International Journal of Theoretical Physics, Vol. 38, No. 7, 1999

Quantum Mechanics from Symmetry and Statistical Modeling

Inge S. Helland[1]

Received November 25, 1998

Quantum theory is derived from a set of plausible assumptions related to the following general setting: For a given system there is a set of experiments that can be performed, and for each such experiment an ordinary statistical model is defined. The parameters of the single experiments are functions of a hyperparameter which defines the state of the system. There is a symmetry group acting on the hyperparameters, and for the induced action on the parameters of the single experiment a simple consistency property is assumed, called permissibility of the parametric function. The other assumptions needed are rather weak. The derivation relies partly on quantum logic, partly on a group representation of the hyperparameter group, where the invariant spaces are shown to be in 1–1 correspondence with the equivalence classes of permissible parametric functions. Planck's constant only plays a role connected to generators of unitary group representations.

1. INTRODUCTION

The two great revolutions in physics at the beginning of this century—relativity and quantum mechanics—still influence nearly all aspects of theoretical physics. Similar as they may be, both in their impact on modern science and in the way they in their time turned conventional ideas upside down, there are also of course great differences—both in origin, appearance, and type of content. Relativity theory was founded by one man, Einstein, while the ideas of quantum theory developed over time through the work of many people, most notably Planck, Bohr, Schrödinger, Heisenberg, Pauli, and Dirac. An equally important—and perhaps related—aspect is the following: Relativity theory can be developed logically from a few intuitively clear, nearly obvious concepts and axioms, essentially only constancy of the speed

[1] Department of Mathematics, University of Oslo, N-0316 Oslo, Norway; e-mail: ingeh@math.uio.no.

0020-7748/99/0700-185 1$16.00/0

of light and invariance of physical laws under change of coordinate system, while quantum theory still has a rather awkward foundation in its abstract concepts for states and observables.

Over the years, several attempts to provide a deeper foundation for quantum theory have been made; some of these we will return to later. Although there is far from a universal agreement on the foundation, today's physicists, both theoretical and experimental, have developed a clear intuition directly connected to states of a system as rays in a complex Hilbert space and observables as self-adjoint operators in the same space. The theory has had success in very many fields—some claim that quantum theory is the most successful physical theory ever advanced, but it has also met problems: difficulties with defining the border between object and observer in von Neuman's quantum measurement theory; difficulties with interpretations requiring many worlds or action at a distance; infinities in quantum field theory requiring complicated renormalization programs; difficulties in reconciling the theory with general relativity and so on. We will of course not try to attack all these problems here. What we will assert, though, is that the fact that such difficulties occur does make it legitimate to look at the foundation of the theory with fresh eyes. One way to do this is to try to find a foundation which is in accordance with common sense. Another way is to compare it with another, apparently unrelated, theoretical area. In this paper we will try to combine both these lines of attack.

A vital clue is the role of probability in quantum theory. In the beginning, this was an aspect that overshadowed all other difficulties in the theory, and that made leading physicists—first of them Einstein himself—skeptical: The new physical laws were then and are still claimed to be probabilistic by necessity. Still some people are looking at hidden variable theories in attempts to avoid the fact that the fundamental laws of nature are stochastic, but after the experiments of Aspect *et al.* (1982) and other overwhelming evidence, most scientists seem to accept stochasticity of nature as an established fact.

At the same time, statistical science has developed methodology that has found applications in an increasing number of empirical sciences, methodology largely based upon stochastic modeling. With this background, the obvious, but apparently very difficult question then is: Why has there been virtually no scientific contact between physicists and statisticians throughout this century? This lack of contact is in fact very striking. At the same time as Dirac was developing a foundation for relativistic quantum theory in Cambridge in England, R. A. Fisher completely independently developed the foundation of statistical inference theory based upon probability models in Rothamsted and London. While modern quantum field theory was being developed by Feynman in Princeton and Schwinger in Harvard, J. Neyman and coworkers were laying the foundation of modern statistical inference

theory in Berkeley. One of the few early contacts that I know of is Feynman's (1951) Berkeley Symposium paper on the interpretation of probabilities in quantum mechanics. Today, quantum theory sometimes has its own session at large international statistical conferences, but the language spoken there is of a nature which is difficult to understand for ordinary statisticians. Meyer's (1993) book on noncommutative probability theory may be seen as an attempt to make a synthesis of the two worlds, but this book does not address the foundation question, at least if we seek a foundation related to common sense. It is also becoming increasingly apparent that there are similarities between advanced probability theory and quantum theory, but these similarities seem to be mostly at the formal level.

The lack of a common ground for modern physics and statistics is even more surprising when we know that the outcome of any single experiment with fixed experimental arrangement can always be described by ordinary probability models, also in the world of particle physics. It is in cases where several arrangements are possible, as when one has the choice between measuring the position and momentum of a particle, that quantum mechanics gives results which cannot be reached by ordinary probability theory. In addition, quantum theory gives definite rules for *computing* probabilities, also in cases where the ordinary probability concept can be used in principle.

We will formulate below an extended experimental setting which includes the possible decisions which must be made before the experiment itself is carried out, at least before the final inference is made. We will look at the situation where strong symmetries exist both within single experiments and in the wider experimental setting between single experiments. This, together with other reasonable assumptions, will lead to probability models of the type found in ordinary quantum mechanics.

The physical implications of this theory, at least in its nonrelativistic variant, are not expected to differ considerably from existing quantum theory; basic mathematical equivalence will be discussed later. Our main contribution here is to propose a simpler and more intuitive foundation. The really interesting further challenge seems to be the question of finding a corresponding relativistic theory. This issue will not be pursued here, but in light of the importance of symmetry groups in high-energy physics, it seems very plausible that it should be possible to develop the approach in this direction, too. A brief discussion on this will be given in the last section.

Our aim is to write the paper in such a way that the main principles can be appreciated by theoretical physicists, mathematicians, and statisticians. Technical details at several points are unavoidable, however. Also, even though we will try to be fairly precise, at least in the main results, there may still be room for improvement in mathematical rigor. What we feel are the

most important assumptions are stated explicitly. Minor technical assumptions are stated in the text.

2. EXPERIMENTAL SETTING

The common statistical framework for analyzing an experiment is a sample space $\mathcal{X}$, listing the possible experimental outcomes, a fixed σ-algebra (Boolean algebra) $\mathcal{F}$ of subsets of $\mathcal{X}$, and a class $\{P_\theta; \theta \in \Theta\}$ of probability measures on the measurable space $(\mathcal{X}, \mathcal{F})$. The parameter θ—or a function of this parameter—is ordinarily the unknown quantity which the statistician aims at saying something about using the outcome of the experiment. A fixed θ, or alternatively, a probability distribution expressing prior knowledge about θ, may also be related to the physicist's concept of 'state.' A simple purpose of a statistical experiment might be to *estimate* θ, which formally means to select a function $\hat{\theta}$ on the sample space $\mathcal{X}$ such that $\hat{\theta}(x)$ is a reasonable estimate of θ when the observation x is given. There is a considerable literature on statistical inference; three good and thorough books with different perspectives are Berger (1985), Lehmann (1983), and Cox and Hinkley (1974). Both at the more specialized and at the more elementary level there are very many books, of course. An important point is that intuition related to statistical methodology in this ordinary sense has been developed in a large number of empirical sciences, also in parts of experimental physics.

A very general approach to statistics assumes a decision-theoretic framework: First, a space D of possible decisions is defined, given the experimental outcome; for instance, D can consist of different estimators of θ. Then a loss function $L(\theta, d(\cdot))$ is specified, giving the loss by taking the decision d when the true parameter is θ. The choice of decision is typically done by minimizing the expected loss. By restricting the class of decision functions so that they posess invariance properties, or so that estimators have the correct expectation, this can often be done uniformly over the unknown parameter. Another possible course—gaining increasing popularity—is to assume a prior probability distribution over the parameter space, and then base inference on the corresponding posterior distribution, given the data. This is called the Bayesian approach.

Essential for what follows is that we will extend the traditional statistical framework to include possible actions taken by an experimenter before or during a given experiment. The most important actions for our purposes are those that label the whole experiment and thus allow different experiments to be done in the same situation. However, we will also allow actions that change the class of probability measures, the space of decisions, and/or the loss function.

Thus we start with a space $\mathcal{A}$ of possible actions, and for each $a \in \mathcal{A}$ we have an experiment $\mathcal{E}_a$, consisting of a probability model $(\mathcal{X}_a, \mathcal{F}_a, \{P_\theta^a; \theta \in \Theta_a\})$, and possibly a loss function $L_a(\cdot, \cdot)$ and a space D_a of potential decisions. In macroscopic experiments such actions are in fact very common. They are not usually explicitly taken into account in the statistical analysis, but are taken as fixed once and for all, an attitude which is fairly obvious in some cases, but absolutely can be discussed in other cases. In fact, more can be said, also in the ordinary statistical setting, for a closer link between the experimental design phase (choosing a) and the statistical analysis phase. A partial list of possibilities for choosing a include the following:

(a) Choice between a number of essentially different experiments that are possible to perform in a given situation.

(b) Choice of target population, and way to select experimental units, including choice of randomization. Choice of conditioning in models is related to these issues.

(c) Choice of treatments. There are many variants: There are lots of examples of medical treatments which are mutually exclusive. In factorial experiments the choice of factors and the levels of these are important issues.

(d) A choice between different statistical models; for instance, one may want to reduce the number of parameters if the model is too complicated to give firm decisions. In fact, this is permissible under certain conditions.

(e) Factorial experiments often turn out to be unbalanced, more specifically, nonorthogonal: If n_{ij} denotes the number of experimental units having level i if factor I and level j of factor J, then the statistical analysis is simple only if $n_{ij} = n_{i.}\, n_{.j}/n_{..}$, where the dots denote summation. If this is not the case, the experiment is said to be nonorthogonal. Then the statistical test for equal effects of the levels of factor I depends on whether or not factor J is included in the model. So in this sense the consecutive order of the factors influences the conclusion; hence this must be decided upon. Simpson's famous paradox [see Gudder (1988) for references and for a discussion from the point of view of quantum theory] is a different, but related issue. Even though factorial experiments with many factors used on the same units—if possible—are known both from experience and from statistical theory to be very efficient, the theoretical discussion in the present paper is more related to the situation where one uses one or a few factors at a time and performs consecutive experiments.

