Relations between Quantum Theory and Symmetry Structures in a Parameter Space for Statistical Experiments.

Inge S. Helland *

Abstract

The aim of this paper is to show a connection between an extended theory of statistical experiments on one hand and the foundation of quantum theory on the other hand. The main aspects of this extension are: One assumes a hyperparameter space $\Phi$ common to several potential experiments, and a basic symmetry group $G$ associated with this space. The parametric function $\theta_a(\cdot)$ of a single experiment is said to be permissible if $G$ induces in a natural way a new group on the image space of the function. If this is not the case, it is arranged for by changing from $G$ to a subgroup $G_a$. The Haar measure of this subgroup is the preferred prior when the parameter is unknown. It is assumed that the hyperparameter itself can never be estimated, only a set of parametric functions. Very much of the ordinary quantum theory formalism can be retrieved from these and a few related assumptions. The scope of the theory is illustrated on the one hand by the example of the spin 1/2 particle and a related EPR discussion, on the other hand by a simple macroscopic example.

Keywords and phrases: Quantum theory; Hilbert space; hyperparameter; parametric invariant space; permissible parameter; statistical experiment; symmetry group; transitive subgroup.

1 Introduction

In Helland (1998a) it was argued tentatively that the basic principles of quantum mechanics in its usual Hilbert space formulation is consistent with an ordinary general

*Department of Mathematics, University of Oslo, P.O.Box 1053 Blindern, N-0316 Oslo, Norway. E-mail: ingeh@math.uio.no
statistical modelling framework which includes the following features: 1) For a given system one has the choice between several experiments that can be performed. 2) There is an unobservable hyperparameter $\phi$ which is common to all potential experiments. 3) There is a symmetry group $G$ acting on $\Phi = \{\phi\}$. (In Helland, 1998a, the notation $\mathcal{G}$ was used, since $G$ was reserved for the corresponding sample space group; in the present paper we will concentrate on parameter transformations, however, so we change notation.) 4) The functions $\theta_a(\phi)$, giving the parameter of experiment $a$, satisfy a natural permissible condition as a function of the hyperparameter $\phi$, namely: $\theta_a(\phi_1) = \theta_a(\phi_2)$ implies $\theta_a(g\phi_1) = \theta_a(g\phi_2)$ for all $g$.

We will show in this paper that in order to derive a version covering ordinary quantum mechanical systems from these assumptions, one has to include the case when the permissible condition is not with respect to $G$, but only with respect to a subgroup $G_a$ of $G$. In each experiment, we let $G_a$ be the maximal subgroup for which $\theta_a$ is permissible. In the case where $\phi$ is unknown, this is expressed relative to each experiment $a$, namely by taking the Haar measure relative to $G_a$ as a prior.

The arguments of Helland (1998a) were based on quantum logic, and gave therefore only an indirect construction of the Hilbert space in question. Here we will give a direct construction using group representation theory. The ordinary formula for probabilities and expectations will be given a general derivation.

Our ultimate aim is to try to make a link between the world of theoretical statistics and the world of quantum physics. Of course, it cannot be expected that all aspects of such a unity-of-science idea can be explored in a single paper like this. If the main idea here is accepted, though, there seems to be room for further developments both within statistics and within physics; some tentative ideas follow.

For theoretical statistics: The present focus is very much on classes of probability measures where each class has a relatively small set of parameters. In many applications this can be complemented by a rather rich and useful theoretical platform focused on a large parameter space where only a few specific parametric functions are of direct interest. Sometimes this same parameter space can then be used in probability models for several potential experiments, perhaps also for single models conditioned in different ways. Of course, this is a much more complicated concept than the standard one, but one can compensate by paying less attention to the detailed properties of the probability distribution chosen. Ideas not far from this are in fact also discussed in good books in experimental design, like Box, Hunter and Hunter (1978) and in books and papers in data analysis by various authors.

The other specific feature of the models in this paper: a close link between probability models and symmetry groups, will be amply discussed below. Such a link was argued for by using purely statistical arguments in Helland (1998b).

For theoretical physics: As many others who try to look at quantum theory from the outside, I first of all miss simplicity of fundamental ideas. Although this
definitely constitutes the standard theory with a huge literature and ample experimental support, it still seems to me to be difficult to accept as the final solution a theory that takes something as abstract as rays in a Hilbert space as its basic state concept, and where the relativistic generalizations have even more abstract foundations. Renouncing this foundation when no satisfying alternative exists, is of course impossible, but there do exist books on quantum theory that have a relatively open attitude towards foundational issues (Isham, 1995; Peres, 1993, and references there), and radically new foundations have at times been proposed.

