',E)$ is in T. By 3) and theorem II.2:5 $p(E'''|E',E) = p(E'''|E')$. By condition 3b) of definition II.1:1 $p(E'''|E',E) = \sum p(E'''|\underline{E}'',E)$ and $p(E'''|E') = \sum p(E'''|\underline{E}'')$ where the sums are over all $\underline{E}'' = (\underline{e}_1,\dots,\underline{e}_n)$, with $E'' = (e_1,\dots,e_n)$, such that $\underline{e}_i = e_i$ if e_i is in E'. But, since according to theorem II.1:2

$$p(E'''|\underline{E}'',E) \le p(E'''|\underline{E}''),$$

the preceding equalities are possible only if all inequalities become equalities and especially

$$p(E'''|E'',E) = p(E'''|E'').$$

Since, by assumption 2b) $p(E'''|E'')$ is nonzero, assumption 4) and theorem II.2:5 gives

$$p(E''|E) = p(E'''|E'', E)/p(E'''|E'') = 1.$$

□

Definition II.2:7. A sequence $(e_1, \dots, e_n)$ of events is called compatible if there exists a process $(e'_1, \dots, e'_m|e_1, \dots, e_n)$ in T with $(e_1, \dots, e_m)$ as outcome. A sequence is called possible if it is compatible and $p(e'_1, \dots, e'_m|e_1, \dots, e_n)$ is nonzero for some $(e'_1, \dots, e'_m)$.

Definition II.2:8. A set of sequences $(e'_1, \dots, e'_m)$, $(e''_1, \dots, e''_n)$, …is called compatible (possible) if there exists a compatible (possible) sequence E containing all $e'_1, \dots, e'_m, e''_1, \dots, e''_n, \dots$.

Definition II.2:9. A set A of (formal) processes $(E'_1|E''_1)$, $(E'_2|E''_2)$, … is called compatible if there exists a compatible sequence E containing $E'_1, E''_1, E'_2, E''_2, \dots$. The set A of (formal) processes is called (formally) possible if there exists a process $(E'''|E)$ in T, such that E contains all the (formal) processes $(E'_1|E''_1), \dots$ and $p(E'''|E'''')$ is nonzero for some subsequence E'''' of E which contains all the premises $E'_1, E'_2, \dots$.

Remark 1. With these definitions, condition 2) of theorem II.2:6 can be expressed as "$(E''|E)$ is formally possible".

Remark 2. If $(E|E')$ is formally possible, then, by theorem II.2:5, the property of theorem II.1:3 holds also for events in the premise E, i.e.

$$p(\dots, e, \dots, e, \dots | \dots) = p(\dots, e, \dots, , \dots | \dots) = p(\dots, , \dots, e, \dots | \dots)$$

(provided that the corresponding $(\dots | \dots)$ are elements in T).

Theorem II.2:10. A subsequence of a compatible (possible) sequence is compatible (possible).

Proof. An immediate consequence of condition 2) of definition II.1:1 and theorem II.1:2. □

Definition II.2:11. A (formal) process Q' is said to be a (formal) subprocess of a process $Q = (E|E')$ if it has one of the forms $Q' = (E, E'_1|E'_2)$ or $Q' = (E'_1|E'_2)$ with $(E'_1, E'_2) \subset$ some $\underline{E}'$, where E'_1 may be empty in the first case.

Theorem II.2:12. Let $Q = (e_1, \dots, e_m|e'_1, \dots, e'_n)$ be a process in the stochastic event structure $S = (E_0, -, T, p)$ and let E'_0 be the set containing

$$e_1, -e_1, \dots, e_m, -e_m, e'_1, -e'_1, \dots, e'_n, -e'_n.$$

Let T' be the set of all subprocesses of Q and let $-'$ and p' be the restrictions of $-$ and p to E'_0 and T' respectively. Then $S' = (E'_0, -', T', p')$ is a stochastic event structure.

Proof. One easily checks that 1), 2), 3a)–e) of definition II.1:1 for the structure S' follow directly from the construction and the corresponding conditions for the structure S. □

Definition II.2:13. The stochastic event structure S' of theorem II.2:12 is said to be spanned in S by Q. It will also be called the restriction of S to Q and be denoted by $S|_Q$. A stochastic event structure of this form is called simple.

A simple stochastic event structure S' is spanned (in itself, S') by one of its processes. The smallest structure containing a given process is given by the following theorem.

Theorem II.2:14. Let $Q = (E|E')$ be a process in a stochastic event structure $S = (E_0, -, T, p)$ and E'_0 be defined as in theorem II.2:12. Let T'' be the set of all subprocesses of Q of the form $Q'' = (E, E'_1|E'_2)$, $E'_1, E'_2 \subset$ some $\underline{E}'$ (E'_1 may be empty), with $p(E|E'_1)$ nonzero if E'_1 is nonempty, and let $-''$ and p'' be the restrictions of $-$ and p to E'_0 and T''. Then $S'' = (E'_0, -'', T'', p'')$ is a stochastic event structure.

Every stochastic event structure which is a substructure of S and contains Q contains S''.

Proof. The same as that of theorem II.2:12. The last statement follows from the construction and condition 3d) of definition II.1:1. □

Definition II.2:15. A stochastic event structure of the form S'' of theorem II.2:14 will be called the minimal simple stochastic event structure spanned by the process of Q.

Definition II.2:16. Let $Q = (E|E')$ be a process and $Q' = (E, E'_1|E'_2)$ or $Q' = (E'_1|E'_2)$, $E'_1, E'_2 \subset$ some $\underline{E}'$ and E'_1 nonempty, be a formal subprocess of Q. Q' is called nonsingular (singular) with respect to Q if $p(E|E'_1)$ is nonzero (zero). A subprocess of the form $(E|E'_1)$ is defined as nonsingular. A simple stochastic event structure S spanned by Q is called nonsingular if all processes in S are nonsingular with respect to Q.

A minimal simple stochastic event structure is nonsingular. A nonsingular subprocess of the form $(E'_1|E'_2)$, $E'_1 \subset$ some $\underline{E}'$, of a process $(E|E')$ is possible since E'_1 is a premise with $p(E|\tilde{E}')$ nonzero. A nonsingular subprocess to $(E|E')$ of the form $(E, E'_1|E'_2)$ is possible if $(E|E'_1, E'_2)$ is formally possible.

The restriction of a stochastic event structure to the set of all nonsingular subprocesses of a given process Q defines a nonsingular simple stochastic event structure spanned by Q. (One easily verifies conditions 1)–3) of definition II.1:1.)

Theorem II.2:17. In a minimal simple (or nonsingular simple) stochastic event structure spanned by a process $Q = (E|E')$, the p-values are uniquely determined by the p-values $p(E|\underline{E}')$ of the subprocesses to Q of the form $(E|\underline{E}')$.

Proof. Follows from condition 3d) of definition II.1:1 and theorem II.2:5. □

Theorem II.2:18. Let E_0 be the set containing the $2(m+n)$ elements

$$e_1, -e_1, \dots, e_m, -e_m, e'_1, -e'_1, \dots, e'_n, -e'_n$$

and let $-$ be the operation on E_0 interchanging e_i and $-e_i$ and e'_i and $-e'_i$. Let T be the set of all tuples $(E, E_1|E_2)$ where $E = (e_1, \dots, e_m)$, $(E_1, E_2) \subset$ some $\underline{E}'$ where $E' = (e'_1, \dots, e'_n)$ and E_1 may be empty, and T_0 be the subset of T of tuples of the form $(E|\underline{E}')$. Let p_0 be a nonnegative function on T_0 with

$$\sum_{\underline{E}'} p_0(E|\underline{E}') = 1.$$

Then there exists a subset T_1 of T containing T_0 and an extension p_1 of p_0 to T_1 such that $S = (E_0, -, T_1, p_1)$ is a minimal simple stochastic event structure spanned by $(E|E')$. T_1 and p_1 are uniquely determined by these conditions.

Proof. Let T_1 be the subset of T of all tuples $(E, E_1|E_2)$ such that either E_1 is empty or $p(E|\underline{E}')$ is nonzero for some $\underline{E}'$ containing E_1. We define

$$p_1(E|E_2) = \sum_{\underline{E}' \supset E_2} p_0(E|\underline{E}') \tag{1}$$

for elements in T_1 of the form $(E|E_2)$ (E_1 empty), where the sum is over all $\underline{E}'$ containing E_2. For elements in T_1 of the form $(E, E_1|E_2)$, E_1 non-empty, we define

$$p_1(E, E_1|E_2) = p_1(E|E_1, E_2)/p_1(E|E_1) \tag{2}$$

(this is possible since $p_1(E|E_1)$ is nonzero in this case). By (1) we have that $(E, E_1|E_2)$ with E_1 nonzero is in T_1 iff $p_1(E|E_1)$ is nonzero. We have to prove conditions 1)–3) of definition II.1:1 for S. 1) and 2) are obvious. The conditions 3c) and 3e) are empty in the present case and thus trivially satisfied.

By summing (1) over all $\underline{E}_2$ we get

$$\sum p_1(E|\underline{E}_2) = \sum_{\underline{E}_2} \sum_{\underline{E}' \supset \underline{E}_2} p_0(E|\underline{E}') = \sum_{\text{all } \underline{E}'} p_0(E|\underline{E}') = 1.$$

If E_1 is nonempty and $(E, E_1|E_2)$ is in T_1, we find by summing (2) over $\underline{E}_2$, that

$$\begin{aligned}
&\sum_{\underline{E}_2} p_1(E, E_1|\underline{E}_2) = \Big(\sum_{\underline{E}_2} p_1(E|E_1, \underline{E}_2) \Big) / p_1(E|E_1) \\
&= \Big(\sum_{\underline{E}_2} \sum_{\underline{E}' \supset (E_1, \underline{E}_2)} p_0(E|\underline{E}') \Big) / p_1(E|E_1) = \Big(\sum_{\underline{E}' \supset E_1} p_0(E|\underline{E}') \Big) / p_1(E|E_1) \\
&= p_1(E|E_1)/p_1(E|E_1) = 1.
\end{aligned}$$

Thus 3a) is proved for both types of elements in T_1.

To prove 3b), suppose $E_2 \subset E_2'$ and $(E, E_1|E_2')$ is in T_1. By (1)

$$p_1(E|E_1, E_2) = \sum_{\underline{E}' \supset (E_1, E_2)} p_0(E|\underline{E}') = \sum_{\underline{E}_2' \supset E_2} \sum_{\underline{E}' \supset (E_1, \underline{E}_2')} p_0(E|\underline{E}')$$
$$= \sum_{\underline{E}_2' \supset E_2} p_1(E|E_1, \underline{E}_2').$$

Putting E_1 empty or dividing by $p_1(E|E_1)$ we get

$$p_1(E, E_1|E_2) = \sum_{\underline{E}_2' \supset E_2} p_1(E, E_1|\underline{E}_2')$$

and 3b) is proved.

For an element in T_1 of the form $(E|E_2)$, condition 3d) coincides with the definition of T_1 and (2). If $(E, E_1|E_2, E_3)$ is in T_1 and $p(E, E_1|E_2)$ is nonzero, then, by (2),

$$\begin{aligned} p_1(E, E_1|E_2, E_3) &= p_1(E|E_1, E_2, E_3)/p_1(E|E_1) \\ &= p_1(E, E_1, E_2|E_3) \cdot p_1(E|E_1, E_2)/p_1(E|E_1) \\ &= p_1(E, E_1, E_2|E_3) \cdot p_1(E, E_1|E_2) \end{aligned}$$

and 3d) is proved also in this case.

Finally, the uniqueness of T_1 follows from definitions II.2:13 and II.2:15 and the uniqueness of p_1 follows from theorem II.2:17 □

Theorem II.2:19. Let $S = (E_0, -, T, p)$ be a minimal simple stochastic event structure spanned by a process $Q = (E|E')$ and let $T' \supset T$ be a set of nonsingular formal subprocesses of Q. A necessary and sufficient condition for that p can be extended to a p' on T' such that $S' = (E_0, -, T', p')$ is a nonsingular simple stochastic event structure is that

1) if $(E_1|E_2)$ is in T' and $E_3 \subset E_2$, then $(E_1|\underline{E}_3)$ is in T'
2) if $(E_1|E_2, E_3)$ is in T' with $(E_1, E_2, E_3) \subset$ some $\underline{E}'$ and $p(E|E_1, E_2)$ is nonzero, then $(E_1, E_2|E_3)$ is in T'
3) if $p(E|E_1, E_2)$ is nonzero and $(E_2|E_3)$ is in T', then $p(E, E_1, E_2|E_3) = p(E, E_2|E_3)$.

If such an extension p' of p exists, it is unique.

Proof. The necessity of 1) and 2) follows from conditions 2) and 3d) of definition II.1:1 (the second statement, $p(E|E_1, E_2)$ nonzero, of 2) implies $p(E_1|E_2)$ nonzero if S' is a stochastic event structure). If S' is a stochastic event structure, condition 3) follows from 3e) of definition II.1:1 by noting that $p(E_2|E_3)$ equals $p(E, E_2|E_3)$ and $p(E, E_1, E_2|E_3)$. The uniqueness of p' follows from theorem II.2:17.

We now prove the sufficiency.

A formal subprocess in T' not in T must be of the form $(E_1|E_2)$ with $(E_1, E_2) \subset$ some $\underline{E}'$. Define $p'(E_1|E_2) = p(E, E_1|E_2)/p(E|E_1)$ in this case ($p(E|E_1)$ is nonzero by the assumption of nonsingularity of the elements in T'). We have to check

the conditions 1)–3) of definition II.1:1 for S'. Condition 1) is obvious and condition 2) follows from assumption 1). To check condition 3) we first note that $p'(E_1|E_2) \leq 1$, since $p(E|E_1, E_2) \leq p(E|E_1)$ by theorem II.1:2. Condition 3a) for $p'(E_1|E_2)$ follows from condition 3b) of definition II.1:1 applied to $p(E|E_1)$ and $p(E|E_1, \underline{E}_2)$ and the definition of $p'(E_1|\underline{E}_2)$.

Similarly, by applying 3b) to $p(E|E_1, \underline{E}_2')$ and $p(E|E_1, \underline{E}_2)$, $\underline{E}_2' \subset E_2$ and dividing by $p(E|E_1)$ we obtain condition 3b) for $p'(E_1|E_2)$. Condition 3c) for $p'(E_1|E_2)$ follows from 3c) for $p(E|E_1, E_2)$.

The first part of 3d) follows from assumption 2). If $p(E|E_1, E_2)$ is nonzero, then, by condition 3d) for p,

$$p(E|E_1, E_2, E_3) = p(E|E_1, E_2) \cdot p(E, E_1, E_2|E_3).$$

Thus, by definition of p',

$$p(E, E_1, E_2|E_3) = p'(E_1, E_2|E_3).$$

Substituting this and dividing by $p(E|E_1)$ we get

$$p'(E_1|E_2, E_3) = p'(E_1|E_2) \cdot p'(E_1, E_2|E_3)$$

and condition 3d) is proved.

To prove condition 3e) for p', assume $(E_2|E_3)$ and $(E_1, E_2|E_3)$ are in T' and $(E_2, E_3) \subset$ some $\underline{E}'$. There are two cases, $E_1 \subset$ some $\underline{E}'$ or $E_1 = (E, E_1')$, $E_1' \subset$ some $\underline{E}'$. Suppose $E_1 \subset$ some $\underline{E}'$. By the preceding argument $p(E, E_1, E_2|E_3) = p'(E_1, E_2|E_3)$ and similarly $p(E, E_2|E_3) = p'(E_2|E_3)$. Assumption 3) then gives $p'(E_1, E_2|E_3) = p'(E_2|E_3)$ and 3e) for p' is proved in this case.

In the other case we have, by the same argument

$$\begin{aligned} p'(E_2|E_3) &= p(E, E_2|E_3) = p(E, E_1', E_2|E_3) = p(E_1, E_2|E_3) \\ &= p'(E_1, E_2|E_3) \end{aligned}$$

and 3e) for p' is proved also in this case. □

We conclude this section by noting that the study of any compatible (possible) system of events or processes can be reduced to the study of a simple (nonsingular simple) stochastic event structure.

II.3 Time-ordered Stochastic Event Structures

Definition II.3:20. A stochastic event structure with time is a structure $(E_0, -, T, p, t)$ such that $(E_0, -, T, p)$ is a stochastic event structure and t is a realvalued function on E_0 such that

1) $t(-e) = t(e)$ for every e in E_0.

2) If $(e_1, \ldots, e_m|e_1', \ldots, e_n')$ is in T, then $t(e_1) \leq \ldots \leq t(e_m) \leq t(e_1') \leq \ldots \leq t(e_n')$.