The main purpose of this paper will be to explore the relationship between macroscopic statistical modeling and the microscopic modeling we find in the quantum mechanical world. In the latter one definitively has the choice between performing different experiments, say, measuring position or momentum or measuring spin in different directions. The quantum mechanical

state of a particle or a particle system can be used to predict the outcome of any given one of these experiments once the choice is made. In this way one might say that the quantum mechanical state vector contains the simultaneous model of a large set of possible statistical experiments $\mathscr{E}_a$ ($a \in \mathscr{A}$). A major reason why this is possible is that the situation contains a high degree of symmetry. To study the connection between quantum mechanics and statistics more directly, it turns out, however, to be useful first to consider a simpler representation than the ordinary Hilbert space representation of quantum mechanics.

3. THE LATTICE APPROACH TO QUANTUM THEORY

Mathematically, a σ-lattice is defined as a partially ordered set $\mathscr{L}$ such that the infimum and supremum (with respect to the given ordering) of every countable subset exist and belong to $\mathscr{L}$. In this section we will summarize the approach to quantum mechanics taking lattices of propositions as points of departure. These can be looked upon as generalizations of the σ-algebras of classical probability, where the propositions are subsets of a given space $\mathscr{X}$, ordered by set inclusion; the term Boolean algebra is sometimes used for the same concept. For Boolean algebras, the partial ordering corresponds to set inclusion, and infimum and supremum correspond to intersection and union, respectively. In the following we will mainly consider (complete) lattices, where the infimum and supremum of all subsets, not only the countable ones, exist.

The lattice approach to quantum mechanics consists in formulating a number of axioms for a lattice which is weaker than the set of axioms needed to define a Boolean algebra, just sufficiently weak that they are satisfied by the set of projections in a Hilbert space, which are the entities that represent propositions in conventional quantum mechanics. We will follow largely Beltrametti and Cassinelli (1981) in presenting these axioms. It is relatively easy to prove that the axioms below actually are satisfied by Hilbert space projections, much more difficult to show that the axioms by necessity imply that the lattice has a Hilbert space representation. We proceed with the necessary definitions; in the next section we will try to relate these definitions to sets of potential experiments as formulated above.

We have already defined a lattice $\mathscr{L}$ as a set of propositions $\{\mathscr{P}\}$ with a partial ordering $\leq$ such that the supremum $\vee_i \mathscr{P}_i$ exists and belongs to $\mathscr{L}$ when all the $\mathscr{P}_i$ belong to $\mathscr{L}$, similarly for infimum. ($\vee i\ \mathscr{P}_i$ is defined as a proposition $\mathscr{P}$ such that $\mathscr{P}_i \leq \mathscr{P}$ for all $\mathscr{P}_i$, and such that $\mathscr{P}_i \leq \mathscr{P}_0$ for all $\mathscr{P}_i$ implies $\mathscr{P} \leq \mathscr{P}_0$. It is easy to show that the supremum, if it exists, is unique if the lattice is assumed to have the property that $\mathscr{P}_1 \leq \mathscr{P}_2$ and $\mathscr{P}_2 \leq \mathscr{P}_1$ implies $\mathscr{P}_1 = \mathscr{P}_2$. Infimum is defined correspondingly.)

It follows that $\mathscr{L}$ contains the infimum of all propositions, denoted **0**, and the supremum of all propositions, denoted **1**.

What corresponds to the complement in a Boolean algebra is here denoted by orthocomplement. An orthocomplementation in a lattice $\mathscr{L}$ is a mapping $\mathscr{P} \to \mathscr{P}^{\perp}$ of $\mathscr{L}$ onto itself such that (i) $\mathscr{P}^{\perp\perp} = \mathscr{P}$, (ii) $\mathscr{P}_1 \leq \mathscr{P}_2$ implies $\mathscr{P}_2^{\perp}$ le $\mathscr{P}_1^{\perp}$, (iii) $\mathscr{P} \wedge \mathscr{P}^{\perp} = \mathbf{0}$, and (iv) $\mathscr{P} \vee \mathscr{P}^{\perp} = \mathbf{1}$. It is easy to see that De Morgan's laws follow from (ii):

$$(\vee \mathscr{P}_i)^{\perp} = \wedge \mathscr{P}_i^{\perp}, \qquad (\wedge \mathscr{P}_i)^{\perp} = \vee \mathscr{P}_i^{\perp}$$

Two propositions $\mathscr{P}_1$ and $\mathscr{P}_2$ are said to be disjoint or orthogonal, written $\mathscr{P}_1 \perp \mathscr{P}_2$, when $\mathscr{P}_1 \leq \mathscr{P}_2^{\perp}$ (or, equivalently, when $\mathscr{P}_2 \leq \mathscr{P}_1^{\perp}$). A subset of $\mathscr{L}$ formed by pairwise disjoint elements is simply called a disjoint subset. The lattice $\mathscr{L}$ is called separable if every disjoint subset of $\mathscr{L}$ is at most countable.

If $\mathscr{P}_1 \leq \mathscr{P}_2$, we will write $\mathscr{P}_2 - \mathscr{P}_1$ for $\mathscr{P}_2 \wedge \mathscr{P}_1^{\perp}$. It is clear that we then have $\mathscr{P}_2 - \mathscr{P}_1 \perp \mathscr{P}_1$. A lattice $\mathscr{L}$ is called orthomodular if $\mathscr{P}_1 \leq \mathscr{P}_2$ implies $\mathscr{P}_2 = \mathscr{P}_1 \vee (\mathscr{P}_2 - \mathscr{P}_1)$. Another way to put this is that the distributive laws hold for the triple $(\mathscr{P}_1, \mathscr{P}_2, \mathscr{P}_1^{\perp})$ when $\mathscr{P}_1 \leq \mathscr{P}_2$.

A much stronger requirement is that the distributive laws should hold for all triples:

$$\mathscr{P}_1 \vee (\mathscr{P}_2 \wedge \mathscr{P}_3) = (\mathscr{P}_1 \vee \mathscr{P}_2) \wedge (\mathscr{P}_1 \vee \mathscr{P}_3),$$
$$\mathscr{P}_1 \wedge (\mathscr{P}_2 \vee \mathscr{P}_3) = (\mathscr{P}_1 \wedge \mathscr{P}_2) \vee (\mathscr{P}_1 \wedge \mathscr{P}_3)$$

An orthocomplemented distributive lattice is in general a Boolean algebra, and can always be realized as an algebra of subsets of a fixed set.

The lattices of quantum mechanics are orthocomplemented, but nondistributive; they only satisfy the weaker requirement of being orthomodular. Much of the pioneering work in the lattice approach to quantum mechanics is due to Mackey (1963). Two books on orthomodular lattices are Beran (1985) and Kalmbach (1983). This represents an approach to quantum mechanics that in some sense is more primitive than most other approaches, but there are still concepts here, like the lattice property or orthomodularity, which are difficult to get an intuitive relation to. Our aim here will be to try to understand these concepts to some extent in terms of ordinary statistical models for a set of potential experiments and in terms of symmetry properties.

In addition to orthomodularity, two further properties are required by Beltrametti and Cassinelli (1981): The lattices are assumed to be atomic and to have the covering property.

A nonzero element $\mathscr{P}_0$ of $\mathscr{L}$ is called an atom if $0 \leq \mathscr{P} \leq \mathscr{P}_0$ implies $\mathscr{P} = 0$ or $\mathscr{P} = \mathscr{P}_0$. The lattice $\mathscr{L}$ is called atomic if there always for $\mathscr{P} \neq 0$ exists an atom $\mathscr{P}_0$ such that $\mathscr{P}_0 \leq \mathscr{P}$. It can be shown that if $\mathscr{L}$ is an

orthomodular, atomic lattice, then every element of $\mathscr{L}$ is the union of the atoms that it contains. If the lattice is separable, this union is at most countable.

We say that $\mathscr{L}$ has the covering property if for every $\mathscr{P}_1$ in $\mathscr{L}$ and every atom $\mathscr{P}_0$ such that $\mathscr{P}_1 \wedge \mathscr{P}_0 = 0$, we have that $\mathscr{P}_1 \leq \mathscr{P}_2 \leq \mathscr{P}_1 \vee \mathscr{P}_0$ implies either $\mathscr{P}_2 = \mathscr{P}_1$ or $\mathscr{P}_2 = \mathscr{P}_1 \vee \mathscr{P}_0$.

In total, the kind of propositional structure which is promoted in Beltrametti and Cassinelli (1981) as a basis for quantum mechanics is an orthocomplemented, orthomodular, separable lattice which is atomic and has the covering property. All these requirements are proved to hold for the lattice of projection operators in a separable Hilbert space, or equivalently, the lattice of all closed subspaces of the Hilbert space.

One can also prove results in the opposite direction, though this is much more difficult: If the given requirements hold for some lattice, then one can construct an isomorphic Hilbert space in the sense that the projections upon subspaces of this Hilbert space are in one-to-one correspondence with the propositions of the lattice, with corresponding ordering. The proof of this last result is only hinted at in Beltrametti and Cassinelli (1981); more details are given in Piron (1976) and in Maeda and Maeda (1970). Related results can also be found in Varadarajan (1985). One problem is to convince oneself that the complex number field is the natural one to choose as a basis for the Hilbert space: One can also construct representations based on the real or quaternion number field.

Finally, the concept of state in quantum mechanics can be defined as a probability measure on the propositions of a lattice. In the Hilbert space representation the famous theorem of Gleason says that all states can be represented by density operators ρ [positive operators with trace $\mathrm{tr}(\rho) = 1$] in the sense that the expectation of every observator, represented by a self-adjoint operator A, is given by

$$\mathrm{tr}(A\rho)$$

As is well known from probability theory, the set of expectations for all variables determines the set of probability distributions for all variables.

4. A SET OF POTENTIAL STATISTICAL EXPERIMENTS

Let us start by turning to a completely different situation where the concept of 'state' is also being used. Consider a medical patient. One way to make precise what is meant by the state of this patient is to contemplate the potential results of all possible tests that the patient can be exposed to, where the word 'test' is used in a very wide sense, possibly including treatments or parts of treatments. Thus in a concrete setting like this we can imagine a large number of potential experiments, some possibly mutually

exclusive, and let the state be defined abstractly as the totality of probability distributions of results from these experiments, or some parameter determining all these probability distributions. A similar concept can be imagined for a set of several medical patients, where randomization and allocation of treatment may be included as part of the potential experiments under consideration. Here, the focus may be on the treatments rather than on the patients, but still the results of the experiments depend upon the state of the patients, and—if we consider a large enough collection of experiments for a large enough collection of patients in the same state—the potential outcomes may hopefully determine these states.