As a contrast to the Hilbert space concept of state, the concept of symmetry, and hence of symmetry groups acting on some given space or set, is a very simple one, basically, but still a concept from which much interesting mathematics can be developed. One hope might be that one day a theoretical physical platform shall emerge based on simple concepts like sets and symmetry groups, and that derived vector space concepts, for instance group representations, will be held for what they really are: extremely useful calculating tools.

Also, the statisticians’ attitude towards mathematical modelling of real phenomena is today somewhat different from what we find in theoretical physics: In statistics, nearly all models are approximations, and they always depend upon unknown, hypothetical parameters, whose value we can never determine completely, only find uncertain estimates of. Physicists seem to be more used to thinking of an absolute, objective world behind their endeavours. In the long run, we may perhaps have something to learn from each other.

Some physicists may feel that the discussion that follows is fairly abstract. I can only give two comments on this: First, the point of departure, models for statistical experiments, may be unfamiliar to physicists; however, for the present paper one does not really need much more from this than the fact that everything depends upon some space of unknown parameters. Secondly, the main mathematical work here is really to demonstrate that a version of the ordinary Hilbert space based quantum theory is indeed consistent with our framework, a framework which in the end turns out to be fairly simple; see the next Section.

The plan of the paper is as follows: The simple example of a spin $1/2$ particle is treated in Section 2, and used to give a short discussion of the EPR paradox. In Section 3 we define and discuss basic concepts like statistical experiments, transformation groups and permissibility of parametric functions. Section 4 gives the relationship between permissible parametric functions and invariant subspaces of a Hilbert space, a basis for the relation between a model point of view and the ordinary quantum theory formalism. Appendix 1 on group representations is also relevant here. Section 5 studies in detail the correspondence between parameter values and Hilbert space vectors. Section 6 contains more about experiments, and in Section 7 structures in a parameter space are related to groups and experiments. Then Section
8 gives a rather thorough discussion of the simplest type of experiments, those where the parameters of interest are permissible with respect to the underlying group. The quantum theoretical formulae for expectation and for change of state are established for this case. This also leads to the rules for change of measure during a measurement. Section 9 generalizes this to the case of several parametric functions. In Section 10 we give the first discussion of experiments with non-permissible parameters, and also give some technical results needed for the final discussion. Finally, in Section 11 we give the main results on comparing the two approaches to modelling and predicting experimental results. The rest of the paper consists of examples and discussion.

Related ideas, namely that symmetries should constitute the main foundation behind quantum theory, have recently been advocated in great detail from a physical point of view by Bohr and Ulffbeek (1995). The present paper aims at making such ideas as precise as possible, which seems to make necessary a point of departure radically different from what is ordinary. However, by using various mathematical tools, most notably the representation theory of groups, the resulting formalism will in the end be very close to ordinary quantum theory.

2 Spin 1/2 particles and the Einstein-Podolsky-Rosen paradox.

As a relatively simple example to illustrate the general theory of this paper, we give a nonstandard description of a spin 1/2 particle, a description which will turn out to be essentially equivalent to the one given by ordinary quantum theory.

A hyperparameter $\phi$ for such a particle may be defined as a unit vector in three dimensional space. The natural associated group $G$ will then be the group of rotations of this vector in $\mathbb{R}^3$, and the invariant (Haar) probability measure will be the normalized area on the sphere. Note that it is impossible according to quantum theory to get full information about this spin vector; all that is possible is to measure the component in some given direction; in fact, for a spin 1/2 particle it is only possible to determine the sign of the spin component in a given direction, and this determines the (pure) state.

If the spin component in some direction defined by a unit vector $c$ has been measured to be $+1/2$, we let the new spin state be defined by $\theta_c = +1$; if it was measured to be $-1/2$, the state is $\theta_c = -1$, where $\theta_c$ is the sign of the cosine of the angle between $\phi$ and $c$. Mixed states are formed by assigning probabilities to these two pure states.

For a given system in a pur or mixed state, and for each fixed unit vector $a$ there is an experiment which consists of making a new measurement of the spin
component of the particle in the direction \(a\), using a Stern Gerlach device oriented in that direction. This is in fact a very general situation both in quantum physics and in large scale experimentation: For a given system we always have a choice of experiments that can be done, and it is always possible to index these experiments by some index \(a\). A meaningful definition of a ‘state’ is then some mathematical quantity from which we in principle can deduce the probability distribution of the outcome of each single such experiment.

Every real experiment is imperfect. In this paper we will for simplicity concentrate on ideal experiments, where the outcome is found from symmetry considerations, and where the system after the experiment is done may be considered to be in a new state with a new parameter. Therefore, we will largely denote the result of a new experiment by a new parameter \(\theta\) instead of by a stochastic variable \(x\), which is common in statistics. This notation is also closer to what is common in quantum physics, where experimental error is usually ignored.