3) If $(e_1,\dots,e_m|e'_1\dots,e'_k,e'_{k+1},\dots,e'_n)$ is in T and $t(e'_{k+1}=t(e'_k))$, then

$$(e_1,\dots,e_m|e'_1,\dots,e'_{k+1},e'_k,\dots,e'_n) \text{ is in } T \text{ and}$$
$$p(e_1,\dots,e_m|e'_1,\dots,e'_k,e'_{k+1},\dots,e'_n)=p(e_1,\dots,e_m|e'_1,\dots,e'_{k+1},e'_k,\dots,e'_n)$$

4) If $(e_1,\dots,e_k,e_{k+1},\dots,e_m|e'_1,\dots,e'_n)$ is in T and $t(e_{k+1})=t(e_k)$, then

$$(e_1,\dots,e_{k+1},e_k,\dots,e_m|e'_1,\dots,e'_n) \text{ is in } T \text{ and}$$
$$p(e_1,\dots,e_k,e_{k+1},\dots,e_m|e'_1,\dots,e'_n)=p(e_1,\dots,e_{k+1},e_k,\dots,e_m|e'_1,\dots,e'_n)$$

$t(e)$ will be called the time of e. If $t(e_1)=t(e_2)$, e_1 and e_2 will be called simultaneous. A sequence $(e_1,\dots,e_n)$ is called time-ordered if $t(e_1)\le t(e_1)\le\dots\le t(e_n)$.

If E_1 and E_2 are two finite time-ordered sequences containing the same elements, then one of them can be obtained from the other by a finite number of interchanges of two consecutive simultaneous events. We shall say in this case that E_1 and E_2 are equivalent.

If $(E_1|E'_1)$ is a process and E_1 and E'_1 are equivalent to E_2 and E'_2 respectively, then by definition II.3:20, $(E_2|E'_2)$ is also a process which has the same p-value as $(E_1|E'_1)$. We shall say in this cae that $(E_1|E'_1)$ and $(E_2|E'_2)$ are equivalent.

Any finite set F of events can be ordered to a a time-ordered sequence, and this sequence is determined uniquely up to equivalence by F. Two sets F_1 and F_2 determines uniquely up to equivalence a process $(E_1|E_2)$ (if this is an element in T) where E_1 and E_2 are time-orderings of F_1 and F_2 respectively. This makes it consistent, when dealing with stochastic event structures with time, to introduce the notation $(F_1|F_2)$ for any of these processes and thus not distinguish between equivalent processes. Thus, e.g., we can write $(E|E_1,E_2)$ without assuming that (E_1,E_2) is time-ordered. The expression "$(E|E_1,E_2)$ is a process" thus means that time-ordering E and the union of E_1 and E_2 results in two sequences which constitute an element in T. t defines an order relation on E_0 which we shall denote by $|$, thus $e_1|e_2$ means $t(e_1)\le t(e_2)$. For sequences (or sets) we shall write $E'|E''$ if $e'|e''$ for all e' in E' and e'' in E''.

To get a general framework for a theory in concordance with the theory of relativity, we also need a more general concept of time-ordering in a stochastic event structure.

Definition II.3:21. By a "time-ordered event structure" we shall mean a structure $(E_0,-,|)$ such that $(E_0,-)$ is an event structure and $|$ is a relation $e_1|e_2$, called time ordering, between elements $e_!$, e_2 in E_0 such that

1) $e_1|e_2$ implies $\underline{e}_1|\underline{e}_2$.

The relation $|$ will generally not be supposed to be transitive. If F_1 and F_2 are sets of events we shall write $F_1|F_2$ if $e_1|e_2$ for all $e_1\in F_1$, $e_2\in F_2$.

If $F=\{e_1,\dots,e_n\}$ is a finite set of events we shall denote by $\underline{F}$ any of the 2^n sets $\{\underline{e}_1,\dots,\underline{e}_n\}$. The sets $\underline{F}$ are called alternatives to F.

Definition II.3:22. By a "general time-ordered stochastic event structure" we shall mean a structure $(E_0,-,T,p,|)$ such that

1) $(E_0, -, |)$ is a time-ordered event structure.

2) T is a set of pairs denoted $(F_1|F_2)$ where F_1 and F_2 are nonempty finite sets of events in E_0 such that

   a) $F_1|F_2$

   b) if $(F_1|F_2) \in T$ and $\emptyset \neq F_3 \subset F_2$, then $(F_1|\underline{F}_3) \in T$ for any alternative $\underline{F}_3$ to F_3.

3) p is a function from T to the closed interval $[0, 1]$ such that

   a) if $(F_1|F_2) \in T$, then $\sum p(F_1|\underline{F}_2) = 1$, where the sum is taken over all alternatives to F_2

   b) if $(F_1|F_2) \in T$ and $\emptyset \neq F_3 \subset F_2$, then

$$p(F_1|F_3) = \sum_{\underline{F}_2 \supset F_3} p(F_1|\underline{F}_2)$$

   where the sum is taken over all alternatives $\underline{F}_2$ to F_2 which coincide with F_3 on this subset.

   c) $p(F_1|F_2) = 0$ if F_2 contains both an event e and its opposite $-e$.

   d) if $(F_1|F_2 \cup F_3) \in T$, F_2, $F_3 \neq \emptyset$, $F_2|F_3$ and $p(F_1|F_2) > 0$, then

$$(F_1 \cup F_2|F_3) \in T \text{ and}$$
$$p(F_1|F_2 \cup F_3) = p(F_1|F_2) \cdot p(F_1 \cup F_2|F_3).$$

   e) if $(F_1 \cup F_2|F_3) \in T$, F_1, $F_2 \neq \emptyset$, $F_1|F_2$ and $(F_2|F_3) \in T$ then

$$p(F_1 \cup F_2|F_3) = p(F_2|F_3).$$

We shall mainly (except in sections II.5 and II.6) restrict ourselves to structures which are linearly time-ordered according to the following definitions.

Definition II.3:23. Let $(E_0, -, |)$ be a given time-ordered event structure. A sequence (ordered n-tuple) of events $(e_1, \ldots, e_n)$ is said to be time-ordered if $e_i|e_j$ for $i < j$. We shall say that a set $F = \{e_1, \ldots, e_n\}$ of events admits of a linear time-ordering if $(e'_1, \ldots, e'_n)$ is time-ordered for some permutation $(e'_1, \ldots, e'_n)$ of $(e_1, \ldots, e_n)$. In this case we shall say that $(e'_1, \ldots, e'_n)$ is a linear time-ordering of F. Two time-ordered sequences containing the same elements are called equivalent.

Definition II.3:24. By a linearly time-ordered stochastic event structure we shall mean a structure $(E_0, -, T, p, |)$ where $(E_0, -, T, p)$ is a stochastic event structure and $|$ is a relation between events such that

1) $(E_0, -, |)$ is a time-ordered event structure

2) if $(e_1, \ldots, e_m|e'_1, \ldots, e'_n) \in T$ then

   a) $e_i|e_j$ if $i < j$

   b) $e'_i|e'_j$ if $i < j$

c) $e_i|e'_j$ for all $i = 1, \ldots, m$, $j = 1, \ldots, n$,

3) If $(E_1|E_2) \in T$, E'_1 is a time-ordered permutation of E_1, and E'_2 is a time-ordered permutation of E_2, then $(E'_1|E'_2) \in T$ and $p(E'_1|E'_2) = p(E_1|E_2)$.
Two processes $(E_1|E_2)$ and $(E'_1|E'_2)$ satisfying condition 3) are called equivalent.

Theorem II.3:25. Let $S = (E_0, -, T, p, |)$ be a general time-ordered stochastic event structure satisfying the condition

1) if $(F_1|F_2) \in T$, then F_1 and F_2 admits of linear time-orderings.

Let T' be the set of all pairs $(E_1|E_2)$ where E_1 and E_2 are linear time-orderings of some F_1 and F_2 respectively with $(F_1|F_2) \in T$ and define $p'(E_1|E_2) = p(F_1|F_2)$ in this case. Then $S' = (E_0, -, T', p', |)$ is a linearly time-ordered stochastic event structure.

Proof. Conditions 2) and 3a)–e) of definition II.1:1 for the structure $(E_0, -, T', p')$ follow from the corresponding conditions 2b) and 3a)–e) of definition II.3:22 for the structure S. Thus $(E_0, -, T', p')$ is a stochastic event structure. Conditions 1)–3) of definition II.3:24 for the structure S' are obvious from the definition of T' and p'. □

Theorem II.3:26. Let $S' = (E_0, -, T', p', |)$ be a linearly time-ordered stochastic event structure. Let T be the set of all pairs $(F_1|F_2)$ such that F_1 and F_2 are finite sets of events and there exists a process $(E_1|E_2) \in T'$ with E_1 and E_2 linear time-orderings of F_1 and F_2 respectively and define $p(F_1|F_2) = p'(E_1|E_2)$ in this case. Then $S = (E_0, -, T, p, |)$ is a general time-ordered stochastic event structure.

Proof. The definition of $p(F_1|F_2)$ is consistent since different representations by time-ordered processes $(E_1|E_2)$ have the same p'-value according to condition 3) of definition II.3:24. Conditions 1), 2) and 3a)–c) of definition II.3:22 for the structure S follow from definition II.3:24 and conditions 1), 2), and 3a)–c) of definition II.1:1. To prove condition 3d) we note that F_2 and F_3, being subsets of $F_2 \cup F_3$ which admits of a linear time-ordering, can be represented by time-ordered sequences E_2 and E_3 respectively. $F_2|F_3$ then implies $E_2|E_3$. Thus the process $(F_1|F_2 \cup F_3)$ can be represented by $(E_1|E_2, E_3)$ where (E_2, E_3) is time-ordered. Condition 3d) of definition II.3:22 then follows from condition 3d) of definition II.1:1 for the structure $(E_0, -, T', p')$. Condition 3e) is proved by a similar argument. □

Lemma II.3:27. Let A be a set and $a_1 R a_2$ a relation between elements a_1, a_2 in A. Let $A' = (a'_1, \ldots, a'_n)$ and $A'' = (a''_1, \ldots, a''_n)$ be two finite sequences of elements in A containing the same elements and ordered with respect to R i.e. $a'_i R a'_j$ and $a''_i R a''_j$ if $i < j$, $i, j = 1, \ldots, n$. Then A'' can be obtained from A' by a sequence of permutations, each consisting in the interchange of two consecutive elements a_1, a_2 for which $a_1 R a_2$ and $a_2 R a_1$.

Proof. We prove it by induction over the number n of elements in A' (A''). The staement is obviously true if $n = 0, 1$ and 2. Suppose it is true for all subsets with

less than n elements. Let $A' = a_1, a'_1, \dots, a'_p, a_2, A_1$ and $A'' = a_2, a''_1, \dots, a''_q, a_1, A_2$ be two arbitrary ordered sequences of the same n elements, where we have chosen a notation to indicate where the first elements in the sequences a_1 and a_2 respectively occurs in the other sequence. (A_1 and A_2 denote sequences which need not be specified further.) By the ordering assumption $a_1 R a_2$, $a_2 R a_1$, $a_2 R a'_i$, $a'_i R a_2$, $i = 1, \dots, p$ and $a_1 R a''_j$, $a''_j R a_1$, $j = 1, \dots, q$. By successively interchanging a_2 with $a'_p, \dots, a'_1, a_1$ we obtain the (ordered) sequence $a_2, a_1, a'_1, \dots, a'_p, A_1$. The two sequences $a'_1, \dots, a'_p, A_1$ and $a''_1, \dots, a''_q, A_2$ are two ordered sequences of the same $n-2$ elements. Thus, by the induction assumption, the latter can be obtained from the former by successively permuting consecutive elements. This is then true also for the two sequences $a_2, a_1, a'_1, \dots, a'_p, A_1$ and $a_2, a_1, a''_1, \dots, a''_q, A_2$. Finally, by interchanging a_1 successively with $a''_1, \dots, a''_q$ we obtain $a_2, a''_1, \dots, a''_q, a_1, A_2$. Thus, the statement is true also for sets of n elements. □

Theorem II.3:28. Suppose that $(E_0, -, T, p)$ is a stochastic event structure and that $|$ is a relation between elements in E_0 satisfying 1) and 2) of definition II.3:24 and such that

1) If $(e_1, \dots, e_m | e'_1, \dots, e'_k, e'_{k+1}, \dots, e'_n)$ is in T and $e'_{k+1} | e'_k$, then $(e_1, \dots, e_m | e'_1, \dots, e'_{k+1}, e'_k, \dots, e'_n)$ is in T and

$$p(e_1, \dots, e_m | e'_1, \dots, e'_k, e'_{k+1}, \dots, e'_n) = p(e_1, \dots, e_m | e'_1, \dots, e'_{k+1}, e'_k, \dots, e'_n)$$

2) If $(e_1, \dots, e_k, e_{k+1}, \dots, e_m | e'_1, \dots, e'_n)$ is in T and $e_{k+1} | e_k$, then $(e_1, \dots, e_{k+1}, e_k, \dots, e_m | e'_1, \dots, e'_n)$ is in T and

$$p(e_1, \dots, e_k, e_{k+1}, \dots, e_m | e'_1, \dots, e'_n) = p(e_1, \dots, e_{k+1}, e_k, \dots, e_m | e'_1, \dots, e'_n).$$

Then $(E_0, -, T, p, |)$ is a linearly time-ordered stochastic event structure.

Proof. Conditions 1) and 2) of definition II.3:24 for $(E_0, -, T, p, |)$ is satisfied by assumption and condition 3) follows from assumptions 1) and 2) by applying lemma II.3:27. □

Theorem II.3:29. Let $(E_0, -, T, p, t)$ be a stochastic event structure with time and $e_1 | e_2$ be the relation $t(e_1) \leq t(e_2)$. Then $(E_0, -, T, p, |)$ is a linearly time-ordered stochastic event structure.

Proof. Conditions 1) and 2) of definition II.3:24 follow from conditions 1) and 2) of definition II.3:20. Condition 3) of definition II.3:24 follows from conditions 3) and 4) of definition II.3:20 by applying theorem II.3:28 (or from the discussion following definition II.3:20). □

II.4 Space and Spacetime Localization of Events

Definition II.4:30. By a stochastic event structure with time and instantaneous space localization we shall mean a structure $(E_0, -, T, p, t, \subset)$ such that

1) $(E_0, -, T, p, t)$ is a stochastic event structure with time.

2) $\subset$ is a relation $e \subset R$ between elements e in E_0 and subsets R of $\mathbb{R}^3$ such that

   a) $e \subset R$ implies $-e \subset R$

   b) $e \subset R \subset R'$ implies $e \subset R'$

   c) $e \subset R$ and $e \subset R'$ implies $e \subset R \cap R'$.

An event e with $e \subset R \subset \mathbb{R}^3$ and $t(e) = t$ can be considered as localized to the "space-time"-region $R' = R \times \{t\} \subset \mathbb{R}^3 \times \mathbb{R}^1 = \mathbb{R}^4$ and we shall write $e \subset R'$

Definition II.4:31. By space-time we shall mean $\mathbb{R}^4$ considered as $\mathbb{R}^3 \times \mathbb{R}^1$. For a point $X = (x,t) = \big((x_1, x_2, x_3), t\big) = (x_1, x_2, x_3, t)$ in space-time, $x = (x_1, x_2, x_3)$ will be called the space coordinates or components and t will be called the time component or simply the time. We define a time-ordering relation $|$ on $\mathbb{R}^4$ in two different ways:

a) "pure time-ordering": for $X' = (x', t')$ and $X'' = (x'', t'')$ in $\mathbb{R}^4$ we define $X'|X''$ to mean $t' \leq t''$

b) "relativistic time-ordering": for $X' = (x'_1, x'_2, x'_3, t')$ and $X'' = (x''_1, x''_2, x''_3, t'')$ in $\mathbb{R}^4$ we define $X'|X''$ to mean that not both

$$(x'_1 - x''_1)^2 + (x'_2 - x''_2)^2 + (x'_3 - x''_3)^2 - (t' - t'')^2 \leq 0$$

and

$$t' > t''$$

In both cases a) and b) we define $R'|R''$ where R' and R'' are subsets of $\mathbb{R}^4$, to mean that $X'|X''$ for any points X' in R' and X'' in R''.

Definition II.4:32. By a space-time localized stochastic event structure we shall mean a structure $(E_0, -, T, p, |, \subset)$ such that

1) $(E_0, -, T, p, |)$ is a general or linearly time-ordered stochastic event structure.

2) $\subset$ is a relation between elements e in E_0 and subsets R of $\mathbb{R}^4 = \mathbb{R}^3 \times \mathbb{R}^1$ such that

   a) $e \subset R$ implies $-e \subset R$

   b) $e \subset R \subset R'$ implies $e \subset R'$

   c) $e \subset R$ and $e \subset R'$ implies $e \subset R \cap R'$

3) $e_1|e_2$ iff $e_1 \subset R_1$ and $e_2 \subset R_2$ for some subsets R_1 and R_2 of $\mathbb{R}^4$ with $R_1|R_2$ where the latter $|$-relation is given by definition II.4:31 a) or b).

There should be no risk of confusion in using the same symbol $|$ for the relation $e_1|e_2$ and $R_1|R_2$.

By the argument following definition II.4:30 a stochastic event structure with time and instantaneous space localization can in a natural way be considered as a space-time localized stochastic event structure with pure time-ordering. On the

other hand, let $S = (E_0, -, T, p, |, \subset)$ be a linearly time-ordered space-time localized stochastic event structure with either pure or relativistic time-ordering. Let E_0' denote the subset of those e for which $e \subset R \times \{t\}$ for some $R \subset \mathbb{R}^3$ and t in $\mathbb{R}^1$. Then, by theorem II.1:4, $S' = (E_0', -', T', p')$ is a stochastic event structure. For e in E_0' and $R \subset \mathbb{R}^3$ we define $e \subset R$ and $t(e) = t$ if $e \subset R \times \{t\}$. If S has pure time-ordering, then obviously $(E_0', -', T', p', t, \subset)$ is a stochastic event structure with time and instantaneous space localization.