In general, now fix some concrete experimental setting, and let $\mathcal{A}$ be a set of potential experiments, in statistical terminology $\mathcal{E}_a = (\mathcal{X}_a, \mathcal{F}_a, \{P_\theta^a: \theta \in \Theta_a\})$ with decision space D_a and loss function $L_a(\theta, d(\cdot))$ for $a \in \mathcal{A}$.

For two experiments $\mathcal{E}_a$ and $\mathcal{E}_{a'}$ it may be crucial whether or not these can both be performed on the experimental unit(s)—say the patient—in such a way that one experiment does not disturb the result of the other. There are many examples where such disturbance takes place, or where even one experiment may preclude the other: Biopsy of a possible beginning tumor may make the evaluation of a simple medical treatment difficult; a psychiatric patient may be treated by psychopharmica or by classical psychoanalysis, but the evaluation of both approaches on the same patient may be impossible. In other settings, say factorial experiments, similar phenomena occur: An industrial experiment with a fixed set of units may be performed with one given set of factors or another set, not orthogonal to the first one, but including both sets will lead to a different experiment. The effect of nitrogen in some fertilizer may be evaluated in a small, fixed experiment with or without potassium present, not both.

We will assume that for any pair a and a' it is always possible to decide whether or not these experiments can be performed simultaneously without disturbing each other. If this is the case, we say that the experiments are *compatible*. Two compatible experiments can always be joined into a compound experiment by taking the Cartesian product: $\mathcal{E}_a \otimes \mathcal{E}_{a'} = (\mathcal{X}_a, \mathcal{F}_a, \{P_\theta^a; \theta \in \Theta_a\}) \otimes (\mathcal{X}_{a'}, \mathcal{F}_{a'}, \{P_\theta^{a'}; \theta \in \Theta_{a'}\})$, similarly for larger sets of experiments. Sometimes the parameter set for the joint experiment can be simplified in such cases.

In fact we will assume that all potential experiments that can be performed on a given set of units depend on a common (multidimensional) parameter ϕ, defined on a space Φ, and connected to the state of the units. Later we will show that it may be natural instead to associate the state concept to a distribution over Φ, but everything that is said below, in particular the ordering of propositions, can be repeated with states as probability measures.

So to keep things reasonably simple, we will keep ϕ fixed in this discussion. (In fact we will do so until Section 10 below.)

Then θ in the experiment $\mathscr{E}_a$ is a function of the common parameter ϕ, say $\theta = \theta_a(\phi)$; it is assumed that the parameter spaces Φ and Θ_a are equipped with σ-algebras and that each function $\theta = \theta_a(\phi)$ is measurable. Furthermore, it is convenient to assume that each parameter is identifiable: For each a and pair (θ, θ'), if $P^a_\theta(E) = P^a_{\theta'}, (E)$ for all $E \in \mathscr{F}_a$, then $\theta = \theta'$.

Since σ-algebras may be coarsened (data reduction), there is a natural partial ordering between the experiments: Say that $\mathscr{E}_{a'} \leq \mathscr{E}_a$ if $\mathscr{X}_{a'} = \mathscr{X}_a$ and $\mathscr{F}_{a'} \subseteq \mathscr{F}_a$. Then the probability models are assumed to be consistent: $\Theta_{a'} \subseteq \Theta_a$ and $P^{a'}_\theta = P^a_\theta\big|_{\mathscr{F}_{a'}}$ for $\theta \in \Theta_{a'}$. We will let $\mathscr{A}_0$ be the extreme set of experiments on a given set of units, so that for $\mathscr{E}_a$ with $a \in \mathscr{A}_0$ there is no $\mathscr{E}_{a'} \neq \mathscr{E}_a$ such that $\mathscr{E}_a \leq \mathscr{E}_{a'}$.

By a *proposition* $\mathscr{P}$ we mean an experiment together with an event from this experiment: $\mathscr{P} = (a, E_a)$, where E_a belongs to the σ-algebra $\mathscr{F}_a$.

A partial ordering of the propositions from the same experiment is first defined as the obvious one: We say that $(a, E_{1a}) \leq (a, E_{2a})$ if $E_{1a} \subseteq E_{2a}$. This ordering will be generalized to some pairs of propositions from different experiments:

Definition 1. We say that $\mathscr{P}_1 = (a_1, E_{1a_1}) \leq \mathscr{P}_2 = (a_2, E_{2a_2})$ iff $P^{a_1}_{\theta_{a_1}(\phi)}(E_{1a_1}) \leq P^{a_2}_{\theta_{a_2}(\phi)}(E_{2a_2})$ for all state parameters ϕ.

In order that $\mathscr{P}_1 \leq \mathscr{P}_2$ and $\mathscr{P}_2 \leq \mathscr{P}_1$ together shall imply $\mathscr{P}_1 = \mathscr{P}_2$, we will here identify $\mathscr{P}_1$ and $\mathscr{P}_2$ if $P^{a_1}_{\theta_{a_1}(\phi)}(E_{1a_1}) = P^{a_2}_{\theta_{a_2}(\phi)}(E_{2a_2})$ for all ϕ. This may lead to somewhat unfortunate situations where unrelated propositions are identified, but mathematically it is convenient. Among other things it is necessary in order that supremums shall be unique.

This definition includes, but may in certain cases also be an extension of, the trivial one when the experiments are the same. Other cases where $\mathscr{P}_1 \leq \mathscr{P}_2$ include (i) $E_{1a_1} = \emptyset$, (ii) $E_{2a_2} = \mathscr{X}_{a_2}$, and (iii) $\mathscr{E}_{a_1}$ is the Cartesian product of $\mathscr{E}_{a_2}$ and another compatible experiment $\mathscr{E}_{a_3}$, so $E_{1a_1} = E_{2a_2} \otimes E_{3a_3}$.

From (i) and (ii) it follows that we have to identify all propositions of the form $(a, \emptyset)$, which will be collected in the single proposition $\mathbf{0}$, and the propositions of the form $(a, \mathscr{X}_a)$, which will be collected in the single proposition $\mathbf{1}$. These will be the infimum and supremum, respectively, of the whole set of propositions.

The orthocomplement of a proposition is defined in a straightforward way from the complement of an event: For $\mathscr{P} = (a, E_a)$ take $\mathscr{P}^\perp = (a, E^c_a)$. It is then clear that $\mathscr{P}^{\perp\perp} = \mathscr{P}$ and that $\mathscr{P}_1 \leq \mathscr{P}_2$ implies $\mathscr{P}_2^\perp \leq \mathscr{P}_1^\perp$.

From the results of the next section it follows that $\mathcal{P} \wedge \mathcal{P}^{\perp} = \mathbf{0}$ and $\mathcal{P} \vee \mathcal{P}^{\perp} = \mathbf{1}$. Thus the properties of an orthocomplementation are satisfied.

Sets of propositions from the same experiments will ordinarily have a natural supremum and infimum, corresponding to the usual unions and intersections. To introduce supremum and infimum for arbitrary sets of propositions, however, one needs more structure. We will in fact add more structure by making symmetry assumptions. But first we will show that the simple structure of sets of experiments under weak extra assumptions implies that the set of propositions always will be an orthocomplemented, orthomodular poset.

5. ORTHOMODULAR PROPOSITIONS FROM EXPERIMENTS

We start with an assumption which seems rather weak in the general setting we are considering, but also as it stands seems difficult to motivate directly from statistical reasoning. It is closely related to Axiom V in Mackey (1963), and has been formulated again in several papers in quantum logic. Mączyński (1973) showed that this assumption is necessary and sufficient in order that an orthocomplemented poset shall possess some natural properties.

We say that k propositions $\mathcal{P}_1, \ldots, \mathcal{P}_k$ with $\mathcal{P}_i = (a_i, E_i)$ are *orthogonal* if the inequality

$$\sum_{i=1}^{k} P_{\theta_i(\phi)}^{a_i}(E_i) \leq 1 \tag{1}$$

holds for all ϕ. We say that these propositions are *pairwise orthogonal* if the same inequality holds for any pair chosen from the k propositions. This is equivalent to $\mathcal{P}_i \leq \mathcal{P}_j^{\perp}$ for all pairs.

It is clear that orthogonality implies pairwise orthogonality. The opposite implication holds for most other orthogonality concepts, but here statistical examples can easily be constructed for which this does not hold. Such cases will be explicitly excluded from our sets of propositions (strictly speaking, we only need this assumption for $k = 3$ here).

Assumption 1. For the propositions under consideration, any set of k pairwise orthogonal propositions is orthogonal.

For events E_i from the same experiment, pairwise orthogonality essentially means $E_i \cap E_j = \emptyset$ for $i \neq j$. This obviously implies orthogonality in the sense given by (1). For general sets of propositions, Assumption 1 amounts in some sense to assuming a certain richness of the set of models.

For the intermediate lemma below, we may also need to extend the set of experiments under consideration. This can always be done artificially, since there are no limitations on the set of experiments under consideration,

in particular, the set of experiments need not be countable. We will see later how the artificial experiments can be deleted when final results are formulated.

Assumption 2. Let $\mathcal{P}_i = (a_i, E_i)$ $(i = 1, \ldots, k)$ be k propositions such that (1) holds for all ϕ. Then there is an experiment $\mathcal{E}_{a'} = (\mathcal{X}_{a'}, \mathcal{F}_a, \{P^{a'}_\theta; \theta \in \Theta_{a'}\})$ and k pairwise disjoint events E'_i in $\mathcal{F}_{a'}$ such that $P^{a'}_{\theta'(\phi)}(E'_i) = P^{a_i}_{\theta_i(\phi)}(E_i)$ for all $\phi \in \Phi$ and $i = 1, \ldots, k$.

In the proof of the lemma below, we will only make use of Assumption 2 for $k = 3$, to begin with only for $k = 2$.

Lemma 1. Let $\mathcal{P}_1 = (a_1, E_1)$ and $\mathcal{P}_2 = (a_2, E_2)$ be two orthogonal propositions in the above sense, i.e.,

$$P^{a_1}_{\theta_1(\phi)}(E_1) + P^{a_2}_{\theta_2(\phi)}(E_2) \leq 1, \qquad \forall \phi$$

Then under Assumptions 1 and 2 there exists a proposition $\mathcal{P} = (a, E)$ on some experiment $\mathcal{E}_a$ such that

$$P^a_{\theta(\phi)}(E) = P^{a_1}_{\theta_1(\phi)}(E_1) + P^{a_2}_{\theta_2(\phi)}(E_2) \tag{2}$$

and we have that $\mathcal{P} = \mathcal{P}_1 \vee \mathcal{P}_2$ for this choice of $\mathcal{P}$.

Proof. The existence of $\mathcal{P}$ follows by first using Assumption 2. This gives $E'_1 \cap E'_2 = \emptyset$ for events E'_1 and E'_2 in some single experiment. We can then take $E = E'_1 \cup E'_2$, so that (2) follows.

It is clear that $\mathcal{P} \geq \mathcal{P}_1$ and $\mathcal{P} \geq \mathcal{P}_2$ for this choice of $\mathcal{P}$. Assume that $\mathcal{P}_0$ is another proposition such that $\mathcal{P}_0 \geq \mathcal{P}_1$ and $\mathcal{P}_0 \geq \mathcal{P}_2$. Then by Assumption 1 and Assumption 2 with $k = 3$ there exists an experiment $\mathcal{E}_{a'}$, and events E'_0, E''_2, and E''_2 in $\mathcal{F}_{a'}$ such that $E'_1 \subseteq E'_0$, $E''_2 \subseteq E'_0$, and $E'_1 \cap E''_2 = \emptyset$. But then it follows that $E''_1 \cup E''_2 \subseteq E'_0$, $P^{a_0}_{\theta_0(\phi)}(E_0) \geq P^{a_1}_{\theta_1(\phi)}(E_1) + P^{a_2}_{\theta_2(\phi)}(E_2)$ for all ϕ, so $\mathcal{P} \leq \mathcal{P}_0$. Hence $\mathcal{P} = \mathcal{P}_1 \vee \mathcal{P}_2$.

The basic property of orthomodularity was defined in Section 3 for lattices. This definition requires only a definition of supremum for pairs of *orthogonal* propositions. Here is the necessary reformulation: Say that a set of propositions is orthomodular if $\mathcal{P}_1 \leq \mathcal{P}_2$ implies that $\mathcal{P}_2 = \mathcal{P}_1 \vee (\mathcal{P}_2^{\perp} \vee \mathcal{P}_1)^{\perp}$. The following result must be seen in relation to Lemma 1. Note that neither Assumption 1 nor Assumption 2 is needed explicitly in the theorem.

Theorem 1. Assume an orthocomplemented poset of propositions based on experiments with the property that pairwise suprema exist and satisfy equation (2) for the pairwise supremum $\mathcal{P} = (a, E)$. Then this poset will be orthomodular.

Proof. Let $\mathcal{P}_i = (a_i, E_i, (i = 1, 2)$ satisfy $\mathcal{P}_1 \leq \mathcal{P}_2$. Then by assumption, $\mathcal{P}_2^{\perp} \vee \mathcal{P}_1$ is equal to $\mathcal{P} = (a, E)$ such that $P^a_{\theta(\phi)}(E) = 1 - P^2_{\theta_2(\phi)}(E_2) + P^1_{\theta_1(\phi)}(E_1)$. By the same assumption $\mathcal{P}_1 \vee \mathcal{P}^{\perp}$ is equal to $\mathcal{P}' = (a', E')$ such that $P^{a'}_{\theta'(\phi)}(E') = P^1_{\theta_1(\phi)}(E_1) + 1 - P^a_{\theta(\phi)}(\mathrm{E}) = P^2_{\theta_2(\phi)}(E_2)$. Hence $\mathcal{P}' = \mathcal{P}_2$, and the orthomodularity property follows.

The concept of an orthocomplemented, orthomodular poset has been central to the lattice approach to quantum mechanics. As stated before, to arrive at a structure that has a Hilbert space representation, further assumptions are needed. We will study the consequences of assuming that there exists a symmetry group connected to the state of a set of experiments.

6. SYMMETRY AND PERMISSIBILITY

We assume that each sample space $\mathcal{X}_a$ is a locally compact topological space, and let $\mathcal{X} = \{(a, x)\colon a \in \mathcal{A}, x \in \mathcal{X}_a\}$ be the collection of all sample spaces. If $\mathcal{X}$ is given the topology composed of unions of sets of the form (a, V_a), where V_a is open in $\mathcal{X}_a$, then $\mathcal{X}$ will also be locally compact.

Now let G be a group of transformations on $\mathcal{X}$. For single experiments it is known (Helland, 1998, and references there) that—if simple consistency properties are satisfied—such a group may both simplify the statistical analysis and allow a considerable strengthening of conclusions. We will here discuss consequences for sets of potential experiments of the assumption of the existence of a symmetry group.

The groups connected to single experiments may be looked upon as subgroups of G: Let $G_a = \{g \in G\colon ga = a\}$. Then, by a slight misuse of notation, G_a will be considered to be a group of transformations of $\mathcal{X}_a$.

Now introduce models, i.e., look at the whole specification of the experiment $\mathcal{E}_a = (\mathcal{X}_a, \mathcal{F}_a, \{P^a_\theta;\ \theta \in \Theta_a\})$. As is common in statistical models under symmetry, it will be assumed that $\mathcal{F}_a$ is closed under the action of the elements of G_a, and that G_a is given a topology such that the mappings $g \to g^{-1}$, $(g_1, g_2) \to g_1 g_2$, and $(g, x) \to gx$ are continuous.

A group $\bar{G}_a$ of transformations of the parameter space Θ_a is introduced in the natural way by

$$P^a_{\bar{g}_a\theta}(E) = P^a_\theta(g_a^{-1} E) \qquad \text{for} \quad E \in \mathcal{F}_a$$

A basic assumption is that the model is closed under the transformations $\bar{g}_a \in \bar{G}_a$. Another assumption is that in each model the parameter can be identified:

$$P^a_{\theta'}(E) = P^a_\theta(E) \quad \text{for all} \quad E \in \mathcal{F}_a \quad \text{implies} \quad \theta' = \theta \tag{3}$$

As earlier, it is assumed that all parameters $\theta = \theta_a$ in Θ_a are functions of a single parameter ϕ from some space Φ. This is the parameter characteriz-

ing the state of the system, and it is natural to assume that all parameter transformations are generated by transformations of ϕ.

Assumption 3. There is a group $\bar{G}_a$ of transformations on Φ such that the elements $\tilde{g}_a$ of $\tilde{G}_a$ all have the form

$$\tilde{g}_a(\theta_a(\phi)) = \theta_a(\bar{g}_a\phi) \tag{4}$$

The special parametric functions $\theta(\cdot) = \theta_a(\cdot)$ that satisfy a relation of the form (4) for some groups $\bar{G}$ and $\bar{G}_a$ played an important role in Helland (1998), where they were called invariantly estimable functions. They will be even more important here, and they will be referred to several times, also when no estimation is involved, so we will simply call these parametric functions *permissible* . The following results summarize some of their main properties:

Lemma 2. (a) For the group $\bar{G}_a$, the parametric function $\theta_a(\cdot)$ is permissible, i.e., there is a set of transformations $\{\bar{g}_a\}$ such that (4) holds if and only if for each pair (ϕ', ϕ)

$$\theta_a(\phi') = \theta_a(\phi) \quad \text{implies} \quad \theta_a(\bar{g}\phi') = \theta_a(\bar{g}\phi) \quad \text{for all} \quad \bar{g} \in \bar{G}. \tag{5}$$

(b) The set of transformations $\tilde{g}_a$ described by the relation (4) will necessarily constitute a group, and this group is the homomorphic image of the group $\bar{G}_a$: If $\bar{g},\bar{g}' \to \tilde{g},\tilde{g}'$, then $\bar{g}\bar{g}' \to \tilde{g}\tilde{g}'$ and $\bar{g}^{-1} \to \tilde{g}^{-1}$.

(c) Assume that $\theta_a(\cdot)$ is permissible relative to $\bar{G}_a$ as above, and assume that in experiment $\mathcal{E}_a$ we have that $\eta(\cdot)$ is a permissible function of θ_a relative to $\tilde{G}_a$. Then $\zeta(\cdot)$ defined by $\zeta(\phi) = \eta(\theta_a(\phi))$ is permissible.

Proof. (a) The general implication (5) is equivalent to the requirement that $\theta_a(\bar{g}\phi)$ is a function of $\theta_a(\phi)$.

(b) Straightforward verification.

(c) We have $\zeta(\bar{g}\phi) = \eta(\bar{g}_a(\theta_a(\phi))) = \tilde{g}_a(\eta(\theta_a(\phi))) = \tilde{g}_a(\zeta(\phi))$ for some $\tilde{g}_a$.

Certain inconsistencies in Bayesian estimation theory are avoided if one concentrates on permissible parametric functions. Under weak additional assumptions on the loss function we also have that the best invariant estimator will be equal to the Bayes estimator under noninformative prior for such parameters (Helland, 1998).

Another homomorphism in the structure described above is $g_a \to \bar{g}_a(G_a \to \bar{G}_a)$. It follows that $\bar{G}_a$ is the homomorphic image both of $\bar{G}$ and of G_a, which is a subgroup of G.

The following properties of permissible parametric functions are borrowed from Helland (1998), but all points are easy to verify:

(i) The full parametric function $\phi \to \phi$ is permissible.

(ii) If η is invariant in the sense that $\eta(\bar{g}\phi) = \eta(\phi)$ for all $\bar{g} \in \bar{G}$ and all $\phi \in \Phi$, then η is permissible. If $\theta = \theta_a(\cdot)$ corresponding to an experiment $\mathscr{E}_a$ is of this type, then θ is constant (on orbits of $\bar{G}$).

(iii) If $\eta(\phi)$ is permissible with range M and γ is a 1–1 function from M onto another space N, then ζ given by $\zeta(\phi) = \gamma(\eta(\phi))$ is permissible. This is a special case of Lemma 2(c).

(iv) If $\{\eta_i;\ i \in I\}$ is any set of permissible parametric functions, then θ given by $\theta(\phi) = (\eta_i(\phi);\ i \in I)$ is permissible.

Note that the functions $\phi \to P_{\theta(\phi)}(E)$ are not necessarily permissible even though the functions θ are. This implication does hold, however, if the parameter $\theta(\cdot)$ is one-dimensional and P depends monotonically on it, since then the relation between θ and P will be 1–1.