If S has relativistic time-ordering, let T' be the subset of those processes $(e_1, \ldots, e_m | e_1', \ldots, e_n')$ in T for which $t(e_1) \leq \ldots \leq t(e_m) \leq t(e_1') \leq \ldots \leq t(e_n')$. One easily finds, checking conditions 1), 2) 3a)–e) of definition II.1:1, that $(E_0', -, T', p)$ is a stochastic event structure ($-$ and p are here understood to mean restrictions to E_0' and T' respectively) and thus $(E_0', -, T', p, t, \subset)$ is a stochastic event structure with time and instantaneous space localization.

II.5 Initial and Boundary Conditions

Let $(E_0, -, T, p, |)$ be a given general time-ordered stochastic event structure.

Definition II.5:33. By a process with boundary conditions, we shall mean a tuple $(E_I; E_B \| E)$ of three sequences E_I, E_B, E such that

1) $E_I | E$ if E_I and E are nonempty

2) $(E_i, E_B | E)$ is formally possible

3) if $E' \subset E$, then, for every E'' for which $(E'' | E_I, E_B, E')$ is a process and $p(E'' | E_I, E_B)$ is nonzero and for every $\underline{E}' = (\underline{e}_1, \ldots, \underline{e}_n)$ where $E' = (e_1, \ldots, e_n)$ the quantity

$$p(E'' | E_I, E_B, \underline{E}')/p(E'' | E_I, E_B)$$

has one and the same value independent of E''. This value is denoted by $p(E_I; E_B \| \underline{E}')$. (We have adopted a notation in terms of sequences instead of the corresponding sets. Thus e.g. E_I, E_B, E' should be interpreted as $F_I \cup F_B \cup F'$ according to definition II.3:22. The definition of "formally possible" for a process in a general time-ordered stochastic event structure is quite analogous to definition II.2:9.)

In $E_I; E_B | E$ E_I will be called the initial conditions, E_B will be called the boundary conditions and E will be called the outcome.

We shall denote this new p-function with the same letter p as that in the structure $(E_0, -, T, p, t)$. There will be no risk for confusion since the two functions act on different types of arguments. It is natural to consider the function $p(E_I; E_B \| E)$ as an extension of the p in $(E_0, -, T, p, |)$.

Theorem II.5:34.

$$0 \leq p(E_I; E_B \| E) \leq 1.$$

Proof. An immediate consequence of the analogue for general time-ordered stochastic event structures of theorem II.1:2 □

It will be practical to have a notation for a process with boundary conditions in which the division into initial conditions and boundary conditions is not explicit. Let E_{IB} denote the combined sequence of initial and boundary conditions. E_{IB} can then be divided into an E_I containing all e in E_{IB} with $e|E$ and $E_B = E_{IB} - E_I$. This makes it consistent to use the notation $(E_{IB}\|E)$ for a process with "initial and boundary conditions" E_{IB} and outcome E. Note that the definition II.5:33, except for 1), actually does not depend on the division of $E_{IB} = (E_I, E_B)$ into E_I and E_B. If E_B happens to be empty, then the process $(E_{IB}\|E)$ can be identified with the process $(E_{IB}|E)$ if it happens to be an element in T. We shall also alternatively call $(E_1\|E)$ a generalized process. A process $(E_1|E)$ in T is a generalized process if $(E_1|E)$ is formally possible.

The generalized processes have properties similar to those of processes given in definition II.1:1.

It is obvious from definition II.5:33 that if $(E_1\|E)$ is a generalized process, then $(E_1\|\underline{E})$ where $\underline{E} = (\underline{e}_1, \ldots, \underline{e}_n)$, $E = (e_1, \ldots, e_n)$, is also a generalized process.

Theorem II.5:35. If $(E_1\|E)$ is a generalized process, then $\sum p(E_1\|\underline{E}) = 1$, where the sum is over all 2^n combinations $\underline{E} = (\underline{e}_1, \ldots, \underline{e}_n)$, where $E = (e_1, \ldots, e_n)$.

Proof. If $(E''|E_1, E)$ is in T and $p(E''|E_1)$ is nonzero, we have according to 3b) of definition II.3:22 that $p(E''|E_1) = \sum p(E''|E_1, \underline{E})$. Thus

$$\sum p(E_1\|\underline{E}') = \sum p(E''|E_1, \underline{E})/p(E''|E_1) = 1.$$

□

Theorem II.5:36. If $(E\|E_1)$ is a generalized process and $E_2 \subset E_1$, then $(E\|E_2)$ is also a generalized process and

$$p(E\|E_2) = \sum p(E\|\underline{E}_1)$$

where the sum is over all $\underline{E}_1 = (\underline{e}_1, \ldots, \underline{e}_n)$, where $E_1 = (e_1, \ldots, e_n)$, such that $\underline{e}_i = e_i$ if e_i is in E_2.

Proof. Suppose $(E\|E_1)$ is a generalized process. Since (E, E_1) is formally possible, $(E|E_2)$ is also formally possible, which proves condition 2) of definition II.5:33 for $(E\|E_2)$ to be a generalized process.

Suppose $E' \subset E_2$, $(E''|E, E')$ is a process (in T) and $p(E''|E)$ nonzero. We have to prove that $p(E''|E, \underline{E}')/p(E''|E)$ is independent of E''. But since $E' \subset E_1$ and $(E\|E_1)$ is a generalized process, this follows from the corresponding condition on $(E\|E_1)$ according to definition II.5:33. Thus $(E\|E_2)$ is a generalized process.

Choose E''' such that $(E'''|E, E_1)$ is a process and $p(E'''|E)$ nonzero. (This is possible since $(E\|E_1)$ is, by assumption, a generalized process.) Then, by 3b) of definition II.3:22

$$\begin{aligned} p(E\|E_2) &= p(E'''|E, E_2)/p(E'''|E) = \sum p(E'''|E, \underline{E}_1)/p(E'''|E) \\ &= \sum p(E\|\underline{E}_1) \end{aligned}$$

and the second statement of the theorem is proved. □

Theorem II.5:37. If $(E\|E')$ is a generalized process and $E'' \subset E'$, then

$$p(E\|E'') \geq p(E\|E').$$

Proof. By theorem II.5:36, $p(E\|E'')$ is a sum of nonnegative terms, one of which is $p(E\|E')$. □

Theorem II.5:38. If $(E\|E_1, E_2)$ is a generalized process and $p(E\|E_1)$ is nonzero, then $(E, E_1\|E_2)$ is also a generalized process and

$$p(E, E_1\|E_2) = p(E\|E_1, E_2)/p(E\|E_1).$$

Proof. Choose E'' such that $(E''|E, E_1, E_2)$ is a process and $p(E''|E)$ is nonzero. Then, $p(E''|E, E_1) = p(E''|E) \cdot p(E\|E_1)$ is nonzero and thus $(E, E_1|E_2)$ is formally possible. This proves condition 2) of definition II.5:33 for $(E, E_1\|E_2)$ to be a generalized process. To prove 3) of definition II.5:33, suppose $E' \subset E_2$, $(E'''|E, E_1, E')$ is a process and $p(E'''|E, E_1)$ nonzero. Since, by theorem II.5:36 $(E\|E_1, E')$ and $(E\|E_1)$ are generalized processes, we have

$$p(E'''|E, E_1, \underline{E}') = p(E'''|E) \cdot p(E\|E_1, \underline{E}') \text{ and}$$
$$p(E'''|E, E_1) = p(E'''|E) \cdot p(E\|E_1).$$

Thus,

$$p(E'''|E, E_1, \underline{E}')/p(E'''|E, E_1) = p(E\|E_1, \underline{E}')/p(E\|E_1)$$

is independent of E''' and $(E, E_1\|E_2)$ is a generalized process. Since the left member of the preceding equality equals $p(E, E_1\|\underline{E}')$, the last statement of the theorem follows if we put $\underline{E}' = E_2$. □

Definition II.5:39. We shall write $E \Longrightarrow E'$ if $(E\|E')$ is a generalized process and $p(E\|E') = 1$.

Theorem II.5:40. Suppose

1) $E_1 \subset E_2$

2) $(E_2|E)$ is formally possible

3) $E_1 \Longrightarrow E$

Then $E_2 \Longrightarrow E$.

Proof. Supose $E' \subset E$, $(E''|E_2, E')$ is a process and $p(E''|E_2)$ is nonzero. We have to show that $p(E''|E_2, \underline{E}')/p(E''|E_2)$ is independent of E'' and equals 1 if $\underline{E}' = E$.

By 3) and theorem II.5:37, $p(E''|E_1, E') = p(E''|E_1)$. By 3b) of definition II.3:22 $p(E''|E_1, E') = \sum p(E''|\underline{E}_2, E')$, $p(E''|E_1) = \sum p(E''|\underline{E}_2)$ where the sum is over all $\underline{E}_2 = (\underline{e}_1, \ldots, \underline{e}_n)$ where $E_2 = (e_1, \ldots, e_n)$ such that $\underline{e}_i = e_i$ if e_i is in E_1. But, since by the analogue for general time-ordered stochastic event structures of theorem II.1:2 $p(E''|\underline{E}_2, E') \leq p(E''|\underline{E}_2)$ this is possible only if $p(E''|\underline{E}_2, E') =$

$p(E''|\underline{E}_2)$ for all $\underline{E}_2$ and especially $p(E''|E_2, E') = p(E''|E_2)$. By 3b) of definition II.3:22, $p(E''|E_2) = \sum p(E''|E_2, \underline{E}')$ where the sum is over all $\underline{E}'$. Combining this with the preceding equality we find that all terms in the sum, except the one for which $\underline{E}' = E'$, equals zero. Thus the quotient $p(E''|E_2, \underline{E}')/p(E''|E_2)$ equals 1 if $\underline{E}' = E'$ and equals zero otherwise and is thus independent of E''. □

Theorem II.5:41. If $E \Longrightarrow E_1$ and $E_2 \subset E_1$, then $E \Longrightarrow E_2$.

Proof. A simple consequence of theorems II.5:34 and II.5:37. □

Theorem II.5:42. Suppose

1) $E \Longrightarrow E_1$

2) $E \Longrightarrow E_2$

3) $(E|E_1, E_2)$ is formally possible.

Then $E \Longrightarrow (E_1, E_2)$.

Proof. By 3) we can choose an E'' such that $(E''|E, E_1, E_2)$ is in T and $p(E''|E)$ is nonzero. Then, by 1),

$$p(E''|E, E_1) = p(E''|E) \cdot p(E\|E_1)$$

is nonzero, thus $(E, E_1|E_2)$ is formally possible. Theorem II.5:40 then gives that $(E_1, E_1') \Longrightarrow E_2$ if $E_1' \subset E_1$ and theorem II.5:41 gives that $(E, E_1') \Longrightarrow E_2'$ if $E_2' \subset E_2$.

Now suppose $E_1' \subset E_1$, $E_2' \subset E_2$, and E'' is an arbitrary sequence such that $(E''|E, E_1', E_2')$ is a process and $p(E''|E)$ is nonzero. We have to prove that

$$p(E''|E, \underline{E}_1', \underline{E}_2')/p(E''|E)$$

is independent of E'' and equals 1 when $\underline{E}_1' = E_1$ and $\underline{E}_2' = E_2$. By 1) and theorem II.5:41 $p(E''|E, E_1')/p(E''|E) = 1$ and thus $p(E''|E, E_1')$ is nonzero. Since $(E, E_1') \Longrightarrow E_2'$, we have $p(E''|E, E_1', E_2')/p(E''|E, E_1') = 1$. Combining these results we find $p(E''|E) = p(E''|E, E_1', E_2')$. By 3b) of definition II.3:22, $p(E''|E) = \sum p(E''|E, \underline{E}_1', \underline{E}_2')$ where the sum is over all $\underline{E}_1'$, $\underline{E}_2'$, and, in view of the preceding equality, this is possible only if all terms in the sum equal zero except the one with $\underline{E}_1' = E_1'$ and $\underline{E}_2' = E_2'$. Thus the quotient $p(E''|E, \underline{E}_1', \underline{E}_2')/p(E''|E)$ equals 1 if $\underline{E}_1' = E_1'$ and $\underline{E}_2' = E_2'$ and equals zero otherwise and is independent of E''. Thus $(E\|E_1, E_2)$ is a generalized process and putting $E_1' = E_1$ and $E_2' = E_2$ we find that $p(E\|E_1, E_2) = 1$. □

Theorem II.5:43. Suppose

1) $E_1 \Longrightarrow E_2$

2) $E_2 \Longrightarrow E_3$

3) $(E_1|E_2, E_3)$ is formally possible

4) $(E_1\|E_3)$ is a generalized process.

Then $E_1 \implies E_3$.

Proof. By 3) we can choose an E such that $(E|E_1, E_2, E_3)$ is in T and $p(E|E_1)$ is nonzero. Then, by 1), $p(E|E_1, E_2) = p(E|E_1)$ and thus $p(E|E_1, E_2)$ is nonzero, and $(E_1, E_2|E_3)$ is formally possible. By theorem II.5:40, and condition 2), $(E_1, E_2) \implies E_3$ and

$$p(E|E_1, E_2, E_3) = p(E|E_1, E_2).$$

Condition 4) and theorem II.5:37 then give

$$\begin{aligned} 1 \geq p(E_1\|E_3) &= p(E|E_1, E_3)/p(E|E_1) \\ &\geq p(E|E_1, E_2, E_3)/p(E|E_1) = 1, \end{aligned}$$

thus $p(E_1\|E_3) = 1$. □

II.6 An Example: Concordance of Measurements

As an example of application of the preceding formalism, we shall prove a simple theorem on "concordance of measurements".

Definition II.6:44. By a "proving registration" or "proving measurement" we shall mean a 3-tuple $(E_{IB}\|E_O; E_R)$ such that

1) $(E_{IB}\|E_O, E_R)$ is a generalized process, and

2) $(E_{IB}, E_R) \implies E_O$.

Alternatively we shall say that $(E_{IB}; E_R)$ is a proving registration (proving measurement) of E_O.

Definition II.6:45. By a "converse proving registration" or "converse proving measurement" or "forcing registration" we shall mean a 3-tuple $(E_{IB}\|E_O; E_R)$ such that

1) $(E_{IB}\|E_O, E_R)$ is a generalized process, and

2) $(E_{IB}, E_O) \implies E_R$.

Alternatively we shall say that $(E_{IB}; E_R)$ is a converse proving registration (measurement) of E_O.

Definition II.6:46. A triple $(E_{IB}\|E_O; E_R)$ is called a double-directed or complete measurement (registration) if it contains both a proving measurement and a converse proving measurement.

Theorem II.6:47. If
1) $(E'_{IB}\|E'_O; E'_R)$ and
2) $(E''_{IB}\|E''_O; E''_R)$
are proving measurements, and
3) $(E'_{IB}, E''_{IB}, E'_R, E''_R, E'_O, E''_O)$

is possible, then

$$(E'_{IB}, E''_{IB}, E'_R, E''_R) \Longrightarrow (E'_O, E''_O).$$

If in addition

4) $(E'_{IB}, E''_{IB} \| E'_O, E''_O, E'_R, E''_R)$

is a generalized process, then

$$(E'_{IB}, E''_{IB} \| E'_O, E''_O; E'_R, E''_R)$$

is a proving measurement.

Proof. By 3) $(E'_{IB}, E''_{IB}, E'_R, E''_R | E'_O)$ is formally possible. 1) and theorem II.5:40 then gives that

$$(E'_{IB}, E''_{IB}, E'_R, E''_R) \Longrightarrow E'_O.$$

Similarly, we get

$$(E'_{IB}, E''_{IB}, E'_R, E''_R) \Longrightarrow E''_O.$$

By 3) also $(E'_{IB}, E''_{IB}, E'_R, E''_R | E'_O, E''_O)$ is formally possible and theorem II.5:42 gives

$$(E'_{IB}, E''_{IB}, E'_R, E''_R) \Longrightarrow (E'_O, E''_O).$$

This proves the first statement of the theorem and also condition 2) of definition II.6:44 for the second staement. Since condition 1) of definition II.6:44 equals assumption 4) of the theorem, the second statement follows. □

If we put $E'_O = \{e\}$, $E''_O = \{-e\}$ we find, since (E'_O, E''_O) is impossible, that the event that both measurements 1) and 2) give affirmative result is impossible, according to definition II.2:7.

II.7 Deterministic Event Structures and Approximate Partial Differential Equations

A special case of stochastic event structure is obtained if all the probabilities $p(\dots | \dots)$ have value 1 or 0. We shall call such a structure a deterministic event structure.

The preceding discussion of initial and boundary conditions generalizes the common initial and boundary value problem of hyperbolic differential equations. Let e.g.

$$f''_{tt} = F(f, f'_x, \dots, f''_{xx}, \dots)$$

be the basic equation of a relativistic field theory, where F is a function of f and its space derivatives. Then the values of f on some spacelike surface $R(t)$, $T_1 < t < T_2$, is uniquely determined by f and f'_t on any spacelike surface $S(t_0)$, $t_0 < T_1$ which encompasses the backward light-cone projection of $R(t)$ on $S(t_0)$. We can eliminate f'_t if we know f at two such surfaces.