7. GROUP REPRESENTATION RELATED TO THE PARAMETER SPACE

One very important aspect of locally compact groups of transformations is that one can define left- and right-invariant measures (Nachbin, 1965): $\mu(\bar{g}D) = \mu(D)$ and $\nu(D\bar{g}) = \nu(D)$, where $\bar{g} \in \bar{G}$ and $D \subseteq \bar{G}$. If Φ is also locally compact, and if the action of $\bar{G}$ on Φ satisfies a weak extra condition, then right Haar measure can also be defined on Φ in a consistent way (Helland, 1998, and references there). It is argued in Helland (1998) that from many points of view this right-invariant measure is the correct one to use as a prior 'distribution' when $\bar{G}$ expresses the symmetry of the problem and no other information is present.

Linear representation of groups has played an increasing role in quantum mechanical calculations in the last decades, and in fact much of the motivation behind the recent development of group representation theory as a mathematical discipline has been taken from quantum theory. Nevertheless, when it comes to the physical and mathematical foundation of quantum theory, little use has been made of group representations. An exception is Bohr and Ulfbeck (1995), where physical aspects are emphasized.

We will concentrate here on the group $\bar{G}$ of transformations on the basic parameter space Φ. Assume that Φ is endowed with a σ-algebra with a separability property (Dunford and Schwartz, 1958, p. 169) so that the space $\mathscr{H} = L^2(\Phi,\nu)$ of complex square-integrable functions on Φ is separable. We take the measure ν as right Haar measure.

The elements $\bar{g} \in \bar{G}$ generate unitary transformations on $\mathscr{H}$ by

$$U(\bar{g})f(\phi) = f(\bar{g}^{-1}\phi) \tag{6}$$

Note that these transformations form a group which is the homomorphic image of the group $\bar{G}$. A major issue in group representation theory is to

study invariant subspaces under such transformations, in particular to look for irreducible invariant subspaces.

Now recall the permissible parametric functions $\theta(\cdot)$ defined on the same space Φ. There is a natural ordering of these functions: *Say that* $\theta'(\cdot) \preccurlyeq \theta(\cdot)$ *if* $\theta'(\cdot)$ *is a function of* $\theta(\cdot)$, *i.e.*, $\theta'(\phi) = \psi(\theta(\phi))$ *for some* ψ. In classes of statistical models/experiments it may be of some interest to find minima under this ordering for certain sets of permissible functions. We will see below that equivalence classes of permissible functions are in one-to-one correspondence with subspaces of $\mathcal{H}$ that are invariant under the group-generated class of unitary transformations (6), and that the ordering above corresponds to the natural ordering of subspaces. This may explain to a certain extent why group representation theory is so important in quantum mechanics.

Theorem 2. (a) Consider a fixed permissible function $\theta(\cdot)$. The set of functions f of the form $f(\phi) = f_0(\theta(\phi))$ constitutes a closed linear subspace of $\mathcal{H}$ which is invariant under the transformations $U(\bar{g})$.

(b) If θ_i is a permissible function and $\theta_1 \preccurlyeq \theta$, then the subspace corresponding to θ_1 is contained in the subspace corresponding to θ. Conversely, if V_1 corresponds to θ_1, V to θ, and V_1 is a subspace of V, then $\theta_1(\cdot) \preccurlyeq \theta(\cdot)$.

(c) Say that two permissible parametric functions $\theta(\cdot)$ and $\theta'(\cdot)$ are equivalent if they are 1–1 functions of each other. Then the set of equivalence classes of permissible functions is in 1–1 correspondence with the subset of invariant subspaces of $\mathcal{H}$ described in (a).

Proof. (a) It is clear that the space is closed under linear combinations, and also under infinite sums that converge in L^2-norm, so the space is closed. If f belongs to the subspace, then

$$U(\bar{g})f(\phi) = f(\bar{g}^{-1}\,\phi) = f_0(\tilde{g}^{-1}(\theta(\phi)))$$

is also in the subspace.

(b) Obvious.

(c) From (b) the invariant subsets of equivalent permissible functions must be contained in each other.

For the use of group representation in physics, see, for instance, Hamermesh (1962); a reference to the more general theory is Barut and Rączka (1977).

In later developments we will need further properties of the Hilbert space $\mathcal{H}$ and those closed subspaces V of $\mathcal{H}$ that are defined by permissible parametric functions. Here is the kind of result that is needed.

Lemma 3. (a) Under weak regularity conditions the parameter group $\bar{G}_a$ of experiment $\mathcal{E}_a$ will be locally compact and have a right Haar measure ν_a.

Let $f(\cdot)$ be a given function on Φ. Then there is a unique (almost everywhere with respect to ν_a) function $f_a(\theta)$ such that for all functions $c(\cdot)$ of θ we have

$$\int c(\theta_a(\phi))f(\phi)\ \nu(d\phi) = \int c(\theta)f_a(\theta)\ \nu_a(d\theta) \tag{7}$$

(b) Let $V_1 \subset V_a \subset \mathcal{H}$, where V_1 corresponds to the permissible function $\theta_1(\cdot)$ and V_a to $\theta_a(\cdot)$. Let f_1 and f_a be the corresponding functions defined in (a). Then, for all $c_1(\cdot)$

$$\int c_1(\theta_1)f_1(\theta_1)\nu_1(d\theta_1) = \int c_1(\theta_1(\theta))f_a(\theta)\nu_a(d\theta) \tag{8}$$

(c) For all functions g on Θ_a we have

$$\int \left|f(\phi) - f_a(\theta_a(\phi))\right|^2 \nu(d\phi) \leq \int \left|f(\phi) - g(\theta_a(\phi))\right|^2 \nu(d\phi) \tag{9}$$

Equality holds here if and only if $g(\theta) = f_a(\theta)$ almost everywhere with respect to the measure ν_a.

Proof. (a) We know that $\bar{G}$ is locally compact, and that $\bar{G}_a$ is the homomorphic image of $\bar{G}$. Assuming that the function describing the homomorphism is continuous, $\bar{G}_a$ will inherit the topology from $\bar{G}$, and then be locally compact. The measure $f(\phi)\nu(d\phi)$ will be absolutely continuous with respect to $\nu_a(d\theta)$; let $f_a(\theta)$ be the Radon-Nikodym derivative.

(b) This follows from well-known properties of Radon-Nikodym derivatives.

(c) By using (7) on $c(\theta) = g(\theta) - f_a(\theta)$, we find

$$\int \left|f(\phi) - g(\theta_a(\phi))\right|^2 \nu(d\phi) - \int \left|f(\phi) - f_a(\theta_a(\phi))\right|^2 \nu(d\phi)$$
$$= \int \left|g(\theta) - f_a(\theta)\right|^2 \nu_a(d\theta)$$

Equations (7) and (9) show that the function $f_a(\theta(\phi))$ may be regarded as the projection of $f(\phi)$ on the space determined by the permissible function $\theta(\cdot)$, and (8) shows that this projection functions as it should under iteration. This appears to be the start of a state vector approach to quantum mechanics based on a Hilbert space representation of $\bar{G}$. We will come back to this approach later, but first we will return to the lattice approach, which leads to a (complementary) Hilbert space representation of propositions.

8. LATTICE PROPERTY AND HILBERT SPACE

From Assumption 3 it is clear that the parameters θ connected to each single experiment are permissible functions of the basic (hyper-) parameter ϕ. An important observation was made in Lemma 2(c): Further parametric functions that are permissible relative to this single experiment are also permissible relative to ϕ. Often it is natural to design subexperiments to estimate such parameters.

The first part of the assumption below can be motivated as follows: If $a(\cdot)$ is a function of a set of parameters and $b(\cdot)$ is an extension of a using more parameters, then usually the values b will be smaller than the values a for some parameter values, but in most cases there are also parameter values for which b is larger than a.

We let the vector space representation of permissible parametric functions from the previous section be understood.

Assumption 4. (a) The supremum of a set of propositions $\{\mathcal{P}_i\}$, if it exists, will be some proposition $\mathcal{P} = (a, E)$ with probability $P^a_{\theta(\phi)}(E)$ where the parameter θ determines a space (Theorem 2) contained in the space spanned by those corresponding to the parameters θ_i in each $\mathcal{P}_i$.

(b) Corresponding to every such parametric function θ there is an experiment $\mathcal{E}$.

Our main aim in this section is to show that the supremum of any set of propositions can be defined.

Theorem 3. Let $\{\mathcal{P}_i = (a_i, E_i); i \in I\}$ be a set of propositions, partially ordered under the ordering given in Definition 1. Let Assumptions 3 and 4 hold. Then the supremum $\mathcal{P}$ of this set exists, in the sense that $\mathcal{P}_i \leq \mathcal{P}$ for all i and $\mathcal{P}_i \leq \mathcal{P}_0$, $\forall i$, implies $\mathcal{P} \leq \mathcal{P}_0$.

Proof. Assume a set of propositions $\mathcal{P}_i = (a_i, E_i)$ with probabilities $P^{a_i}_{\theta_i(\phi)}(E_i)$. What we are after, is a proposition $\mathcal{P} = (a, E)$ such that

$$P^{a_i}_{\theta_i(\phi)}(E_i) \leq P^a_{\theta(\phi)}(E) \tag{10}$$

for all i and for all ϕ, and such that $\mathcal{P}$ is the minimal proposition satisfying this. By Assumption 4(a) we can restrict our search to propositions $\mathcal{P}$ with parametric function θ associated with a definite linear space V, the one spanned by the linear spaces of the parameters θ_i.

Now (10) for each ϕ is equivalent to

$$\sup_i P^{a_i}_{\psi_i(\theta)}(E_i) \leq P^a_\theta(E) \tag{11}$$

where $\psi_i(\theta(\phi)) = \theta_i(\phi)$, which makes sense by the nesting of the linear spaces and by Theorem 2 (b).

By Assumption 4(b) there exists at least one experiment with parameter θ. Pick one such experiment. Then we can choose the E which minimizes the right-hand side of (11) under this constraint for all θ's. Propositions that give the same solution are identified by earlier assumptions. It is then clear that the supremum found is unique.

As a consequence of Theorems 1 and 3 we now have that, with the assumptions made, the set of propositions, constructed in a natural way from sets of potential experiments subject to symmetry conditions, form an orthomodular, orthocomplemented lattice. This is the basic entity of the quantum logic approach to quantum mechanics. We feel that most of the assumptions made are relatively innocent; they may perhaps be improved slightly in detail, but they have a more concrete interpretation than axioms for quantum mechanics in most existing approaches.

The remaining conditions from Section 3, atomicity, covering property, and separability, are more technical, and some of them have been controversial in some of the quantum logic literature. From a statistical point of view it is obvious that each sample space is the union of its atoms, but the corresponding assumption is more problematic if events with the same probability for each θ are identified, as we have chosen to do here. We also have difficulties with the covering properties in models where atoms are undefined. On the other hand, with the identification of events, it is usually not problematic to assume that unions of disjoint events in each sample space are at most countable, and then by Assumption 2, the separability property follows.

A simple solution is to assume that all experiments are discrete, that is, that each sample space $\mathscr{X}_a$ is countable. Then all conditions are satisfied, and by arguments in Piron (1976), Maeda and Maeda (1970), Varadarajan (1985), and other places the ordinary Hilbert space model of quantum mechanics follows.

Heuristically, continuous sample spaces may be approximated by discrete sample spaces. In fact, the situation at this point may be seen as a reflection of the situation in ordinary quantum mechanics, where it is well known that precise treatment of continuous variables requires other concepts than the ordinary Hilbert space approach, say based on C^*-algebras. Note that our own underlying framework with sets of models for experiments subject to symmetry is conceptually simple in the continuous case, too.

From the discussion in Beltrametti and Cassinelli (1981) we deduce the following result.

Theorem 4. In the case where all experiments $\mathscr{E}_a$ are discrete, and the assumptions above hold, there is a complex, separable Hilbert space $\mathscr{H}_0$ such that (assuming that the dimension of $\mathscr{H}_0 \geq 3$) each proposition $\mathscr{P} = (a, E)$

can be associated uniquely with a projection operator $\Pi_{a,E}$ in $\mathcal{H}_0$ in the sense that

$$P^a_{\theta(\phi)}(E) = \text{trace}(\rho\Pi_{a,E}) \tag{12}$$

where $\rho = \rho(\phi)$ is a density operator, a nonnegative operator of trace 1.

Here we have combined the theorem on existence of the Hilbert space with the famous Gleason theorem, which states that if the dimension of the Hilbert space is 3 or larger, every probability distribution over propositions can be computed in this way. Gleason's theorem is very cumbersome to prove (Varadarajan, 1985). We will indicate later that the state vector representation of Section 7 probably may be used to give a simpler proof of this result.

Corollary 1. The conclusion of Theorem 4 holds under only the Assumptions 1, 3, and 4.

Proof. As it stands, Theorem 4 is deduced using all Assumptions 1–4. As remarked earlier, however, it is always possible to extend the set of propositions so that Assumption 2 holds. This extended set of propositions does not occur in the statement (12).

All that has been done up to now could also have been done for a simple case of a distributive lattice. Then as in Theorem 4 we would still have been given a Hilbert space, but this Hilbert space would have been trivial. Hence, a further question to ask is if the quantum logic derived from sets of experiments is nondistributive in general. In some sense this is obvious, since if it were distributive, we would have an ordinary Boolean algebra, which by a well-known theorem by Stone would imply that everything could be represented as one single experiment. Here is a simple example of nondistributivity. More complicated examples can be constructed from this.

Example 1. Look at two experiments $\mathcal{E}_a = (\mathcal{X}, \mathcal{F}, \{P_\theta\})$ and $\mathcal{E}_b = (\mathcal{Y}, \mathcal{G}, \{\mathcal{Q}_\psi\})$. Let A, B $\in \mathcal{F}$, $A, B \neq \mathcal{X}$, but $A \cup B = \mathcal{X}$, and let $C \in \mathcal{G}$, $C \neq \emptyset$. Assume that the two experiments are unrelated in the sense that θ and ψ depend on the common parameter ϕ in disjoint parts of its range $\mathbf{\Phi}$. Also assume that $P_\theta(A) = P_\theta(\mathrm{B}) = \mathcal{Q}_\psi(C) = 0$ in the areas where these are independent of the parameters. Then $(a, A^c) \vee (b, C^c)$ is some proposition (e, D) whose probability for all ϕ should dominate $\max(P_{\theta(\phi)}(A^c), Q_{\psi(\phi)}(C^c))$. But, from the assumptions made, this maximum must be 1 for all ϕ. Hence $(a, A^c) \vee (b, C^c) = \mathbf{1}$, so $(a, A) \wedge (b, C) = \mathbf{0}$. Similarly, $(a, B) \wedge (b, C) = \mathbf{0}$, so $((a, A) \wedge (b, C)) \vee ((a, B) \wedge (b, C)) = \mathbf{0}$. On the other hand, $((a, A) \vee (a, B)) \wedge (b, C) = (a, A \cup B) \wedge (b, C) = \mathbf{1} \wedge (b, C) = (b, C)$. So these three events do not satisfy the distributive law.

Finally, one can ask if the quantum logic derived from sets of potential experiments is wide enough to cover everything that is of interest in quantum mechanics. Of course, it is very ambitious to try to give an answer to such a question, but it is an encouraging thought that virtually every attempt that can be made to verify any theory on the quantum level has to be based on experiments. Hence it seems very difficult to transcend beyond this frame. (However, this argument perhaps does not hold for quantum cosmology.)

9. OBSERVABLES

In Section 7 we introduced a group representation of the transformation group $\bar{G}$ in the parameter space, and looked at some concrete interpretations of that representation in Theorem 2. The Hilbert space of Theorem 4 can in some sense be regarded as the corresponding representation of the sample space group G. Even though the consequences of this latter representation are discussed in all books in quantum mechanics, we will give some remarks on it here.

Since we have assumed that all sample spaces $\mathcal{X}_a$ are discrete (this is not necessary for our basic model, but convenient for the Hilbert space representation), we might as well assume that $\mathcal{X}_a = \{1, 2, \ldots\}$ for each a. The σ-algebra is then the obvious one, and there is a parametric class of probabilities $\{P^a_{\theta(\phi)}(x);\ \theta \in \Theta_a\}$, adding to 1 for each a. As elsewhere it is assumed that the parameter in each experiment is a function of a state parameter $\phi \in \Phi$. The most important assumptions that we have made, apart from the assumption that the class of experiments should be rich enough, is that pairwise orthogonality should imply total orthogonality in the sense of (1) and that there exists a group structure on Φ with the consistency requirement $\theta_a(\bar{g}\phi) = \bar{g}_a(\theta_a(\phi))$ for some $\bar{g}_a$. We will also assume here that $\bar{G}$ is transitive on Φ; the case with several orbits corresponds to superselection rules: Index the orbits with some parameter τ. Then τ is conserved during all symmetry transformations. Using a reasonable theory of time development, it can also be shown to be conserved over time. Physical variables with this property might be charge, mass, or hypercharge. In the further discussion there is nothing lost by keeping τ fixed, which is the same as sticking to one particular orbit, so that $\bar{G}$ is transitive.

The simplest quantum mechanical interpretation of (12) is now that each primitive event (a, x) can be represented in a fixed complex separable Hilbert space $\mathcal{H}_0$ by a one-dimensional projection: For each a we may take a set of orthonormal vectors $\{e_{a,x}\}$ so that the event (a, x) is represented by the projection $e_{a,x}\, e^\dagger_{a,x}$. In concrete terms this means that according to Gleason's theorem there exists a density operator ρ such that $P^a_{\theta(\phi)}(x) = e^\dagger_{a,x}\, \rho e_{a,x}$. It is not difficult to see that it is always possible to find one such ρ for each

experiment; the strong part of this result is that the same ρ can be chosen for all the experiments.

In each experiment one can of course introduce random variables in the usual statistical sense: $Y(\cdot)$, $Z(\cdot)$, . . .are measurable functions on some $\mathcal{X}_a$, and the distributions of these are determined in the usual way from $\{P^q_{\theta(\phi)}(x)\}$ above. Again, the main thing that is new in quantum theory is that there is some connection between variables defined in different experiments, and this connection may have rather large consequences. It may be natural, since probabilities are summarized through projection operators, to associate random variables originally defined in an experiment $\mathcal{E}_a$ with operators, too: If $Y(x) = y_x$ for $x \in X_a$, then take

$$\hat{Y} = \sum_x y_x e_{a,x} e^{\dagger}_{a,x} \tag{13}$$

The implication from Gleason's theorem is then that in any state ρ the expectation of Y will be

$$\langle Y \rangle = \text{trace}(\hat{Y}\rho) \tag{14}$$

Again take $\rho = ee^{\dagger}$ for some unit vector e. Then an easy standard calculation shows that the variance $\langle (Y - \langle Y \rangle)^2 \rangle$ of Y vanishes if and only if e is an eigenvector of $\hat{Y}$ with some eigenvalue λ. Well-known results from quantum theory follow.

With a more general sample space, the random variable Y is a measurable function from $(\mathcal{X}_a, \mathcal{F}_a)$to the Borel sets $\mathcal{B}$ on the real line, meaning that $E_a = Y^{-1}(B)$ belongs to $\mathcal{F}_a$ whenever $B \in \mathcal{B}$. In quantum logic, an observable is defined more generally as a measurable map from $\mathcal{B}$ to the set of propositions $\{(a, E_a)\colon a \in \mathcal{A}, E_a, \in \mathcal{F}_a\}$. This is a formal way of generalizing the notion of random variables in such a way that it makes sense for several experiments, a conceptual idea that may be of interest in ordinary statistics, too. Some technical problems associated with this notion of observables are discussed in Gudder (1978).

10. STATES

It is now time to leave the habit of only associating the concept of state with fixed values of the hyperparameter ϕ. Two fundamental observations lead to this:

1. So far, nothing has been said about the state space Φ, except that there should be defined a group $\bar{G}$ on it. From that point of view, there is nothing to prevent us from replacing Φ by a larger space Φ', constructed such that each ϕ is a function on Φ', in the earlier notation: $\phi \preccurlyeq \phi'$. Then

little in what has been said is changed if each ϕ is replaced by any parameter value that is mapped upon ϕ, or more generally, by a prior measure on these ϕ'.

2. Since the density operator ρ in (12) is a weighted average of pure state density operators, it is reasonable to have the left-hand side also as a measure over "pure states" in some sense expressed by ϕ. It is this argument that we now will try to make more precise.

Again, since in the derivation of (12) we assumed that each $\mathscr{X}_a$ is discrete. we might as well replace E by a singleton $\{x\}$. In Section 9 we used $\Pi_{a,x} = e_{a,x}\, e_{a,x}^{\dagger}$ but this is not needed here. We will assume that x is nontrivial in the sense that $P^{a}_{\theta_a(\phi)}(x) > 0$ for at least one ϕ.