The field can be described by "events" (statements on f) e_i of the form

$$|f(t, x_i) - a_i| \leq b_i, \, i = 1, 2, \dots$$

By using suitable regularity assumptions or a priori estimates on the space derivatives of f we see that a finite set of such events describes f to a sufficient approximative accuracy on the surface $R(t)$. From the hyperbolic equation we can then derive expressions $e_i = G_i(e'_{i1}, \dots)$ which gives the e_i:s as uniquely determined by suitable $e'_{i1} \dots$ in the initial condition region R_I. Defining formally

$$p(e'_{i1}, \dots | e_i) = 1$$

we then get a deterministic event structure.

The discussion of boundary conditions in the preceding section means, for the case of classical field theories, that the boundary conditions on the boundary of R_S should in principle be derived from the combined set of field equations describing the system and its environment.

The concept of stochastic event structure to be used for general quantum mechanics, generalizes the classical initial and boundary value problem in two senses. Firstly, it is a generalization to stochastic transitions of which deterministic evolutions can be considered as special cases. Secondly, it describes boundary conditions in some space region surrounding the system rather than as values on a boundary surface. This is a consequence of the fact that quantum systems, due to the uncertainty principle, always have a finite extension in space. However, if we derive from the underlying quantum mechanics an approximate differential equation describing some macroscopic field, then this boundary region may be approximated by a surface.

Initial and Boundary Conditions

The basic problem in any physical situation is to describe the behaviour of a system under given initial and boundary conditions. In a stochastic event structure with the events equipped with time ordering and space localization, we can give a general precise formulation of this problem.

Let the system under consideration be confined to the space region $R(t)$ (which may depend on time) and suppose that we are interested in studying the system during the time interval (T_1, T_2). The behaviour of the system will be described by a series of events $e_1, \dots, e_n$ localized within the space-time region R_S defined by $R(t)$ with t varying in the interval (T_1, T_2) (see fig. 1). The initial conditions will be described by a series of events $e'_1, \dots, e'_m$, all associated to times before T_1. The behaviour of the system can then be predicted by studying processes of the form

$$(e'_1, \dots, e'_m | e_1, \dots, e_n) \tag{1}$$

Although the behaviour of the system can be predicted if we take into account initial conditions in a sufficiently large space-time region R_I, we often want to separate out the course of events in the environment and treat this as given boundary conditions. The course of events in the environment – boundary in a general sense – is described by a series of events $e''_1, \dots, e''_p$ localized in a suitable space-time region R_B during the time interval (T_1, T_2). With a sufficiently large set

$e'_1, \ldots, e'_m$ we then have a process say (compare fig. 1)

$$(e'_1, \ldots, e'_m | e_1, e''_1, e_2, {}' e''_2, \ldots) \tag{2}$$

describing a possible course of events of the system under concern and its boundary.

It can now be the case that the course of events $e''_1, \ldots, e''_p$ in the boundary region is correlated to the initial conditions in such a way that it can be used to replace some of the information given by the large set $e'_1, \ldots, e'_m$ of initial conditions. Thus we can, from the underlying process (2), construct and derive the properties of generalized kinds of processes

$$(e'''_1, \ldots, e'''_q; e''_1, \ldots, e''_p \| e_1, \ldots, e_n) \tag{3}$$

where $e'''_1, \ldots, e'''_q$ is a smaller set of initial conditions (e.g. a subset of $e'_1, \ldots, e'_m$) compensated by the series of boundary conditions $e''_1, \ldots, e''_p$. The mathematical formulation and derivation of boundary conditions in a quantum-mechanical stochastic event structure will be the central question in any application of this theory. It has to be worked out in concrete form separately for every special class of problems, by using the concrete form of the underlying dynamics described by the Schrödinger equation.

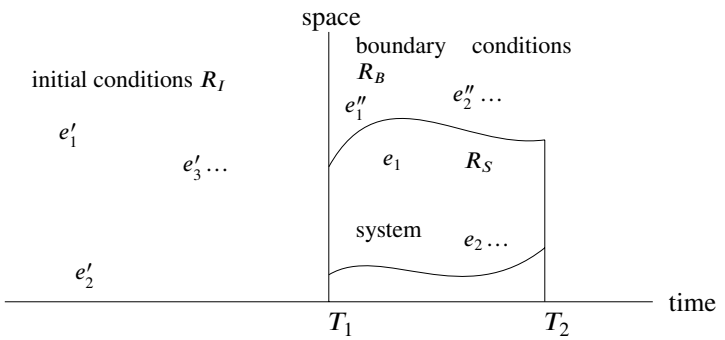


fig. 1

II.8 Approximation of Stochastic Event Structures

Lemma II.8:48. Let $(S_1 = (E_0, -, T_1, p_1))$ and $S_2 = (E_0, -, T_2, p_2)$ be two non-singular simple stochastic event structures spanned by $(E|E')$, $E' = (e'_1, \ldots, e'_n)$. Suppose

1) $T_1 \subset T_2$

and for every $\underline{E}'$

2) $|p_1(E|\underline{E}') - p_2(E|\underline{E}')| \leq \varepsilon_0$

and

3) either $p_1(E|\underline{E}') = 0$ or $p_1(E|\underline{E}') \geq p_0$, $p_0 > 0$.

Then

1° $|p_1(E|E_1) - p_2(E|E_1)| \leq 2^{n-1}\varepsilon_0$ for every subprocess in T_1 of the form $(E|E_1)$, $E_1 \subset$ some $\underline{E}'$

2° $|p_1(E_1|E_2) - p_2(E_1|E_2)| \leq \varepsilon = \frac{2^n}{p_0}\varepsilon_0$ for every $(E_1|E_2)$ in T_1.

If $p_1(E|\underline{E}') = p_2(E|\underline{E}')$ except for at most m values of $\underline{E}'$, then the right members of 1° and 2° can be replaced by $m\varepsilon_0$ and $2m\varepsilon_0/p_0$ respectively.

If S_1 is a minimal simple stochastic event structure and $p_0 > \varepsilon_0$, then the assumption $T_1 \subset T_2$ is unnecessary (follows from 2) and 3)).

Proof. For an element in T_1 of the form $(E|E_1)$ we get, by 3b) of definition II.1:1

$$p_i(E|E_1) = \sum_{\underline{E}' \supset E_1} p_i(E|\underline{E}_1), \quad i = 1,2 \tag{1}$$

and

$$|p_1(E|E_1) - p_2(E|E_1)| = \Big| \sum_{\underline{E}' \supset E_1} \varepsilon(\underline{E}') \Big|$$

where $\varepsilon(\underline{E}') = p_1(E|\underline{E}') - p_2(E|\underline{E}')$, $|\varepsilon(\underline{E}')| \leq \varepsilon_0$ and $\sum_{\text{all } \underline{E}'} \varepsilon(\underline{E}') = 0$. Since

$$s = \Big| \sum_{\underline{E}' \text{ in } F} \varepsilon(\underline{E}') \Big| = \Big| \sum_{\underline{E}' \text{ not in } F} \varepsilon(\underline{E}') \Big|$$

where one of the sums contains at most 2^{n-1} terms, we have $s \leq 2^{n-1} \cdot \varepsilon_0$ and thus $|p_1(E|E_1) - p_2(E|E_1)| \leq 2^{n-1}\varepsilon_0$. If $p_1(E|E_1) = p_2(E|E_1)$ except for at most m values of $\underline{E}'$, we can obviously replace the factor 2^{n-1} by m.

For an element in T_1 of the form $(E, E_1'|E_2')$, E_1' nonempty, we have $p_i(E, E_1'|E_2') = p_i'/p_i''$ where $p_i' = p_i(E|E_1', E_2')$, $p_i'' = p_i(E|E_1')$, $i = 1,2$. Using the identity

$$\frac{p_1'}{p_1''} - \frac{p_2'}{p_2''} = \frac{1}{p_1''}\left[(p_1' - p_2') + \frac{p_2'}{p_2''}(p_2'' - p_1'')\right] \tag{2}$$

$p_2'/p_2'' \leq 1$ and $1/p_1'' \leq 1/p_0$ by (1) and assumption 3), we get

$$\big|p_1(E, E_1'|E_2') - p_2(E, E_1'|E_2')\big| \leq 2 \cdot \frac{1}{p_0} \cdot 2^{n-1}\varepsilon_0 = \frac{2^n}{p_0} \cdot \varepsilon_0.$$

If $(E_1'|E_2')$ is in T_1 with $E_1' \subset$ some $\underline{E}'$, then, by 4e) of definition II.1:1, $p_i(E_1'|E_2') = p_i(E, E_1'|E_2')$ so 2° is satisfied also in this case.

Suppose S_1 is a minimal simple stochastic event structure. If $p_0 > \varepsilon_0$, then $p_1(E|E_1')$ nonzero implies $p_1(E|\underline{E}')$ nonzero for some $\underline{E}' \supset E_1'$, and by assumptions 2) and 3), $p_2(E|\underline{E}')$ nonzero, i.e. $p_2(E|E_1')$ nonzero. $T_1 \subset T_2$ follows from this using theorem II.2:14 □

Remark. When we approximate the p-values in a stochastic event structure by values differing less than a given small ε, we will consider $1-\varepsilon$ as a "confidence level". This means that p-values differing less than ε are considered as equal ("indistinguishable"), p-values less than ε are considered as zero, i.e. probabilities of impossible outcomes and p-values greater than the confidence level $1-\varepsilon$ are considered probabilities of certain outcomes. This also means that p:s which are supposed to be nonzero in theorems on stochastic event structures must, when approximation of p:s are considered, in critical cases be supposed to be significantly different from zero i.e. greater than some $p_0 > \varepsilon$.

Lemma II.8:49. Let $S = (E_0, -, T, p)$ be a nonsingular simple stochastic event structure spanned by $(E|E')$, $E' = (e'_1, \dots, e'_n)$, satisfying

1) for every $\underline{E}'$ either $p(E|\underline{E}') \geq p_0$ or $p(E|\underline{E}') \leq \varepsilon$, where $p_0 > 2(2^n - 1)\varepsilon$.

Then there exists a minimal simple stochastic event structure $S' = (E_0, -, T', p')$ spanned by $(E|E')$ satisfying

1° $T' \subset T$

2° $(E, E_1|E_2)$ is in T' if $p(E|E_1) \geq p_0$

3° for every $E_1 \subset$ some $\underline{E}'$ either $p'(E|E_1) = 0$ or $p'(E|E_1) \geq p'_0 = p_0 - (2^n - 1)\varepsilon$

4° $\left|p'(E|\underline{E}') - p(E|\underline{E}')\right| \leq (2^n - 1)\varepsilon$ for every $\underline{E}'$.

Proof. Choose $p'(E|\underline{E}') = 0$ if $p(E|\underline{E}') \leq \varepsilon$ and change the other $p(E|\underline{E}')$:s by a small amount to a $p'(E|\underline{E}')$ such that $\sum_{\text{all } \underline{E}'} p'(E|\underline{E}') = 1$. This can obviously be done in such a way that 4° is satisfied. Let S' be determined by p' according to theorem II.2:18. 3° is satisfied by construction and 1° follows from lemma II.8:48 (since $p_0 - 2(2^n - 1)\varepsilon > 0$). If $p(E|E_1) \geq p_0$, then, by 4°, $p'(E|E_1) \geq p_0 - (2^n - 1)\varepsilon > 0$ and 2° follows from 2d) of definition II.1:1. □

Chapter III

Equiangular Sequences of Projections

As in chapter I.1 we denote by $M, M_1, M', \ldots$ closed subspaces and by $P, P_1, P', \ldots$ projections in a given complex separable Hilbert space H. With projections (or projection operators) we always mean orthogonal projections. We shall denote by $\underline{P}$ any of the two operators P and $P^c = 1 - P$. Subspaces which are not necessarily closed will be denoted by $N, N_1, N', \ldots$.

III.1 P-equiangular Projections

Definition III.1:1. Let P be a projection and N be a nonzero (not necessarily closed) subspace. P is called equiangular with respect to N (or N-equiangular) if

$$\|Pu\|/\|u\| = \|Pu'\|/\|u'\|$$

for any nonzero u, u' in N, and the common value of $\|Pu\|^2/\|u\|^2$ is denoted by $p(N\,|\,P)$. If $N = 0$, any P is N-equiangular and we define $p(N\,|\,P) = 0$.

To distinguish from ε-equiangularity (approximate equiangularity) defined below we shall sometimes say exactly equiangular or 0-equiangular.

Obviously $0 \leq p(N\,|\,P) \leq 1$ and $p(N\,|\,P) = 1$ iff $0 \neq N \subset R(P)$. $p(N\,|\,P_2) = 0$ iff $N \subset R(P)^c = R(P^c)$.

Definition III.1:2. Let P_1 and P_2 be projections. P_2 is called P_1-equiangular if P_2 is M_1-equiangular with $M_1 = R(P_1)$ and we write $p(P_1\,|\,P_2)$ for $p(M_1\,|\,P_2)$

Theorem III.1:3. If P is N-equiangular, then P^c is also N-equiangular and if N is nonzero, then

$$p(N\,|\,P^c) = 1 - p(N\,|\,P).$$

Proof. If P is N-equiangular, then

$$\frac{\|P^c u\|}{\|u\|} = \left[1 - \frac{\|Pu\|^2}{\|u\|^2}\right]^{1/2} = \left[1 - p(N\,|\,P)\right]^{1/2}$$

is independent of u in N, u nonzero. □

Lemma III.1:4. If u, u' are vectors in H and $\|u+au'\|^2 = \|u\|^2 + \|u'\|^2$ for every complex number a with $|a| = 1$, then u and u' are orthogonal.

Proof. From

$$\begin{aligned}\|u\|^2 + \|u'\|^2 = \|u+au'\|^2 &= \langle u+au', u+au'\rangle \\ &= \|u\|^2 + \|u\|^2 + 2\,\mathrm{Re}\left[a\langle u,u'\rangle\right]\end{aligned}$$

follows $\mathrm{Re}\left[a\langle u,u'\rangle\right] = 0$ for every a with $|a| = 1$. Thus $\langle u,u'\rangle = 0$ and u and u' are orthogonal. □

Theorem III.1:5. If u, u' are orthogonal vectors in N and P is N-equiangular, then Pu and Pu' are orthogonal.

Proof. Suppose P is N-equiangular, u, u' orthogonal in $N \neq 0$ and set $p = p(N\,|\,P)$. (The theorem is trivial if $N = 0$.) Then, for arbitrary complex a with $|a| = 1$ we have

$$\|Pu + aPu'\|^2 = \|P(u+au')\|^2 = p\|u\|^2 + p\|u'\|^2 = \|Pu\|^2 + \|Pu'\|^2.$$

Lemma III.1:4 then gives that Pu and Pu' are orthogonal. □

More generally, if $p(N\,|\,P)$ is nonzero, then P on N is angle-preserving, i.e. the angle between Pu and Pu' equals that between u and u'. This can be seen from the fact that $U = P/\left[p(N\,|\,P)\right]^{1/2}$ defines an isometric mapping of N onto PN. This also proves the following converse to theorem III.1:5

Theorem III.1:6. If P is N-equiangular with nonzero $p(N\,|\,P)$, u and u' are in N and Pu and Pu' are orthogonal, then u and u' are orthogonal.

Theorem III.1:7. Suppose that Pu and Pu' are orthogonal for any orthogonal u, u' in N. Then P is N-equiangular.

Proof. Suppose that Pu' and Pu'' are orthogonal for any orthogonal u', u'' in N. For arbitrary nonzero u', u'' in N we can choose u in N orthogonal to u' with $\|u'\| = \|u\|$ and $u'' = au' + bu$, a and b complex numbers. Then, $\langle u+u', u-u'\rangle = \|u\|^2 - \|u'\|^2 = 0$, thus $u+u'$ and $u-u'$ are orthogonal. Then $P(u+u')$ and $P(u-u')$ are orthogonal and $0 = \langle Pu+Pu', Pu-Pu'\rangle = \|Pu\|^2 - \|Pu'\|^2$ since Pu and Pu' are orthogonal, thus $\|Pu\|^2 = \|Pu'\|^2$ and

$$\begin{aligned}\|Pu''\|^2 = \|aPu' + bPu\|^2 &= |a|^2\|Pu'\|^2 + |b|^2\|Pu\|^2 = \left(|a|^2 + |b|^2\right)\|Pu'\|^2 \\ &= \|u''\|^2\|Pu'\|^2/\|u'\|^2.\end{aligned}$$

Thus $\|Pu''\|^2/\|u''\|^2 = \|Pu'\|^2/\|u'\|^2$ for arbitrary nonzero u', u'' in N and P is N-equiangular. □

Theorem III.1:8. If P_2 is P_1-equiangular and $M_1 = R(P_1)$, then P_2M_1 is closed.

Proof. Set $p = p(P_1\,|\,P_2)$. If $p = 0$, then $P_2M = 0$. If $p > 0$, then $M_1 \subset_\varepsilon R(P_2)$ with $\varepsilon = (1-p)^{1/2} < 1$ and P_2 is closed according to theorem I.1:12 □

Theorem III.1:9. If P is N-equiangular, $N' \subset N$, $N'' \subset N$ and N' and N'' are orthogonal, then PN' and PN'' are orthogonal.