If we now replace ϕ by an average according to some prior probability measure $\tau_\rho(\cdot)$ on Φ, the equation reads

$$\int P^{a}_{\theta_a(\phi)}(x)\, \tau_\rho\, (d\phi) = \operatorname{trace}(\rho \Pi_{a,x}) \tag{15}$$

Look first at the case where $\rho = e_i\, e_i^{\dagger}$ is a pure state. Then (15) is

$$\int P^{a}_{\theta_a(\phi)}(x)\, \tau_i(d\phi) = e_i^{\dagger} \Pi_{a,x} e_i \tag{16}$$

Since it is clear from (15) and (16) that $\tau_\rho(\cdot) = \int \lambda(di)\, \tau_i(\cdot)$ when $\rho = \int \lambda(di)\, e_i e_i^{\dagger}$, we may concentrate on (16).

Now turn to the arbitrariness in the choice of Φ as remarked in point 1 above. A natural requirement is that Φ should be chosen in some minimal way, subject to the requirement that it should serve as a hyperparameter space for all the experiments in question. In fact the concept of minimality can be made completely precise we have the ordering $\preccurlyeq$ of parametric functions.

We will for simplicity assume that each τ_i is absolutely continuous with respect to the right G-Haar measure $\nu(\cdot)$ on Φ.

Theorem 5. (a) There is (under technical assumptions) a unique minimal hyperparametric space Φ so that (16) holds for each a, x, and i.

(b) For each i such that (16) is positive for at least one (a, x) there is a unique hyperparameter ϕ_i such that the measure τ_i is concentrated on ϕ_i. The correspondence between the pure state vectors e_i and these hyperparameters ϕ_i is given by

$$P^{a}_{\theta_a(\phi_i)}(x) = e_i^{\dagger} \Pi_{a,x} e_i \tag{17}$$

Proof. (a) Start with one Φ', and with respect to this hyperparameter space consider the Hilbert space $\mathscr{H}' = L^2(\Phi', \nu')$ as discussed in Section 7. When $\tau_i'(\cdot)$ is absolutely continuous with respect to Haar measure, the left-hand side of (16) can be written

$$\int P^{a}_{\theta'_a(\phi')}(x) f'_i(\phi')\, \nu'(d\phi') \tag{18}$$

for some function f'_i. Using the projection defined in Lemma 3, each such expression can be projected onto a space V_a corresponding to the parametric function θ'_a. Take $\mathcal{H}$ as the minimal linear space spanned by $\{V_a: a \in \mathcal{A}\}$. We will assume regularity conditions [cf. Theorem 2(d)] such that $\mathcal{H}$ corresponds to a unique (hyper-) parametric function ϕ (ϕ'). This means, again using Lemma 3, that all primes can be removed from the expression (18). When $\tau'_i(\cdot)$ is not absolutely continuous, a limiting argument must be used.

(b) Look at the left-hand side of equation (16) when Φ is minimal, and fix i. Assume that $\tau_i(\cdot)$ cannot be chosen as Dirac measure. Then, by projecting upon all the spaces corresponding to the $\theta_a(\cdot)$, there is a nontrivial measure left which does not concern any of the parameters, contradicting the assumption that Φ is minimal. Thus $\tau'_i(\cdot)$ has to be Dirac on some value ϕ_i. This means that (17) holds.

The result of Theorem 5 establishes a connection between the two Hilbert spaces of this paper, and is in some sense close to giving an independent proof of Gleason's theorem. And it also gives a link to ordinary statistical models.

Suppose that an ideal measurement has been done. This means that some experiment $\mathcal{E}_a$ has been performed, and this has been so accurate that we afterward can assume that the corresponding parameter is exactly determined: $\theta_a(\phi) = \theta_0$. From a statistical point of view it is clear that such a measurement must have consequences for what is predicted in future measurement, and hence also on the state.

Assume that the state before measurement is given by a density matrix ρ, hence a prior measure $\tau_\rho(\cdot)$, so that the result x in experiment $\mathcal{E}_a$ is given by both sides of (15). After the measurement is done, the new state—i.e., the new measure—should be restricted to $\Phi_a = \{\phi : \theta_a(\phi) = \theta_0\}$. By suitably parametrizing Φ_a (we can find a permissible parametrization since θ_a is permissible) this again amounts to a projection in the sense of Lemma 3. The well-known Gleason theorem solution emerges, namely to replace ρ by $\pi\rho\pi$, normalized, where π projects on the set of state vectors corresponding to those ϕ for which $\theta_a(\phi) = \theta_0$.

In some early quantum mechanical literature, such results were comprehended as somewhat mysterious, as reflected in the term "collapse of the wave packet." From a statistical point of view it is well known that models must be changed when new information is obtained. Bayesian statistics has put this way of thinking into a system, also for nonideal

measurements. However, the idea is well known also in non-Bayesian statistics.

11. MEASUREMENT THEORY: STATISTICAL INFERENCE THEORY

The measurement theory of von Neumann (1955) has been criticized by several authors, mainly because of difficulties with giving a precise division between the microscopic system and the macroscopic measuring device.

The statistical approach can offer a well-documented, well-understood, and extremely well tested solution to the measurement problem: Suppose that some specific experiment $\mathscr{E}_a$ has been performed. Statistical inference theory to estimate the parameter θ of such an experiment is well developed (see, for instance, Lehmann, 1983). A difficulty is that this does not usually determine the state ϕ. This may, however, be solved using an objective Bayes approach; see below. In Section 4 the extreme set $\mathscr{A}_0$ of experiments was defined. A reasonable conjecture is that under regularity conditions, when an experiment belongs to $\mathscr{A}_0$, there is 1–1 correspondence between θ and ϕ. This seems to be supported by observing the Hilbert space solution for this case.

Malley and Hornstein (1993) discuss statistical inference specifically for the quantum situation.

In the present context it is natural to make use of the symmetry properties of the quantum situation, also in the inference phase. This was done in a general setting in Helland (1998). We will only briefly recapitulate one of the main results adapted to the quantum theory solution proposed here.

Assume that to start with we have no information about the system, so that the Haar prior ν on Φ with respect to the basic group $\bar{G}$ is used. Then we perform an experiment $\mathscr{E}_a$ with parametric function $\theta(\phi)$. The experiment results in the measurement of a random variable X—which is taken as a point in the sample space, and can be multidimensional or even more general. We assume that the group G_a on the sample space $\mathscr{X}_a$ of the experiment is known. For simplicity it is assumed that this group is transitive, and also that $\bar{G}$ is transitive on Φ (no superselection rules). In general one can condition upon orbits. The probability density on the sample space $\mathscr{X}_a$ is known as a function of ϕ; for simplicity this is taken as a density $p_\phi(x)$ with respect to the measure ν_a on $\mathscr{X}_a$ generated by Haar measure ν_a' on G_a: that is, $P^a_{\theta(\phi)}(E) = \int_E p_\phi(x)\, \nu_a(dx)$ for all E.

The main question now is how shall ϕ be estimated; i.e., one wants an estimator $\hat{\phi}(X)$ which is as close as possible to ϕ. To specify what we mean

by closeness, we need to specify a loss function. Assume for simplicity that ϕ is a vector parameter (in some $\mathbb{R}^k$), and that quadratic loss is used. Then the solution is

$$\hat{\phi}(x) = \frac{\int \phi p_\phi(x)\, \nu(d\phi)}{\int p_\phi(x)\, \nu(d\phi)} \tag{19}$$

This solution is best in two different ways: It is the best equivariant ("permissible") estimator, and it minimizes the Bayes loss, that is, the expected loss with respect to the *a posteriori* distribution when the prior is Haar as above.

By consistency, $\theta = \theta(\phi)$ must have a similar formula with $p_\phi(x)$ replaced by $q_\theta(x)$, where $p_\phi(x) = q_{\theta(\phi)}(x)$. On the other hand, it seems reasonable, and it is in fact optimal, to estimate any $\eta(\phi)$ by the formula (19), where ϕ under the integral in the numerator is replaced by $\eta(\phi)$. This may seem to give two different formulas for $\hat{\theta}(x)$, but the two formulas are equal by a straightforward use of Lemma 3. Note, however, that we do not have $\hat{\theta} = \theta(\hat{\phi})$; this relation is only approximately satisfied when one has large amounts of data.

12. TIME EVOLUTION AND THE ROLE OF PLANCK'S CONSTANT

So far the description has been static; we will now try briefly to take time development into account. The point of departure is that we have defined the state of the system as some parameter ϕ in some space Φ. In practice this state can develop with time in many different ways. A simple and physically plausible assumption is that there is a continuous group K acting on Φ such that ϕ after time t changes to $k_t\phi$, where $k_t \in K$.

In Section 7 we discussed a unitary representation of the symmetry group $\bar{G}$ of Φ on a Hilbert space $\mathcal{H}$. Assume that one can find a unitary representation of K also on $\mathcal{H}$. Since K is a continuous group, a well-known theorem by Stone implies that the representation then must be of the form

$$U(t) = e^{iAt}$$

for some self-adjoint operator A, perhaps after rescaling time. As is demonstrated in several books, taking $A = -H/\hbar$ here, where $\hbar$ is Planck's constant and H is the Hamiltonian, leads to the Schrödinger equation.

This is the first time that Planck's constant appears in this paper, a rather significant observation. Everything that has been said on stochastical models and most of what we have said about symmetry apply also to large-scale problems. Planck's constant appears only when we meet the following fact,

most recently commented on by Alfsen and Schultz (1998): Physical variables play a dual role, as observables and as generators of transformation groups. It is when observators are used as generators of groups that the proportionality constant $\hbar$ occurs. One example is the Hamiltonian and the Schrödinger equation, another example is the translation group in some direction, say the x axis, whose unitary representation is of the form $\exp(ip_x/\hbar)$, with p_x being the momentum in the x direction.

13. A LARGE-SCALE EXAMPLE

Most so-called paradoxes of quantum mechanics seem to disappear in the statistical setting described here. Take the Einstein–Podolsky–Rosen paradox, for example (in the form proposed by David Bohm): A composite particle with spin 0 disintegrates into two single particles, whose spin components in any direction then necessarily add to 0, hence are opposite. Sometimes the fact that an observation of one particle implies the knowledge of the spin component of the other particle in the same direction is taken as some action at a distance. One point is that the observer of the first particle is free to choose the direction in which he wants to measure. In the language of the present paper this observer chooses one particular experiment. No paradox appears when we have a statistical interpretation of experiments in mind.

The following macroscopic example seems to be fairly similar: An organizer O sends two envelopes I and II to each of two persons A and B, situated at different places. In envelope I he has the choice between a red card or a black card: either red to A or black to B or the opposite. In envelope II he has the choice between a card with the letter 1 or a card with the letter 2: either 1 to A and 2 to B or the opposite. Now O for simplicity chooses to have probability 1/2 for red to A in envelope I, and he also chooses probability 1/2 for 1 to A in envelope II. However, he may insist that there should be some correlation γ between the two envelopes to A, for instance, defined by

$$P(\text{red}, 1) = \tfrac{1}{4}(1 + \gamma)$$

where $-1 \leq \gamma \leq 1$. Correlation at B, defined in the same way, will then necessarily also be γ.

The initial state of the system is then determined, since we know the probability distribution of any experiment that A and B could choose to do. To make completely concrete that some randomness is involved, one can imagine several independent pairs $(A_1, B_1), \ldots, (A_n, B_n)$, all in the same state, that is, under the same joint probability distribution generated by O.

Imagine now that the following happens: A is given the order that he should open one and only one envelope. Once he has opened one envelope,

the other one is destroyed (by some mechanism which is irrelevant here). A is therefore given the choice between two noncompatible experiment, exactly the situation discussed in this paper. If he opens envelope I and finds a red card, he can make one set of predictions concerning the content of B's two envelopes; if he opens envelope II and sees the number 2, he can make another set of predictions, etc. The situation is at least in some sense similar to the EPR experiment. There is no action at a distance, only predictions conditioned on different information.

As the experiment is described above, however, the well-known inequalities of Bell (1966) will always be satisfied, since the knowledge posessed by O can be regarded as a hidden variable. But the example can be extended in several directions, thus making it more equal to its quantum mechanical counterparts.

The following concept is of interest in general, and specifically in this example: Pick a set $\mathcal{A}_1$ of potential experiments. Then the set $\{(a, \theta_a)\colon a \in \mathcal{A}_1\}$ may be called permissible if there is a group G_1 on $\mathcal{A}_1$ such that (i) $g_1 a \in \mathcal{A}_1$ whenever $a \in \mathcal{A}_1$ and $g_1 \in G_1$, and (ii) $(g_1 a, \theta_{g_1 a}(\bar{g}\phi))$ is a function of $(a, \theta_a(\phi))$ for all choices of $g_1, \bar{g}$. This is equivalent to permissibility of the parameter in a compound experiment, where ϕ is augmented by the "parameter" a, and the group of this new ϕ is given by $\{(g_1, \bar{g})\}$.

In the present example, the three transformations corresponding to $P_\theta(r) \to P_\theta(b)$ (envelope I), $P_\theta(1) \to P_\theta(2)$ (envelope II), and I $\to$ II (change envolope) constitute such a structure. The smallest homomorphic image of S_4 (the permutation group of four elements) which supports this structure is S_3, which is noncommutative and has an interesting two-dimensional irreducible representation.

Several concepts of this paper may be illustrated on this simple example, if necessary by using the structure just described. Alternatively, the example could have been made more complicated to start with, so that each experiment supported a group with a nontrivial irreducible representation. These (not quite trivial) exerciseswill be left to the interested reader.

14. DISCUSSION

Throughout this paper we have stated some assumptions: in addition there are several implicit assumptions of more technical/mathematical type. Though most of these assumptions are rather plausible, we are not completely sure that all of them will be satisfied in all details in a future hypothetical theory. Their main function here has been to pave the way to ordinary quantum mechanics from what is our basic framework: a set of models for experiments with a common state space, subject to some symmetry group.

Our conceptual framework works both for continuous and for discrete variables, and, apart from the discussion of time evolution, there is nothing much nonrelativistic about it. A very interesting—but probably not trivial—question is therefore to what extent it can be developed further in a relativistic setting, perhaps even with large symmetry groups of the kind that are natural to think of in the contexts of general relativity.

Of some interest in this connection are the views expressed by Dirac (1978): He wanted relativistic quantum mechanics to take another direction, and explicitly mentioned group representations of the Poincaré group as a clue to this direction.

In the same paper, Dirac expressed the following opinion: "One should keep the need for a sound mathematical basis dominating one's search for a new theory. Any physical or philosophical idea that one has must be adjusted to the mathematics. Not the other way around."

I strongly agree with Dirac that any fundamental physical theory should have a sound mathematical basis. In fact, much of the motivation behind the present paper has been to find just that. However, every mathematical theory building must be based on a set of axioms. It may well be that the axioms found by mathematicians using aesthetic or quasilogical criteria are among the right ones to serve as a basis for a physical theory, but these criteria alone often leave one with too much arbitrariness. In fact, and unfortunately, much formalistic theory has been the result of only taking such points of departure. To find the basis for a new theory, one must endeavor to look in all directions, draw upon all sources whether they relate to common sense, experience, mathematics, philosophy, or large-scale physics.

This implies extremely difficult requirements on the developer of theories, and intuitive reasoning and even imprecise arguments seem to be unavoidable at intermediate stages of the development. Having said that, however, at the final stage mathematical rigor should be imperative.

It can be inferred from the ideas of this paper that much of what has been said about modeling in the microcosmos also can be transferred to the macrocosmos. In fact, large-scale examples showing quantum mechanical behavior have recently been demonstrated in the physical literature (Aerts, 1996). As indicated above, several examples of this type can be constructed.

Having stressed the similarities between models of quantum theory and those of statistics in this paper, it should also of course be underlined that there are basic differences. In statistics the parameters are always regarded as unknown and subject to inference; in quantum physics the states/parameters can often be regarded as known. Also, the rich set of symmetries that is typical for all applications of quantum theory seldom find their counterpart, at least not to a similar degree, in large-scale statistics.

Nevertheless, even though it is not a main issue here, it should be mentioned that some of the ideas behind this paper also seem to have impact for issues discussed in ordinary statistical science. The question of how to condition statistical models gives another meaning to the label 'a' in $\mathcal{E}_a = (\mathcal{X}_a, \mathcal{F}_a, \{P_\theta^a; \theta \in \Theta_a\})$ (Helland, 1995, and references there). The link between this issue and quantum mechanics was suggested already by Barndorff-Nielsen (1995).

In this paper, and in most statistical papers, a mathematical structure like $\mathcal{E}_a$ above is connected to a real, physical or otherwise, experiment. But in addition, since $\{P_\theta^a\}$ is included in the structure, it can be used to denote different models for the same real experiment. This idea together with the symmetry discussion of the present paper can in fact be used to discuss model reduction in statistics. We hope to develop these ideas further elsewhere.

REFERENCES

Aerts, S. (1996). Conditional probabilities with a quantal and a Kolmogorovian limit. *Int. J. Theoret. Phys.* **35**, 2245–2261.

Alfsen, E. M., and F. W. Schultz (1998). Orientation in operator algebras. *Proc. Natl. Acad. Sci. USA* **95**, 6596–6601.

Aspect, A., J. Dalibard, and G. Roger (1982). Experimental test of Bell inequalities using time-varying analysers. *Phys. Rev. Lett.* **49**, 1804ff.

Barndorff-Nielsen, O. E. (1995). Diversity of evidence and Birnbaum's theorem. *Scand. J. Statist.* **22**, 513–522.

Barut, A. O., and R. Rączka (1977). *Theory of Group Representations and Applications*. Polish Scientific Publishers, Warsaw.

Bell, J. S. (1966). On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447–452.

Beltrametti, E. G., and G. Cassinelli (1981). *The Logic of Quantum Mechanics*. Addison-Wesley, London.

Beran, L. (1985). *Orthomodular Lattices. Algebraic Approach*. Reidel, Dordrecht.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Bohr, A., and O. Ulfbeck (1995). Primary manifestation of symmetry. Origin of quantal indeterminacy. *Rev. Mod. Phys.* **67**, 1–35.

Cox, D. R., and D. V. Hinkley (1974). *Theoretical Statistics*. Chapman and Hall, London.

Dirac, P. A. M. (1978). The mathematical foundations of quantum theory. In Marlow, A. R., ed., *Mathematical Foundations of Quantum Theory*. Academic Press, New York.

Dunford, N., and J. T. Schwartz (1958). *Linear Operators. Part I: General Theory*. Interscience, New York.

Feynman, R. P. (1951). The concept of probability in quantum mechanics. In *Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, California.

Gudder, S. P. (1978). Some unsolved problems in quantum logic. In Marlow, A. R., ed., *Mathematical Foundations of Quantum Theory*. Academic Press, New York.

Gudder, S. P. (1988). *Quantum Probability*. Academic Press, Boston.

Hamermesh, M. (1962). *Group Theory and its Applications to Physical Problems*. Addison-Wesley, Reading, Massachusetts.

Helland, I. S. (1995). Simple counterexamples against the conditionality principle. *Am. Statist.* **49**, 351–356; Correction/discussion, *Am. Statist.* **50**, 382–386.

Helland, I. S. (1998). Statistical inference under a fixed symmetry group.*Statistical Research Report No. 7*, Department of Mathematics, University of Oslo.

Kalmbach, G. (1983). *Orthomodular Lattices*. Academic Press, London.

Lehmann, E. L. (1983). *Theory of Point Estimation*. Springer-Verlag, New York.

Mackey, G. W. (1963). *The Mathematical Foundations of Quantum Mechanics*. Benjamin, New York.

Mączyński, M. J. (1973). The orthogonality postulate in axiomatic quantum mechanics. *Int. J. Theor.* **8**, 353–360.

Maeda, F., and S. Maeda (1970). *Theory of Symmetric Lattices*. Springer-Verlag, Berlin.

Malley, J. D., and J. Hornstein (1993). Quantum statistical inference. *Stat. Sci.* **8**, 433–457.

Meyer, P.-A. (1993). *Quantum Probability for Probabilists*. Springer-Verlag, Berlin.

Nachbin, L. (1965). *The Haar Integral*, Van Nostrand, Princeton, New Jersey.

von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton, New Jersey.

Piron, C. (1976). *Foundation of Quantum Physics*. Benjamin, New York.

Varadarajan, V. S. (1985). *Geometry of Quantum Theory*. Springer, New York.