Proof. An immediate consequence of theorem III.1:5. □

III.2 P-equiangular n-tuples of Projections

We shall now consider equiangularity in connection with ordered n-tuples $(P_1, \dots, P_n)$. Let us first consider successive equiangularity in the case of $n = 2$.

Theorem III.2:10. Suppose
1) P_1 is N-equiangular
2) P_2 is $P_1 N$-equiangular
3) P_2 is $P_1^c N$-equiangular
4) $P_2 P_1 N$ and $P_2 P_1^c N$ are orthogonal.

Then P_2 is N-equiangular.

Proof. Suppose 1)–4) are satisfied and that u is a nonzero vector in N. From the identity $P_1 + P_1^c = 1$ follows $P_2 u = P_2 P_1 u + P_2 P_1^c u$ where $P_2 P_1 u$ and $P_2 P_1^c u$ are orthogonal according to 4). Thus $\|P_2 u\|^2 = \|P_2 P_1 u\|^2 + \|P_2 P_1^c u\|^2$. By 2) and 1)

$$\|P_2 P_1 u\|^2 = p(P_1 N \,|\, P_2)\|P_1 u\|^2 = p(P_1 N \,|\, P_2) p(N \,|\, P_1)\|u\|^2$$

Similarly, by 3) and 1)

$$\|P_2 P_1^c u\|^2 = p(P_1^c N \,|\, P_2) p(N \,|\, P_1^c)\|u\|^2.$$

Thus

$$\frac{\|P_2 u\|^2}{\|u\|^2} = p(N \,|\, P_1) p(P_1 N \,|\, P_2) + p(N \,|\, P_1^c) p(P_1^c N \,|\, P_2)$$

is independent of u in N and P_2 is N-equiangular. □

This makes it natural to introduce the following definition

Definition III.2:11. If conditions 1) – 4) of theorem III.2:10 and the condition

5) $P_2^c P_1 N$ and $P_2^c P_1^c N$ are orthogonal,

are satisfied, we shall say that (P_1, P_2) is N-equiangular. If $N = R(P)$ we shall also say that (P_1, P_2) is P-equiangular.

Theorem III.2:12. If (P_1, P_2) is N-equiangular and N is nonzero, then the number $\|P_2 P_1 u\|/\|u\|$ u nonzero vector in N, is independent of u.

Proof. Follows from the proof of theorem III.2:10. □

Definition III.2:13. If (P_1, P_2) is N-equiangular (P-equiangular), N (P) nonzero, then the number $\|P_2 P_1 u\|^2/\|u\|^2$, u nonzero vector in N ($R(P)$), is denoted by $p(N \,|\, P_1, P_2)$ or $p(P \,|\, P_1, P_2)$ respectively. If $N = 0$ ($P = 0$) we define $p(N \,|\, P_1, P_2) = 0$ and $p(P \,|\, P_1, P_2) = 0$.

Theorem III.2:14. If (P_1, P_2) is P-equiangular, then so is (P_1^c, P_2) and

1) $p(P|P_1,P_2) = p(P|P_1)\cdot p(P_1M|P_2)$, $M = R(P)$
2) $p(P|P_2) = p(P|P_1,P_2) + p(P|P_1^c,P_2)$.

Proof. Follows from definition III.2:13 and the proof of theorem III.2:10. □

If (P_1,P_2) is P-equiangular, then by theorem III.1:3, so is (P_1,P_2^c), thus all four combinations $(\underline{P}_1,\underline{P}_2)$, $\underline{P}_1 = P_1$ or P_1^c, $\underline{P}_2 = P_2$ or P_2^c are P-equiangular. This can be generalized to n-tuples of projections.

Definition III.2:15. We shall say that the ordered sequence $(P_1,\dots,P_n)$ is N-equiangular if
1) For every integer $i = 1,\dots,n$ P_i is equiangular with respect to all the 2^{i-1} subspaces $\underline{P}_{i-1}\underline{P}_{i-2}\dots\underline{P}_1N$ (where $\underline{P}_j$ is either P_j or P_j^c). If $i = 1$, this shall mean that P_1 is N-equiangular.
2) All the 2^n subspaces $\underline{P}_n\underline{P}_{n-1}\cdots\underline{P}_1N$ are orthogonal.

If $N = R(P)$ we shall also say that $(P_1,\dots,P_n)$ is P-equiangular.

Theorem III.2:16. If $(P_1,\dots,P_n)$ is N-equiangular (P-equiangular), then so is any $(\underline{P}_1,\dots,\underline{P}_n)$.

Proof. Follows immediately from definition III.2:15 □

Theorem III.2:17. Suppose $(P_1,\dots,P_n)$ is N-equiangular, N nonzero, and let $(\underline{P}_1,\dots,\underline{P}_n)$ be any sequence with $\underline{P}_j = P_j$ or P_j^c, $j = 1,\dots,n$. Then for any integer $i = 1,\dots,n$ the quantity $\|\underline{P}_i\underline{P}_{i-1}\cdots\underline{P}_1u\|/\|u\|$, u nonzero vector in N, is independent of u.

Proof. By condition 1) of definition III.2:15 we have (u nonzero vector in N)

$$\|\underline{P}_i\underline{P}_{i-1}\cdots\underline{P}_1u\|^2 = p(\underline{P}_{i-1}\cdots\underline{P}_1N|P_i)\dots\|\underline{P}_{i-1}\cdots\underline{P}_1u\|^2.$$

Repeating successively this argument on the last factor we get

$$\|\underline{P}_i\underline{P}_{i-1}\cdots\underline{P}_1u\|^2 = p(\underline{P}_{i-1}\cdots\underline{P}_1N|P_i)\cdots p(N|P_1)\cdot\|u\|^2$$

and the result follows. □

Definition III.2:18. If $(P_1,\dots,P_n)$ is N-equiangular and u nonzero vector i N, then the number $\|P_iP_{i-1}\cdots P_1u\|^2/\|u\|^2$ is denoted by $p(N|P_1,\dots,P_n)$.
If $N = 0$, we define $p(N|P_1,\dots,p_n) = 0$.
If $N = R(P)$ we shall also denote this number by $p(P|P_1,\dots,P_n)$.

Theorem III.2:19. If $(P_1,\dots,P_n)$ is N-equiangular, then

$$p(N|\underline{P}_1,\dots,\underline{P}_i) = p(N|\underline{P}_1,\dots,\underline{P}_{i-1})p(\underline{P}_{i-1}\cdots\underline{P}_1N|\underline{P}_i)$$

for any integer $i = 1,2,\dots,n$.

Proof. Follows immediately from the proof of theorem III.2:17 and definition III.2:18. □

Theorem III.2:20. Suppose $(P_1, \dots, P_n)$ is N-equiangular and N nonzero. Then

$$\sum p(N|\underline{P}_1, \dots, \underline{P}_n) = 1,$$

where the sum is over all the 2^n combinations $\underline{P}_j = P_j$ or P_j^c, $j = 1, \dots, n$.

Proof. Repeated use of the identities $1 = P_j + P_j^c$, $j = 1, \dots, n$ gives

$$u = \sum \underline{P}_n \underline{P}_{n-1} \cdots \underline{P}_1 u$$

where the sum is over all combinations $\underline{P}_j$. If $(P_1, \dots, P_n)$ is N-equiangular and u is in N, then all 2^n terms in this sum are orthogonal according to condition 2) of definition III.2:15. Since N is nonzero we can assume that $\|u\| = 1$. Then

$$1 = \sum \|\underline{P}_n \underline{P}_{n-1} \cdots \underline{P}_1 u\|^2 = \sum p(N|\underline{P}_1, \dots, \underline{P}_n)$$

by definition III.2:18. □

Theorem III.2:21. Suppose $(P_1, \dots, P_n)$ is N-equiangular and let $(i_1, \dots, i_p)$ be an ordered subsequence of $(1, \dots, n)$ and $(P'_1, \dots, P'_p)$ the corresponding subsequence of $(P_1, \dots, P_n)$. Then $(P'_1, \dots, P'_p)$ is N-equiangular and

$$p(N|P'_1, \dots, P'_p) = \sum p(N|\underline{P}_1, \dots, \underline{P}_n)$$

where the sum is over all $(\underline{P}_1, \dots, \underline{P}_n)$ which contain $(P'_1, \dots, P'_p)$ as subsequence. i.e. such that $\underline{P}_k = P_k = P'_j$ if $k = i_j$, $j = 1, \dots, p$.

Proof. The theorem is trivially true if $N = 0$. Suppose N is nonzero. Using the identity $P_k^c + P_k = 1$ for all $k = 1, \dots, n$ not equal to any of i_j, we get the identity

$$P'_p \underline{P}'_{p-1} \cdots \underline{P}'_1 u = \sum{}' \underline{P}_n \underline{P}_{n-1} \cdots \underline{P}_1 u$$

where the sum is over all $\underline{P}_1, \dots, \underline{P}_n$ such that $\underline{P}_k = P_k = P'_p$ if $k = i_p$ and $\underline{P}_k = \underline{P}'_q$ if $k = i_q$, $q = 1, \dots, p-1$. If $(P_1, \dots, P_n)$ is N-equiangular and u is in N, the terms in the sum are orthogonal by condition 2) of definition III.2:15. Then, if $\|u\| = 1$,

$$\|P'_p \underline{P}'_{p-1} \cdots \underline{P}'_1 u\|^2 = \sum{}' \|\underline{P}_n \underline{P}_{n-1} \cdots \underline{P}_1 u\|^2 = \sum{}' p(N|\underline{P}_1, \cdots, \underline{P}_n) \qquad (1)$$

so $\|P'_p \underline{P}'_{p-1} \cdots \underline{P}'_1 u\|$ is independent of u. By a similar argument, $\|\underline{P}'_{p-1} \cdots \underline{P}'_1 u\|$ is independent of u. Thus, either $\underline{P}'_{p-1} \cdots \underline{P}'_1 u = 0$ in which case $\underline{P}'_{p-1} \cdots \underline{P}'_1 N = 0$, or $\|P'_p \underline{P}'_{p-1} \cdots \underline{P}'_1 u\| / \|\underline{P}'_{p-1} \cdots \underline{P}'_1 u\|$ is independent of u. Since every vector in $\underline{P}'_{p-1} \cdots \underline{P}'_1 N$ is of the form $\underline{P}'_{p-1} \cdots \underline{P}'_1 u$ with u in N, this proves that P'_p is equiangular with respect to $\underline{P}'_{p-1} \cdots \underline{P}'_1 N$. Repeating the argument for every $(\underline{P}'_1, \cdots, \underline{P}'_q)$, $q = p-2, \dots, 1$, we find that condition 1) of definition III.2:15 for the N-equiangularity of $(P'_1, \dots, P'_p)$ is satisfied.

From the identity

$$\underline{P}'_p \cdots \underline{P}'_1 u = \sum_S \underline{P}_n \underline{P}_{n-1} \cdots \underline{P}_1 u$$

where the sum is over the set S of all $(\underline{P}_1, \ldots, \underline{P}_n)$ with $\underline{P}_k = P_k = P'_j$ if $k = i_j$, $j = 1, \ldots, p$ i.e. those $\underline{P}_1, \ldots, \underline{P}_n$ which contain $P'_1, \ldots, P'_p$ as subsequence, we get

$$\underline{P}'_p \cdots \underline{P}'_1 N \subset \bigoplus_S \underline{P}_n \cdots \underline{P}_1 N$$

where the direct sum is over the same set S. For different choices of $(\underline{P}'_1, \ldots, \underline{P}'_p)$, the corresponding $\underline{P}'_p \cdots \underline{P}'_1 N$ are included in direct sums over disjoint sets S. The orthogonality of all $\underline{P}'_p \cdots \underline{P}'_1 N$ then follows from the orthogonality of all $\underline{P}_n \cdots \underline{P}_1 N$ according to condition 2) of definition III.2:15 for the N-equiangularity of $(P_1, \ldots, P_n)$. This proves condition 2) of definition III.2:15 for the N-equiangularity of $(\underline{P}'_1, \ldots, \underline{P}'_p)$.
Having proved the N-equiangularity of $(\underline{P}'_1, \ldots, \underline{P}'_p)$, the formula (1) above, with all $\underline{P}'_j = P_j$, gives

$$p(N \mid P'_1, \ldots, P'_p) = \sum_S p(N \mid \underline{P}_1, \ldots, \underline{P}_n)$$

and the last part of the theorem is proved. □

Theorem III.2:22. If $(P_1, \ldots, P_n)$ is N-equiangular and $N' \subset N$, then $(P_1, \ldots, P_n)$ is also N'-equiangular and if N' is nonzero, then

$$p(N' \mid P_1, \ldots, P_n) = p(N \mid P_1, \ldots, P_n).$$

Proof. Suppose $N' \subset N$. By definition III.1:1 it is obvious that for any P we have that if P is N-equiangular, then P is also N'-equiangular. Now suppose $(P_1, \ldots, P_n)$ is N-equiangular. Then conditions 1) and 2) of definition III.2:15 for N'-equiangularity follow from the corresponding conditions for N-equiangularity by noting that

$$\underline{P}_j \underline{P}_{j-1} \cdots \underline{P}_1 N' \subset \underline{P}_j \underline{P}_{j-1} \cdots \underline{P}_1 N$$

for $j = 1, \ldots, n$.
The last part of the theorem follows directly from definition III.2:18. □

Theorem III.2:23. Suppose $(P_1, \ldots, P_m, P'_1, \ldots, P'_n)$ is N-equiangular. Then $(P'_1, \ldots, P'_n)$ is equiangular with respect to $N_1 = P_m P_{m-1} \cdots P_1 N$ and

$$p(N \mid P_1, \ldots, P_m, P'_1, \ldots, P'_n) = p(N \mid P_1, \ldots, P_m) \cdot P(N_1 \mid P'_1, \ldots, P'_n).$$

Proof. Suppose $(P_1, \ldots, P_m, P'_1, \ldots, P'_n)$ is N-equiangular and N nonzero, the case $N = 0$ being trivial. Let u be a vector in N with $\|u\| = 1$. Then, by theorem III.2:17 $\|P'_k \underline{P}'_{k-1} \cdots \underline{P}'_1 P_m \cdots P_1 u\|$ is independent of u for any $k = 1, \ldots, n$. Similarly, $p = \|\underline{P}'_{k-1} \ldots \underline{P}'_1 P_m \cdots P_1 u\|$ is independent of u. If $p = 0$, then $N' =$

$\underline{P}'_{k-1} \dots \underline{P}'_1 P_m \cdots P_1 N = 0$ and P'_k is N'-equiangular. Suppose N' nonzero. Since every nonzero vector u' in N' is of the form $u' = \underline{P}'_{k-1} \cdots \underline{P}'_1 P_m \cdots P_1 u''$ with u'' in N, we have that

$$\frac{\|P'_k u'\|}{\|u'\|} = \frac{\|P'_k \underline{P}'_{k-1} \cdots \underline{P}'_1 P_m \cdots P_1 u''\|}{\|\underline{P}'_{k-1} \cdots \underline{P}'_1 P_m \cdots P_1 u''\|}$$

is independent of u' and thus P'_k is N'-equiangular. Thus $(P'_1, \dots, P'_n)$ satisfies condition 1) of definition III.2:15 for equiangularity with respect to N_1.

Since the set of all $\underline{P}'_n \cdots \underline{P}'_1 N_1$ is a subset of the set of all $\underline{P}'_n \cdots \underline{P}'_1 \underline{P}_m \cdots \underline{P}_1 N$, the condition 2) of definition III.2:15 for $(P'_1, \dots, P'_n)$ to be N_1-equiangular follows from that of $(P_1, \dots, P_m, P'_1, \dots, P'_n)$ to be N-equiangular.

Having thus proved that $(P'_1, \dots, P'_n)$ is N_1-equiangular, we have

$$\begin{aligned} &p(N|P_1, \dots, P_m, P'_1 \dots, P'_n) = \|P'_n \cdots P'_1 P_m \cdots P_1 u\|^2 \\ &= p(N_1|P'_1, \dots, P'_n) \cdot \|P_m \cdots P_1 u\|^2 = p(N_1|P'_1, \dots, P'_n) \cdot p(N|P_1, \dots, P_m) \end{aligned}$$

and the last part of the theorem is proved. □

The following theorem is a partial converse of theorem III.2:23 and generalizes theorem III.2:10

Theorem III.2:24. Suppose
1) $(P_1, \dots, P_m)$ is N-equiangular
2) $(P'_1, \dots, P'_n)$ is N'-equiangular for every $N' = \underline{P}_m \underline{P}_{m-1} \cdots \underline{P}_1 N$
3) all the 2^{m+n} subspaces $\underline{P}'_n \underline{P}'_{n-1} \cdots \underline{P}'_1 \underline{P}_m \underline{P}_{m-1} \cdots \underline{P}_1 N$ are orthogonal.
Then $(P_1, \dots, P_m, P'_1, \dots, P'_n)$ is N-equiangular.

Proof. Condition 3) of the theorem is identical with condition 2) of definition III.2:15 for N-equiangularity of $(P_1, \dots, P_m, P'_1, \dots, P'_n)$. From condition 1) of the theorem follows that P_i is equiangular with respect to all $\underline{P}_{i-1} \dots \underline{P}_1 N$, $i = 1, \dots, m$. Thus it suffices to prove condition 1) of definition III.2:15 for $i = m+1, \dots, m+n$, i.e. that P'_i is equiangular with respect to $\underline{P}'_{i-1} \cdots \underline{P}'_1 \underline{P}'_m \cdots \underline{P}_1 N$. By condition 1) of the theorem and theorem III.2:21, P'_1 is equiangular with respect to $\underline{P}_m \cdots \underline{P}_1 N$ so condition 1) of definition III.2:15 is satisfied for $i = m+1$. From condition 2) of the theorem and theorems III.2:16 and III.2:23 follows that $(P'_2, \dots, P'_n)$ is equiangular with respect to $\underline{P}'_1 \underline{P}_m \dots \underline{P}_1 N$. Thus we can repeat the argument for $i = m+2, \dots, m+n$ and the theorem is proved. □

Theorem III.2:25. Suppose $(P_1, \dots, P_n)$ is N-equiangular. Then $p(N|P_1, \dots, P_n) = 1$ iff $0 \neq N \subset N_1 \subset N_2 \subset \dots \subset N_n$ where $N_j = P_j P_{j-1} \dots P_1 N$, $j = 1, \dots, n$.

Proof. Suppose $p(N|P_1, \dots, P_n) = 1$. It follows from theorem III.2:23 that $1 = p(N|P_1, \dots, P_n) = p(N|P_1) \cdot p(N_1|P_2, \dots, P_n)$. Since both factors lie in the interval $[0, 1]$ it follows that $p(N|P_1) = 1$ and $p(N_1|P_2, \dots, P_n) = 1$ which implies that $0 \neq N \subset N_1$. Repeating the argument on $p(N_j|P_{j+1}, \dots, P_n)$ for $j = 2, \dots, n-1$ we find that $N_1 \subset N_2 \subset \dots \subset N_n$.
The converse of the theorem is obvious from definition III.2:18. □

Theorem III.2:26. Suppose $(P_1, \dots, P_n)$ is N-equiangular and $P_i = P_j^c$ for some i, j, $1 \le i, j \le n$. Then $p(N \mid P_1, \dots, P_n) = 0$.

Proof. Suppose $P_i = P_j^c$ and $i < j$. Then, by theorem III.2:21 (P_i, P_i^c) is N-equiangular and

$$p(N \mid P_i, P_i^c) = p(N \mid P_i, P_j) \ge p(N \mid P_1, \dots, P_n)$$

since $p(N \mid P_i, P_j)$ is a sum of nonnegative terms, one of which is $p(N \mid P_1, \dots, P_n)$. But, by definition III.2:18 $p(N \mid P_i, P_i^c) = 0$ and the theorem follows. □

Theorem III.2:27. Suppose $(P_1, \dots, P_k, P_{k+1}, \dots, P_n)$ is N-equiangular and P_k and P_{k+1} commute. Then $(P_1, \dots, P_{k+1}, P_k, \dots, P_n)$ is N-equiangular and

$$p(N \mid P_1, \dots, P_k, P_{k+1}, \dots, P_n) = p(N \mid P_1, \dots, P_{k+1}, P_k, \dots, P_n)$$

Proof. Under the assumptions of the theorem we have to prove conditions 1) and 2) definition III.2:15 for the N-equiangularity of $(P_1, \dots, P_{k+1}, P_k, \dots, P_n)$. Condition 2) is obviously satisfied. Condition 1) is also obvious if $i \ne k+1$ and it suffices to prove that P_k is equiangular with respect to $N_1 = P_{k+1}P_{k-1} \dots, P_1 N$. This is obvious if $N_1 = 0$ so we can assume that $N_1 \ne 0$. Let u be an arbitrary vector in N with $\|u\| = 1$. Then $\|P_k P_{k+1} P_{k-1} \dots P_1 u\| = \|P_{k+1} P_k P_{k-1} \dots P_1 u\|$ is independent of u. Also $\|P_{k+1} P_{k-1} \dots P_1 u\|$ is independent of u since $(P_1, \dots, P_{k-1}, P_{k+1})$ is N-equiangular by theorem III.2:21. Thus $\|P_k P_{k+1} P_{k-1} \cdots P_1 u\| / \|P_{k+1} P_{k-1} \cdots P_1 u\|$ is independent of u and since every vector u_1 in N_1 is of the form $u_1 = P_{k+1} P_{k-1} \cdots P_1 u$ with u in N, this shows that P_k is N_1-equiangular.

Having thus proved that $(P_1, \dots, P_{k+1}, P_k \dots P_n)$ is N-equiangular, the second part of the theorem is obvious from definition III.2:18. □

III.3 Equiangular (m, n)-tuples of Projections

Definition III.3:28. We shall say that $(P_1', \dots, P_n')$ is equiangular with respect to $(P_1, \dots, P_m)$ (or $(P_1, \dots, P_m)$-equiangular) if $(P_1', \dots, P_n')$ is equiangular with respect to $N = P_m P_{m-1} \cdots P_2 R(P_1)$. We shall then also say that the (m, n)-tuple $(P_1, \dots, P_m \mid P_1', \dots, P_n')$ is equiangular and define

$$p(P_1, \dots, P_m \mid P_1', \dots, P_n') = p(N \mid P_1', \dots, P_n').$$

Theorem III.3:29. If the (m, n)-tuple $(P_1, \dots, P_m \mid P_1', \dots, P_n', P_1'', \dots, P_p'')$ is equiangular, then $(P_1, \dots, P_m, P_1', \dots, P_n' \mid P_1'', \dots, P_p'')$ is equiangular and

$$\begin{aligned} &p(P_1, \dots, P_m \mid P_1', \dots, P_n', P_1'', \dots, P_p'') \\ &= p(P_1, \dots, P_m \mid P_1', \dots, P_n') \cdot p(P_1, \dots, P_m, P_1', \dots, P_n' \mid P_1'', \dots, P_p''). \end{aligned}$$

Proof. A reformulation of theorem III.2:23 by means of definition III.3:28. □

Theorem III.3:30. Suppose $(P_1, \dots, P_m \mid P_1', \dots, P_n')$ is equiangular. Then $(P_1'', \dots, P_k'', P_1, \dots, P_m \mid P_1', \dots, P_n')$ is equiangular and, if $P_m \cdots P_1 \cdot P_k'' \cdots P_2'' R(P_1'')$

is nonzero, then

$$p(P_1'', \ldots, P_k'', P_1, \ldots, P_m | P_1', \ldots, P_n') = p(P_1, \ldots, P_m | P_1', \ldots, P_n').$$

Proof. Follows from theorem III.2:22 observing that $P_m \cdots P_1 \cdot P_k'' \cdots P_2'' R(P_1'') \subset P_m \cdots P_2 R(P_1)$. □

Theorem III.3:31. Suppose $(P_1, \ldots, P_k, P_{k+1}, \ldots, P_m | P_1', \ldots, P_n')$ is equiangular and that P_k and P_{k+1} commute. Then $(P_1, \ldots, P_{k+1}, P_k, \ldots, P_m | P_1', \ldots, P_n')$ is equiangular and

$$p(P_1, \ldots, P_k, P_{k+1}, \ldots, P_m | P_1', \ldots, P_n') = p(P_1, \ldots, P_{k+1}, P_k, \ldots, P_m | P_1', \ldots, P_n').$$

Proof. Follows directly from definition III.3:28 noting that $P_m \cdots P_{k+1} P_k \cdots P_2 R(P_1) = P_m \cdots P_k P_{k+1} \cdots P_2 R(P_1)$. □

III.4 Stochastic Event Structures of Projections

Theorem III.4:32. Suppose E is a set of projections, closed under the operation of orthogonal complement (i.e. P in E implies P^c in E). Let T be a set (m,n)-tuples $(P_1, \ldots, P_m | P_1', \ldots, P_n')$, $m, n = 1, 2, \ldots$ such that

a) all $P_1, \ldots, P_m, P_1', \ldots P_n'$ are in E

b) $(P_1, \ldots, P_m | P_1', \ldots, P_n')$ is equiangular

c) $P_m P_{m-1} \cdots P_2 R(P_1)$ is nonzero.

Suppose also that the set T has the properties

d) if $(P_1, \ldots, P_m | P_1', \ldots, P_n')$ is in T, so is $(P_1, \ldots, P_m | \underline{P}_1'', \ldots, \underline{P}_r'')$, where $P_1'' \ldots, P_r''$ is any subsequence of $P_1', \ldots, P_n'$.

e) if $P_1, \ldots, P_m | P_1', \ldots, P_n', P_1'', \ldots, P_k''$ is in T and $p(P_1, \ldots, P_m | P_1', \ldots, P_n')$ is nonzero,
then $(P_1, \ldots, P_m, P_1', \ldots, P_n' | P_1'', \ldots P_k'')$ is in T.

Let $p : T \to [0,1]$ be defined as in definition III.3:28 and let $-P$ stand for the operation of orthogonal complement $-P = P^c$. Then $(E, -, T, p)$ is a stochastic event structure.

Proof. We have to prove conditions 1)–3) of definition II.1:1. 1) is obviously satisfied, E being closed under the operation $-$ meaning orthogonal complement. Condition 2) of definition II.1:1 follows from assumptions a) and d). (Condition d) of the the theorem is consistent with b) according to theorem III.2:21.) Condition 3a) of definition II.1:1 follows from assumptions b) and c) and theorem III.2:20. Condition 3b) of definition II.1:1 follows from assumptions b) and theorem III.2:21. Condition 3c) of definition II.1:1 follows from the assumption b) and theorem III.2:26. The first part of the condition d) of definition II.1:1 is identical with assumption e) and the second part of condition d) follows from theorem III.3:29. Finally, condition 3e) of definition II.1:1 follows from theorem III.3:30

and assumption c). (Condition e) of the theorem is consistent with b) according to theorem III.3:29.) □

Theorem III.4:33. Let E be a set of projections, closed under the operation of orthogonal complement. Let $|$ be a relation on E such that

1) if $P_1|P_2$ then $\underline{P}_1|\underline{P}_2$

2) if $P_1|P_2$ and $P_2|P_1$ then P_1 and P_2 commute.

Let T be the set of all (m,n)-tuples of projections in E satisfying conditions a) b) and c) of theorem III.4:32 and the condition

3) If $(P_1,\dots,P_m|P'_1,\dots,P'_n)$ is in T then

   a) $P_i|P_j$ if $i<j$, $i,j=1,\dots,m$

   b) $P'_i|P'_j$ if $i<j$, $i,j=1,\dots,n$

   c) $P_i|P'_j$ for all $i=1,\dots,n$, $j=1,\dots,n$.

Define $-$ and p as in theorem III.4:32. Then $(E,-,T,p,|)$ is a linearly time-ordered stochastic event structure.

Proof. We first use theorem III.4:32 to prove that $(E,-,T,p)$ is a stochastic event structure. We have to prove that conditions d) and e) of theorem III.4:32 are satisfied, the other conditions being obvious.

Suppose $(P_1,\dots,P_m|P'_1,\dots,P'_n)$ is in T and that $P''_1,\dots,P''_r$ is a subsequence of $P'_1,\dots,P'_n$. Then, by theorem III.2:21 $(P_1,\dots,P_m|\underline{P}''_1,\dots,\underline{P}''_r)$ is equiangular and obviously satisfies the statements corresponding to assumptions a) b) and c) of theorem III.4:32 and conditions 3a)–c) of the present theorem. Since T is the set of all tuples with these peoperties, it follows that $(P_1,\dots,P_m|\underline{P}''_1,\dots,\underline{P}''_r)$ is in T and condition d) of theorem III.4:32 is proved.

Similarly, to prove condition e), suppose that $(P_1,\dots,P_m|P'_1,\dots,P'_n,P''_1,\dots,P''_k)$ is in T and $p(P_1,\dots,P_m|P'_1,\dots,P'_n)$ nonzero. Then, by definitions III.3:28 and III.2:18, $P'_n\cdots P'_1P_m\cdots P_2R(P_1)$ is nonzero and, by theorem III.3:29 it follows that $P_1,\dots,P_m,P'_1,\dots,P'_n|P''_1,\dots,P''_k$ is equiangular. Then $(P_1,\dots,P_m,P'_1,\dots,P'_n|P''_1,\dots,P''_k)$ obviously satisfies conditions a) b) and c) of theorem III.4:32 and conditions 3a)–c) of the present theorem and thus belongs to T. Thus condition e) of theorem III.4:32 is satisfied. Thus $(E,-,T,p)$ is a stochastic event structure.

We now prove the conditions of theorem II.3:28 for $(E,-,T,p,|)$ to be a linearly time-ordered stochastic event structure. Conditions 1) and 2) of definition II.3:24 are identical with assumptions 1) and 3) respectively. To prove condition 1) of theorem II.3:28, we note that if $(P_1,\dots,P_m|P'_1,\dots,P'_k,P'_{k+1},\dots,P'_n)$ is in T and $P'_{k+1}|P'_k$, then P'_k and P'_{k+1} commute according to assumption 2). Condition 1) of theorem II.3:28 then follows from theorem III.2:27. Similarly, condition 2) of theorem II.3:28 follows from theorem III.3:31 □

III.5 Equiangularity and Factorization. Representation of a Minimal Simple Stochastic Event Structure

If P_1 is M-equiangular and $p = p(M|P_1) \neq 0$, then, by definition III.1:1, $U_1 = \frac{1}{\sqrt{p}}P_1$ restricted to M gives an isometric mapping of M onto P_1M and, if also $p \neq 1$, then $U_2 = \frac{1}{\sqrt{1-p}}P_1^c$ restricted to M gives an isometric mapping of M onto P_1^cM. $U_{12} = U_2U_1^{-1}$ is then an isometric mapping of P_1M onto P_1^cM.

Conversely, we have the following construction of a pair M, P such that P is M-equiangular.

Theorem III.5:34. Suppose the subspaces $M_1 = R(P_1)$ and $M_2 = R(P_2)$ are orthogonal and let U_{12} be an isometric mapping of M_1 onto M_2. Set $U = \sqrt{p} + \sqrt{1-p}U_{12}$ and $M = UM_1$. Then any P with $M_1 \subset R(P) \subset M_2^c$ is M-equiangular and $p(M|P) = p$.

Proof. Since M_1 and $M_2 = U_{12}M_1$ are orthogonal, U is an isometric map of M_1 onto M. Then any u in M is of the form $u = \sqrt{p}u_1 + \sqrt{1-p}u_2$ where u_1 is in $M_1 \subset R(P)$ and $u_2 = U_{12}u_1$ is in $M_2 \subset R(P^c)$. Thus $Pu = \sqrt{p}u_1$ where $\|u\| = \|u_1\|$ and the result follows. □

Theorem III.5:35. P' is M-equiangular with $p(M|P') = p$ iff there exist mappings $U_1 : M \to P'M$ and $U_2 : M \to P'^cM$ onto $P'M$ and P'^cM respectively, which are either isometric or zero, such that

$$\sqrt{p}U_1 + \sqrt{1-p}U_2 = 1.$$

In this case $U_1 = \frac{1}{\sqrt{p}}P'$ on M if $p > 0$ and $U_2 = \frac{1}{\sqrt{1-p}}P'^c$ on M if $p < 1$.

Proof. A simple reformulation of the proof of theorem III.5:34 and the preceding discussion. □

Theorem III.5:36. Suppose P' and P'' are projections in a Hilbert space $\mathscr{H}$ and that the corresponding subspaces $M' = R(P')$ and $M'' = R(P'')$ are infinite dimensional. P'' is M'-equiangular and $M'' - P''M'$ and $M''^c - P''^cM'$ are each infinite or zero dimensional iff there exists a tensor product decomposition $T :\otimes \mathscr{H}_1 \times \mathscr{H}_2 \to \mathscr{H}$ with $\dim \mathscr{H}_1 \leq 4$, a vector u_M in $\mathscr{H}_1$ and a subspace M_1'' in $\mathscr{H}_1$ with $\dim M_1'' \leq 2$ such that $M' = u_M \otimes_T \mathscr{H}_2$ and $M'' = M_1'' \otimes_T \mathscr{H}_2$.

Proof. If such a representation of M' and M'' by u_M and M_1'' exists, the M'-equiangularity of P'' and the condition that $M'' - P''M'$ and $M''^c - P''^cM'$ are infinite or zero dimensional are obvious.

Suppose P'' is M'-equiangular and that $M'' - P''M'$ and $M''^c - P''^cM'$ are infinite or zero dimensional. By theorem III.5:35 there exist mappings $U_1 : M' \to M_1 = P''M'$ and $U_2 : M' \to M_2 = P''^cM'$ onto M_1 and M_2 which are either

isometric or zero. Since $M'' - P''M'$ and $M''^c - P''^c M'$ are either infinite or zero dimensional, there exist mappings $U_3 : M' \to M_3 = M'' - P''M'$ and $U_4 : M' \to M_4 = M''^c - P''^c M'$ onto M_3 and M_4 which are either isometric or zero. Let I be the subset of $\{1, 2, 3, 4\}$ such that U_i is nonzero for i in I. Then lemma I.1:34 gives with $\mathscr{H}_1 = l^2(I)$ and $\mathscr{H}_2 = M'$ a factorization $T :_{\otimes} \mathscr{H}_1 \times \mathscr{H}_2 \to \mathscr{H}$. Suppose all $U_1, \dots, U_4$ nonzero i.e. $I = \{1, 2, 3, 4\}$. Since $M'' = M_1 \oplus M_3$ we have $M'' = M_1'' \otimes_T \mathscr{H}_2$ where M_1'' is the subspace of $\mathscr{H}_1 = l^2(I) = \mathbb{C}^4$ spanned by the vectors $(1,0,0,0)$ and $(0,0,1,0)$. Set $p = p(M'|P'')$ and $u_M = (\sqrt{p}, \sqrt{1-p}, 0, 0)$. Then, by the construction of T according to lemma I.1:34 we have for arbitrary u in $M' = \mathscr{H}_2$

$$u = P''u + P''^c u = \sqrt{p}U_1 u + \sqrt{1-p}U_2 u = u_M \otimes_T u,$$

thus $M' = u_M \otimes_T M' = u_M \otimes_T \mathscr{H}_2$.

If some of the $U_1, \dots, U_4$ are zero, we just have to drop the corresponding dimensions in $l^2(\{1, 2, 3, 4\})$ and get a Hilbert space $\mathscr{H}_1 = l^2(I)$ with dimension 1, 2 or 3 (both U_1 and U_2 cannot be zero). □

The following straightforward generalization of theorem III.5:36 is useful for a construction of an M-equiangular sequence $(P_1, \dots, P_n)$.

Theorem III.5:37. The following conditions 1) and 2) are equivalent

1) a) M_j, $j = 1, \dots, J$ are infinite dimensional and orthogonal
   b) P is M_j-equiangular, $j = i, \dots, J$
   c) all $2J$ subspaces $\underline{P}M_j$ are orthogonal,
   d) $R(P) - \left(\oplus_j PM_j \right)$ and $R(P^c) - \left(\oplus_j P^c M_j \right)$ are each infinite or zero-dimensional

2) There exists a tensor product decomposition $T_{\otimes} : \mathscr{H}_1 \times \mathscr{H}_2 \to \mathscr{H}$ with $\dim \mathscr{H}_1 \leq 2J+2$, $\dim \mathscr{H}_2 = \infty$, orthogonal vectors u_j, $j = 1, \dots, J$ in $\mathscr{H}_1$ and a subspace M in $\mathscr{H}_1$ with $\dim M \leq J+1$ such that

$$M_j = u_j \otimes_T \mathscr{H}_2,\ j = 1, \dots, J$$
$$R(P) = M \otimes_T \mathscr{H}_2$$

   and all $2J$ vectors $\underline{P}_1 u_j$ are orthogonal, where P_1 is the projection (in $\mathscr{H}_1$) on M

Proof. It is easily checked that 2) implies 1). Suppose 1) is satisfied. Let $\mathscr{H}_2$ be any infinite-dimensional separable Hilbert space. Since all M_j are infinite-dimensional (separable) Hilbert spaces there exist isometric onto mappings

$$U_j : \mathscr{H}_2 \to M_j.$$

By theorem III.5:35, there exist isometric or zero, onto mappings

$$U_{j,1} : M_j \to M'_{j,1} = PM_j,$$
$$U_{j,2} : M_j \to M'_{j,2} = P^c M_j,$$

and, by assumption d), there exist isometric or zero mappings

$$U_3 : \mathscr{H}_2 \to M_3' = R(P) - \left(\oplus_j P M_j \right)$$
$$U_4 : \mathscr{H}_2 \to M_4' = R(P^c) - \left(\oplus_j P^c M_j \right).$$

Setting $\mathscr{H}_1 = \mathbb{C}^{2J+2}$, $p_j = p(M_j|P)$, $u_j = (0, 0, \ldots, \sqrt{p_j}, \sqrt{1-p_j}, \ldots, 0, 0)$ and M being spanned by the $J+1$ vectors

$$(1, 0, \ldots, 0)$$
$$(0, 0, 1, 0, \ldots, 0)$$
$$\ldots$$
$$(0, 0, \ldots, 1, 0),$$

the statements of condition 2) follow as in the proof of theorem III.5:36. (If some of the p_j or $1-p_j$ equal zero, or if M_3 or M_4 equals zero, we drop the corresponding dimensions in $\mathscr{H}_1$.) □

Theorem III.5:38. Let S be a minimal simple stochastic event structure spanned by the process $(e_1, \ldots, e_m|e_1', \ldots, e_n')$. Then there exists a minimal simple stochastic event structure S' of projections, spanned by an equiangular sequence

$$(P_1, \ldots, P_m|P_1', \ldots, P_n')$$

with the same p-values (e.g. $p(P_1, \ldots, P_m|\underline{P}_1', \ldots, \underline{P}_n') = p(e_1, \ldots, e_m|\underline{e}_1', \ldots, \underline{e}_n')$) i.e. S and S' are isomorphic.

Proof. By definition III.3:28, it obviously suffices to show the existence of an M-equiangular sequence $(P_1, \ldots, P_n)$ with given p-values. We prove this by induction and assume that there exist M, $P_1, \ldots, P_{n-1}$ in $\mathscr{H}$ such that M is infinite-dimensional, $(P_1, \ldots, P_{n-1})$ is M-equiangular and $p(M|\underline{P}_1, \ldots, \underline{P}_{n-1}) = p(e_1, \ldots, e_m|\underline{e}_1', \ldots, \underline{e}_{n-1}')$ (this being trivial if $n = 1$). Let M_j, $j = 1, \ldots, 2^{n-1}$ be the 2^{n-1} orthogonal subspaces $\underline{P}_{n-1} \cdots \underline{P}_1 M$ (which are zero or infinite-dimensional) and p_j the corresponding values of

$$p(e_1, \ldots, e_m|\underline{e}_1', \ldots, \underline{e}_{n-1}', e_n')/p(e_1, \ldots, e_m|\underline{e}_1', \ldots, \underline{e}_{n-1}')$$

where we define $p_j = 0$ if $p(e_1, \ldots, e_m|\underline{e}_1', \ldots, \underline{e}_{n-1}') = 0$. (if $n = 1$ we take $p_1 = p(e_1, \ldots, e_m|e_1')$). By the proof of theorem III.5:37 there exist P' and orthogonal M_j' with $\dim M_j' = \dim M_j$, $j = 1, \ldots, 2^{n-1}$, and $\dim\left(H - (\oplus_j M_j')\right) = \dim\left(H - (\oplus_j M_j)\right)$, such that P' is M_j'-equiangular with $p(M_j'|P') = p_j$ and with all 2^n $\underline{P}' M_j'$ orthogonal. There exists a unitary transformation in H such that $M_j = U M_j'$, $j = 1, \ldots, 2^{n-1}$. Set $P_n = U P' U^{-1}$. Then P_n is M_j-equiangular with $p(M_j|P_n) = p_j$ and all $\underline{P}_n M_j$ are orthogonal. Then, by theorem III.2:24, $(P_1, \ldots, P_n)$ is M-equiangular and by the definition of p_j and theorem III.2:23, we have

$$p(M|\underline{P}_1, \ldots, \underline{P}_n) = p(e_1, \ldots, e_m|\underline{e}_1', \ldots, \underline{e}_n').$$

The equality for all p-values in S and S' follows from theorem II.2:17 □

III.6 Approximately Equiangular (p_0, ε)-equiangular Sequences

Definition III.6:39. P_2 is called ε-equiangular with respect to N_1 if

$$\left| \|P_2 u'\|^2 - \|P_2 u''\|^2 \right| \leq \varepsilon$$

for any u', u'' in N_1 with $\|u'\| = \|u''\| = 1$. If P_2 is ε-equiangular with respect to N_1, we denote by $p(N_1|P_2)$ any of the values $\|P_2 u\|^2$, u in N_1, $\|u\| = 1$. If $N_1 = 0$ we define any P_2 to be ε-equiangular with respect to N_1 and we define $p(N_1|P_2) = 0$. If $N_1 = R(P_1)$ we also say that P_2 is ε-equiangular with respect to P_1 and that $(P_1|P_2)$ is ε-equiangular and write $p(P_1|P_2)$ for $p(N_1|P_2)$.

Note that $p(N_1|P_2) = p(P_1|P_2)$ is here an approximately defined quantity. It is only defined to the accuracy ε. Any of these values is called a representative of the approximate quantity. An equality between two approximate quantities, as in the following theorem, is then to be understood to hold for at least some representatives. For $\varepsilon = 0$, the concepts of ε-equiangularity and equiangularity (0-equiangularity) (definitions III.1:1 and III.1:2) coincide.

Theorem III.6:40. If P_2 is ε-equiangular with respect to N_1, then so is P_2^c and $p(N_1|P_2^c) = 1 - p(N_1|P_2)$.

Proof. Follows directly from the identities $\|P_2^c u'\|^2 = 1 - \|P_2 u'\|^2$ and $\|P_2^c u'\|^2 - \|P_2^c u''\|^2 = \|P_2 u''\|^2 - \|P_2 u'\|^2$, u', u''' in N_1, $\|u'\| = \|u''\| = 1$. □

Theorem III.6:41. If $M_2 = R(P_2)$ and $N_1 \subset_\varepsilon M_2$, then P_2 is ε^2-equiangular with respect to N_1 and $p(N_1|P_2) \geq 1 - \varepsilon^2$.

Proof. Follows directly from theorem I.1:15 □

Conversely we have:

Theorem III.6:42. If P_2 is ε-equiangular with respect to N_1 and $p(N_1|P_2) \geq 1 - \varepsilon_1$, then $N_1 \subset_{\varepsilon_2} R(P_2)$ with $\varepsilon_2^2 = \varepsilon + \varepsilon_1$.

Proof. Suppose u in N_1, $\|u\| = 1$. Then $\left| \|P_2 u\|^2 - p(N_1|P_2) \right| \leq \varepsilon$ by definition III.6:39. Thus $\|P_2 u\|^2 \geq p(N_1|P_2) - \varepsilon \geq 1 - \varepsilon - \varepsilon_1 = 1 - \varepsilon_2^2$ and the result follows from theorem I.1:15. □

The following theorem generalizes theorem III.1:8

Theorem III.6:43. Suppose $(P_1|P_2)$ is ε-equiangular and that $M_1 = R(P_1)$. If $p(P_1|P_2) > \varepsilon$, then $P_2 M_1$ is closed. If $p(P_1|P_2) < 1 - \varepsilon$, then $P_2^c M_1$ is closed.

Proof. Set $p = p(P_1|P_2)$ and suppose u is in M_1. If $p > \varepsilon$, then

$$\|P_2 u\|^2 \geq (p - \varepsilon)\|u\|^2.$$

Thus, by theorems I.1:15, I.1:12 and definition I.1:14, $\operatorname{dist}(M_1, R(P_2)) < 1$ and $P_2 M_1$ is closed. Interchanging P_2, P_2^c, the second statement follows. □

Definition III.6:44. $(P_1, \ldots, P_n)$ is (p_0, ε)-equiangular with respect to N if $p_0 > \varepsilon$ and

1) $\left| \|\underline{P}_n \cdots \underline{P}_1 u'\|^2 - \|\underline{P}_n \cdots \underline{P}_1 u''\|^2 \right| \leq \varepsilon$
for every u', u'' in N with $\|u'\| = \|u''\| = 1$

2) for every $\underline{P}_1, \ldots, \underline{P}_n$ we have either

   a) $\|\underline{P}_n \cdots \underline{P}_1 u\|^2 \geq p_0$ for every u in N with $\|u\| = 1$, or

   b) $\|\underline{P}_n \cdots \underline{P}_1 u\|^2 \leq \varepsilon$ for every u in N with $\|u\| = 1$

3) all subspaces $\underline{P}_n \cdots \underline{P}_1 N$ with $(\underline{P}_1, \ldots, \underline{P}_n)$ satisfying 2a) are ε-orthogonal.

$(P_1', \ldots, P_m'|P_1, \ldots, P_n)$ is called (p_0, ε)-equiangular if $(P_1, \ldots, P_n)$ is (p_0, ε)-equiangular with respect to $N = P_m' \cdots P_2' R(P_1')$.

If $(P_1, \ldots, P_n)$ is equiangular (0-equiangular) with respect to N and $0 < p_0 \leq$ all possible values of $p(N|\underline{P}_1, \ldots, \underline{P}_n)$, then $(P_1, \ldots, P_n)$ is obviously $(p_0, 0)$-equiangular. Conversely, if $(P_1, \ldots, P_n)$ is $(p_0, 0)$-equiangular with respect to N it is also equiangular with respect to N. To see this, condition 2) of definition III.2:15 is obvious and condition 1) then follows from 1) of definition III.6:44 by an argument similar to the proof of theorem III.2:21 For the case $n = 1$, P_2 is (p_0, ε)-equiangular with respect to N_1 if P_2 is ε-equiangular with respect to N_1 and $\varepsilon < p_0 \leq$ the least positive value of $p(N_1|P_2) - \varepsilon$ and $p(N_1|P_2^c) - \varepsilon$.

Theorem III.6:45. Set $N_j = P_j \cdots P_1 N_0$, $j = 1, \ldots, n$. If $N_{j-1} \subset_\varepsilon R(P_j)$, $j = 1, \ldots, n$, $n\varepsilon^2 < 1$, and $n\varepsilon^2 < p_0 \leq 1 - n\varepsilon^2$, then $(P_1, \ldots, P_n)$ is $(p_0, n\varepsilon^2)$-equiangular with respect to N_0.

Proof. Let u be an arbitrary vector in N_0 with $\|u\| = 1$. Then, by theorem I.1:15,

$$\begin{aligned} 1 &\geq \|P_n \cdots P_1 u\|^2 \geq (1 - \varepsilon^2)\|P_{n-1} \cdots P_1 u\|^2 \geq \ldots \\ &\geq (1 - \varepsilon^2)^n \|u\|^2 \geq (1 - n\varepsilon^2) \geq p_0. \end{aligned}$$

If some $\underline{P}_1, \ldots, \underline{P}_n$ is P_k^c, then

$$\|\underline{P}_n \cdots \underline{P}_1 u\|^2 \leq \|P_k^c \underline{P}_{k-1} \cdots \underline{P}_1 u\|^2 \leq \varepsilon^2 \|\underline{P}_{k-1} \cdots \underline{P}_1 u\|^2 \leq \varepsilon^2.$$

Conditions 1)–3) of definition III.6:44 obviously follow from this. □

III.7 Stochastic Event Structures of (p_0, ε)-equiangular Sequences

We shall now modify theorem III.4:33 to take into account processes, build on a given time-ordered set E_0 of projections, which are only (p_0, ε)-equiangular.

Let us recall (see section II.3) that when we are discussing time-ordered stochastic event structures, processes which can be obtained from each other by permutations, within the premise and within the outcome, conserving the time order, are considered as equivalent. By a time-ordered simple stochastic event structure spanned by a process $(E|E')$ we shall mean a structure S, which, together with a subprocess Q of $(E|E')$, contains all formal processes equivalent to Q and such that all processes in S are of this form.

Theorem III.7:46. Let $(E_0, |)$ be a time-ordered set of projections satisfying 1) and 2) of theorem III.4:33 and $(E|E') = (P_1, \dots, P_m|P'_1, \dots, P'_n)$ be a (p_0, ε)-equiangular (m, n)-tuple of projections in E_0 with $2^{n+1}\varepsilon \leq 1$, which is time-ordered i.e. satisfies 3) of theorem III.4:33. Let $-P$ denote P^c. Suppose u is a vector in $N = P_m \cdots P_2 R(P_1)$ with $\|u\| = 1$. Then there exists a linearly time-ordered minimal simple stochastic event structure $S = (E_0, -, T, p, |)$ spanned by $(E|E')$ such that

1° $p(P_1, \dots, P_m|\underline{P}'_1, \dots, \underline{P}'_n) = \|\underline{P}'_n \cdots \underline{P}'_1 u\|^2 + \delta$, $\delta \leq 2\varepsilon$

and, for every u' in N with $\|u'\| = 1$

2° $\left| p(P_1, \dots, P_m|\underline{P}'_1, \dots, \underline{P}'_n) - \|\underline{P}'_n \cdots \underline{P}'_1 u'\|^2 \right| \leq \varepsilon_1 = \varepsilon + \delta$

3° $\left| p(P_1, \dots, P_m|\underline{P}''_1, \dots, \underline{P}''_k) - \|\underline{P}''_k \cdots \underline{P}''_1 u'\|^2 \right| \leq \varepsilon_2 = 2^{n-1}(3\varepsilon + \delta)$ for every $(P''_1, \dots, P''_k) \subset E' = (P'_1, \dots, P'_n)$

4° $\left| p(P_1, \dots, P_m, P''_1, \dots, P''_k|P'''_1, \dots, P'''_l) - \|P''_l \cdots P''_1 P''_k \cdots P''_1 u'\|^2 / \|P''_k \cdots P''_1 u'\|^2 \right|$
$\leq \varepsilon_3 = \dfrac{2\varepsilon_2}{p_0}$ for every $(P''_1, \dots, P''_k, P'''_1, \cdots, P'''_l) \subset$ some $\underline{E}'$, with
$p(P_1, \dots, P_m|P''_1, \dots, P''_k) \geq p_0$.

1°– 4° also hold if $(P_1, \dots, P_m)$ or $(P'_1, \dots, P'_n)$ are replaced by any time-ordered permutations.

Assuming $p_0 > \delta + 2(2^n - 1)\varepsilon_1$ there also exists a time-ordered minimal simple stochastic event structure $S' = (E_0, -, T, p', |)$ spanned by $(E|E')$ with p' satisfying 2°, 3° and 4° with ε_1, ε_2 and ε_3 replaced by $\varepsilon'_1 = 2^n\varepsilon_1$, $\varepsilon'_2 = 2^n(2^n - 1)\varepsilon_1 + \varepsilon_2$ and $\varepsilon'_3 = 2\varepsilon'_2/p_0$ and such that

5° for every $\underline{E}'$, either $p'(E|\underline{E}') \geq p'_0 = p_0 - \delta - (2^n - 1)\varepsilon_1$ or $p'(E|\underline{E}') = 0$.

Proof. Let $p_1(P_1, \dots, P_m|\underline{P}'_1, \dots, \underline{P}'_n) = \|\underline{P}'_n \cdots \underline{P}'_1 u\|^2 + \delta$ with

$$\sum_{\underline{E}'} p_1(P_1, \dots, P_m|\underline{P}'_1, \dots, \underline{P}'_n) = \sum_{\underline{E}'} \|\underline{P}'_n \cdots \underline{P}'_1 u\|^2 + 2^n\delta = 1. \tag{1}$$

For every $(E''|E''')$ where E'' and E''' are time-ordered permutations of E and E' respectively, we define $p_1'(E''|\underline{E}''') = p_1(E|\underline{E}')$. By theorem II.2:18, p_1 can then, for every $(E''|E''')$ be extended to a p_1' such that $(E_0, -, T_0', p_1')$ is a minimal simple stochastic event structure and where T_0' is a set of formal subprocesses of $(E''|E''')$. By the definition of the p_1':s and theorem II.2:17, subprocesses, which are equivalent with respect to the time-ordering then have the same p-values so $S = (E_0, -, T, p, |)$, with T equal to the union of all the T_0' and p joining all the p_1', is a time-ordered simple stochastic event structure satisfying 1°. 2° then follows from 1° and the definition III.6:44. For δ see below.

To prove 3°, we set for a given $E'' = (\underline{P}_1'', \dots, \underline{P}_k'') \subset$ some $\underline{E}'$, $s = \underline{P}_k'' \cdots \underline{P}_1'' u' = \sum' u_i$, where u_i, $i = 1, \dots, 2^n$, denotes the 2^n vectors $\underline{P}_n' \cdots \underline{P}_1' u'$ and the sum is over all $\underline{E}' = (\underline{P}_1', \dots, \underline{P}_n') \supset E''$.

Let n_s be the number of terms in the sum. Obviously we can suppose $k < n$ and $n_s \leq 2^{n-1}$, the case $k = n$ being covered by 2°.

For the sum we get, by 3) of definition III.6:44

$$\|s\|^2 = \sum{}' \|u_i\|^2 + R \tag{2}$$

where $R = \sum_{i \neq j}' \langle u_i, u_j \rangle$ and

$$|R| \leq \varepsilon \sum{}_{i \neq j}' \|u_i\| \cdot \|u_j\| \leq (n_s - 1)\varepsilon \sum{}' \|u_i\|^2 = (n_s - 1)\varepsilon \Big(\|s\|^2 - R \Big)$$

and, since obviously $\|s\| \leq 1$,

$$|R| \leq \frac{(n_s - 1)\varepsilon}{1 - (n_s - 1)\varepsilon} \leq 2(n_s - 1)\varepsilon \tag{3}$$

since $(n_s - 1)\varepsilon \leq 2^{n-1}\varepsilon \leq \frac{1}{2}$ by assumption.

But

$$\Big| \sum{}' \|u_i\|^2 - \sum{}' \|\underline{P}_n' \cdots \underline{P}_1' u\|^2 \Big| \leq n_s \varepsilon \tag{4}$$

by 1) of definition III.6:44, and

$$\begin{aligned} \sum{}' \|\underline{P}_n' \cdots \underline{P}_1' u\|^2 + n_s \varepsilon &= \sum{}' p(P_1, \dots, P_m | \underline{P}_1', \dots, \underline{P}_n') \\ &= p(P_1, \dots, P_m | \underline{P}_1'', \dots, \underline{P}_k'') \end{aligned} \tag{5}$$

by 1) and 2b) of definition II.1:1. Putting (1)–(5) together, we get

$$\begin{aligned} &\Big| p(P_1, \dots, P_m | \underline{P}_1'', \dots, \underline{P}_k'') - \|\underline{P}_k'' \cdots \underline{P}_1'' u'\|^2 \Big| \leq \\ n_s \delta + \Big| \sum{}' \|\underline{P}_n' \cdots \underline{P}_1' u\|^2 - \sum{}' \|u_i\|^2 \Big| + \Big| \sum \|u_i\|^2 - \|\underline{P}_k'' \cdots \underline{P}_1'' u'\|^2 \Big| &\leq \\ \leq n_s \cdot \delta + n_s \cdot \varepsilon + 2(n_s - 1)\varepsilon &\leq 2^{n-1}(3\varepsilon + \delta). \end{aligned}$$

Using the preceding result and formula (2) in the proof of lemma II.8:48, we get

for the left member of 4°

$$\left| \frac{p(P_1,\dots,P_m|P_1'',\dots,P_k'',P_1''',\dots,P_l''')}{p(P_1,\dots,P_m|P_1'',\dots,P_k'')} - \frac{\|P_l'''\cdots P_1'''P_k''\cdots P_1''u'\|^2}{\|P_k''\cdots P_1''u'\|^2} \right| \leq 2\cdot\frac{1}{p_0}\cdot\varepsilon_2 = \varepsilon_3$$

Using the arguments for the proof of condition 3°, putting $u' = u$, E'' empty and thus $n_s = 2^n$, $\|s\| = 1$ and $2^n\delta$ taking the place of $|R|$ according to (1) we get by (1), (2) and (3) $2^n\delta \leq \varepsilon(2^n-1)\cdot(1+2^n\delta)$ and, since by assumption $2^n\varepsilon \leq \frac{1}{2}$, $2^n\delta \leq 2\cdot(2^n-1)\varepsilon$ and $\delta \leq 2\varepsilon$.

By the assumed commutation property 1) and 2) of theorem III.4:33, 1°–4° are still valid if $(P_1,\dots,P_m)$ or $(P_1',\dots,P_n')$ are replaced by any time-ordered permutations.

To prove the existence of S', we note that, by 1° and by the assumption of (p_0,ε)-equiangularity, every $p(E|\underline{E}')$ is either $\geq p_0-\delta$ or $\leq \varepsilon+\delta = \varepsilon_1$. We can then use lemma II.8:49 to construct an S' such that 5° is satisfied. This is possible since, by assumption, $p_0-\delta > 2(2^n-1)\varepsilon_1$. 2° and 3° then follow from 4° of lemma II.8:49 and 1° of lemma II.8:48 respectively and 4° follows from 3° as above. □

III.8 Physical Remarks

Differences to the Conventional Quantum-Mechanical Formalism

The present theory can be considered as embedded in the conventional quantum-mechanical formalism. Our series of successive events (projections P) could be considered as a series of conventional quantum-mechanical measurements, with a special prescription for preparation of a state after measurement, namely that measuring P on a "state" u results in "state" Pu. We are then considering apparently special series of measurements of localization observables, (leaving out most of the conventional formalism such as canonical commutation relations, complete sets of commuting variables, density operators, etc.). The restriction to equiangular sequences then means that any state vector (or density operator), which can be prepared from some other previous state by successive measurements of the observables describing the initial conditions, will give (approximately) the same probabilities for the following measurements, so these probabilities can be calculated from a knowledge of these initial observables only.

However, a different point of view is to consider the restriction to equiangular sequences of localization statements as an extra dynamical postulate (a "principle of equiangularity"), restricting the possible combinations of events. This extra postulate is lacking in the conventional formalism, where any state vector or any density operator is a possible state and any selfadjoint operator is a possible observable, which can be measured at any time by applying a suitable external measuring equipment.

In conventional quantum-mechanical formalism, the measurement process has received a central position, connected with the interpretation of the theory. In the present theory, a stochastic quantum transition is considered as a fundamental

objective occurrence in reality – it is not considered, as in conventional quantum mechanics, to be a disturbance caused by a measurement. In those cases, where we actually have a measuring equipment, measuring a certain observable P, we can include the measurement apparatus in a larger process containing both the object of measurement and the measurement apparatus. That an object can be forced to undergo a quantum transition is not a phenomenon reserved for measurement situations. Quite generally, the initial and boundary conditions forces a system to undergo quantum transitions. This is described by the concept of stochastic event structure.

To summarize, the present theory differs from the conventional formalism in the following respects:

1° The only operators postulated to correspond to observables are the projection operators corresponding to space localization. All other physical quantities will be indirectly defined in terms of these.

2° We avoid use of the concept of "state", described by state vectors or density operators. Instead the theory is based on a description of initial conditions by means of a series of previously occurred events at different times described by the observables according to 1°. The restriction to equiangular sequences of projections makes it possible to determine transition probabilities completely from the projections describing the initial conditions.

3° We consider the "collapse" of the wave functions (transition from u to Pu in our cases) as an integrated part of the formalism. It is not pushed aside to an interpretation of the theory connected with a measurement process. The concept of equiangular sequences of projections describes a series of successive, really occurred "collapses". The observables thus describe objectively occurred events.

4° Instead of allowing more or less arbitrary state vectors or density operators and measurement of observables corresponding to arbitrary selfadjoint operators, the restriction to equiangular sequences of projections puts a strong restriction on which series of events are possible.

5° Instead of introducing an extra statistical distribution (density operator), extra assumptions concerning this distribution and assumptions that certain mean values describe macroscopic systems, the present theory handles macroscopic systems in a direct and deductive way. The confidence estimates can be used at different levels of description. Macroscopic estimates concerning gross distributions of large number of particles can be derived directly from the wave equations just as e.g. a cross-section formula or an estimate of a bound-state energy level can be.

Summary

General principles for deductive physical theories are discussed. It is claimed that a deductive physical theory should in principle be a pure mathematical theory (or a set of coupled mathematical theories) together with an identification of certain quantities/concepts ("observables") in the theory and corresponding observable entities in the real world. This identification – the "interpretation" of the theory – should be unproblematic, both for the theoretician and the experimentalist.

A general basis for a deductive physical theory, comprising both classical and quantum physics in a unified way, is proposed. The theory is based on successive confidence estimates on quantum-mechanical wave functions corresponding to space-localizations of particles. This allows a direct and simple way of describing both macroscopic and microscopic phenomena by means of the same basic concepts. Especially, this gives a simple, direct, kinetic, radical alternative to the ensemble methods of classical and quantum statistical mechanics.

The theory takes as its starting point the general basic ideas and problematics of quantum theory that was formulated in the 1920:s. The theory is thus consistent with conventional quantum mechanics in the sense that it is based on the same mathematical formalism of Hilbert spaces, projection operators, the Schrödinger equation (and its relativistic generalization, the Schrödinger-Schwinger-Tomonaga equation), etc., together with the – although from an axiomatic point of view, as formulated, unsatisfactory – original primitive statistical interpretation.

However, it is claimed that this is only half of the theory – half the set of conditions in a complete set of axioms. This leaves a manifest and obvious ambiguity. It is claimed that this is the root of the controversial interpretation problems and paradoxes of the conventional expositions of quantum theory.

Central in the axiomatics of the outlined theory is the concept of equiangular sequences of projections (projection operators). It describes a successive sequence of "collapses of the wave function". It is proposed that the restriction of general physical processes to fit an underlying structure of equiangular sequences – a "principle of equiangularity" – together with the restriction to projections corresponding to space-localizations of particles could give the extra conditions, constituting the other half of the theory.

From equiangular sequences of projections is abstracted the general structure of "stochastic event structure". It gives an axiomatization of the ordinary (classical!) probability theory (based on classical – not "quantum" – logic) and, at the same time, an axiomatization of the concept of causality, which generalizes the ordinary "deterministic" causality to what we call "stochastic causality". It can be applied to problems far beyond physics.

Characteristic of the outlined theory is the avoidance of the concepts of "states" and "systems" as basic concepts. The basic concept of the theory is the concept of events, represented by projection operators corresponding to confidence estimates of localization of particles to many-particle space regions.

An event can be characterized as a "partial statement" about the actual physical situation. Thus any concrete physical situation is described by a more or less exhausting set of partial statements, complementing each other. This description

can often be complemented by other events, for instance on a deeper level of description. The set of – simultaneously often "overlapping" – partial statements can, in the quantum domain, generally not be reduced to the classical concepts of "system" and "state" and the incomplete instantaneous specification of the situation is complemented by giving events at different times. This is an expression of the so called "quantum unity".

The avoidance of the concept of state as basic concept means that the wave functions, and the corresponding vectors (or rays) in the Hilbert space are not (generally) given the status or meaning as states. The term "state vector" is thus abandoned. The time-dependent wave functions are instantaneously coupled to events of instantaneous space localizations. At other times they are to be considered as auxiliary dynamical quantities for determining probabilities. This is in accordance with the original Heisenberg idea of the S-matrix.

If we detail a description of a process described by events at times t_1 and t_2, $t_1 < t_2$, by intercalating extra events $e', \ldots$ at times $t', \ldots$, between t_1 and t_2, the resulting time-dependent wave function(s) extrapolated to t_2 will, due to the collapses corresponding to $e', \ldots$ change. The axioms of equiangular sequences guarantee the consistency of such intercalations of events.

In the cat paradox case the process can be detailed by a description of what really happens – when the poisoning capsule explodes, when and how the cat dies etc., in case of a finally found dead cat – or, what the cat did during the process, in case it comes out living.

On the other hand, quantum mechanics puts strong limits on what can be detailed with respect to the time development. In a two-split experiment with a single electron we cannot say that the electron has passed through one and not the other slit. An arrangement that would determine through which slit the electron passes is *incompatible* with the two-split arrangement in the sense of the definition of *compatibility* given in the theory.

The limits on detailization is determined by the restrictions of equiangularity and the ultimate restriction to space-localizations together with the dynamics of wave-mechanics.

Another characteristic of the outlined theory is that it is not – contrary to the conventional quantum theory – based on the concept of measurement. The events are to be considered as really occurred "elements of reality" irrespectively of whether or not a systematic measurement or observation is coupled to the object system under concern. For a discussion of measurements in the proposed theory, see Part III.

The concept of equiangular sequences of projections can be considered as a generalization of the concept of S-matrix (and its factorizations/subdivisions into subprocesses) to finite regions in space and time.

It is also a characteristic of the proposed theory that it is fundamentally indeterministic/stochastic. This is contrary to common ideas that physical laws are fundamentally deterministic and time-reversible. This determinism and time-reversibility is formal and concerns only the one half of the theory mentioned above.

The proposed theory gives a basis for a general theory of irreversible processes based directly on quantum mechanics. It gives an alternative definition of entropy

and an alternative derivation of entropy increase in irreversible processes. It shows a deep relation between thermodynamics and quantum theory.

Irrespective of physical applicability, the concepts of confidence estimates on L^2-functions and their Fourier transforms, equiangular sequences of projections and stochastic event structures have interesting properties that deserve a separate, pure mathematical study, see Part III.

References to Parts I and II

(1) T. Blomberg; *Successive Confidence Estimates on Solutions to the Many-Particle Schrödinger Equation. Basic Concepts*. Stockholm 1987 (unpublished). Included in this book as Part III.
(2) D. Bohm: *Quantum Theory*. Constable and Company Ltd. 1954.
(3) J.M. Jauch: *Foundations of Quantum Mechanics*. Addison-Wesley 1968.
(4) A.I. Khinchin: *Mathematical Foundations of Statistical Mechanics*.
(5) W. Gibbs: *Elementary Principles of Statistical Mechanics*. Yale University Press, 1902.
(6) R.C. Tolman: *The Principles of Statistical Mechanics*. Oxford, Clarendon Press, 1938.
(7) E.T. Jaynes: *Information Theory and Statistical Mechanics*. Physical Review, 106, 620 (1957.
(8) S.S. Schweber: *An Introduction to Relativistic Quantum Field Theory*. Harper & Row, 1961.

References to Part III

(1) T. Kato: *Perturbation Theory for Linear operators*. Springer-Verlag, Berlin 1966

Buy your books fast and straightforward online - at one of the world's fastest growing online book stores! Environmentally sound due to Print-on-Demand technologies.

Buy your books online at

www.get-morebooks.com

Kaufen Sie Ihre Bücher schnell und unkompliziert online – auf einer der am schnellsten wachsenden Buchhandelsplattformen weltweit!
Dank Print-On-Demand umwelt- und ressourcenschonend produziert.

Bücher schneller online kaufen

www.morebooks.de

OmniScriptum Marketing DEU GmbH
Heinrich-Böcking-Str. 6-8
D - 66121 Saarbrücken
Telefax: +49 681 93 81 567-9

info@omniscriptum.com
www.omniscriptum.com

OMNIScriptum