In the spin case, assume that the state is known as \(\theta_c = \pi\). Then for the particular experiment indexed by \(a\), we define the parameter \(\zeta_a\) to be the dot product between \(\pi_c\) and \(a\), i.e., \(\zeta_a\) is the cosine of the angle between the two vectors. The probability distribution of the outcome \(x\) of the Stern Gerlach experiment turns out to depend only upon \(\zeta_a\), and is given by

\[
P(x = -\frac{1}{2}) = \frac{1}{2}(1 - \zeta_a), \quad P(x = +\frac{1}{2}) = \frac{1}{2}(1 + \zeta_a).
\]

The standard quantum theoretical derivation of these formulae may be found in several textbooks. An alternative motivation can be found from a symmetry argument: Looking at a simple figure, it may be reasonable to assume that the two probabilities are proportional to the Lebesgue measure of the parts of the diameter along \(a\) determined by the division point corresponding to \(\zeta_a\).

Note that it follows from equation (1) that \(E(x|\phi) = \frac{1}{2}\zeta_a\), a result that will be needed below.

It is easily seen that the parametric function \(\theta_\phi = \theta_\phi(\phi)\) is non-permissible according to the definition given in the introduction. In fact, the largest group \(G_a\) for which it is permissible, is the group of rotations around \(a\) together with a reflection in the plane perpendicular to \(a\). (This is the group of elements that take the sets of the form \(\{\theta_\phi = \text{const.}\}\) into sets of the same type.)

This group has several orbits: For each \(k \in (0, 1]\), one orbit is given by \(\theta_\phi = k\) and \(\theta_\phi = -k\); for \(k = 1\) these are degenerated. In addition there is an orbit for \(k = 0\).

With these observations we are ready to introduce the prior distribution of the spin vector \(\phi\) in connection to a measurement along the direction \(a\). It is crucial that we are willing to let this prior depend on the direction \(a\). As a general point, a point that can also be advocated from statistical arguments; see Helland, 1998b, it is
natural to use as a prior any Haar measure with respect to a group $G_a$ for which $\theta_a$ is permissible. Note that this dependence of prior upon the choice of measurement makes the present theory different from ordinary hidden variable theories.

In the spin case we also have an extra complication: The group $G_a$ is non-transitive, i.e., has several orbits. This implies that the Haar measure of this group is non-unique. With a single particle this does not matter much; in fact, any such measure over $\phi$ will give $E(\theta_a) = 0$, and hence from (1):

$$P(x = -\frac{1}{2}) = P(x = +\frac{1}{2}) = \frac{1}{2}.$$ 

Nevertheless, we will make a special choice of Haar measure also for the non-transitive case: We will choose the orbit for which the Fisher information (see Frieden, 1995 and references there) for the distribution of $x$, given $\phi$, is as small as possible. In the present case this can be found without any calculation at all: For Fisher information $I_\phi$, the distribution with density $I_\phi^{-1}$ is the so-called Jeffrey prior, in general a noninformative prior, and in the present case with compact orbits this is proportional to the Haar density. And this density is clearly maximized over orbits as $k \rightarrow 1$. Hence from this argument it is natural to use the following prior connected to a spin measurement in the direction $a$: Probability 1/2 on $\theta_a = -1$ and 1/2 on $\theta_a = +1$.

Consider now the Einstein-Podolsky-Rosen situation (as modified by D. Bohm), where two particles previously have been together in a spin 0 state, so that they - in our notation - later have opposite spin-vectors $\phi$ and $-\phi$. This situation has been very much discussed in the literature, where the focus often has been on the fact that one here gets correlated spin component measurements even at great distance.

This in itself is not really surprising, however, since the two particles have a common history. Correlation through a common history can be found in many instances also in large-scale statistics. A much more serious difficulty is that Bell’s inequalities preclude explanation of the correlations via ordinary hidden variable models.

But even this difficulty is solved by letting the prior depend on the choice of measurement as described above: Let an observer $A$ measure the spin of one particle along a direction $a$, and at some distance from this, let another observer $B$ measure the spin of the other particle along some direction $b$. Let $u$ be the angle between the directions $a$ and $b$.

Let us first describe the situation from the point of view of the observer $A$. As explained above, this observer will have a prior on $\phi$ given by a probability 1/2 on $\theta_a = -1$ and a probability 1/2 on $\theta_a = +1$, where $\theta_a$ is the cosine of the angle between $a$ and $\phi$. This is equivalent to a prior on $\phi$ with probability 1/2 on each of $a$ and $-a$. Let $b^\perp$ be a vector in the plane determined by $a$ and $b$, perpendicular to $b$. 

6
Then, from the point of view of $A$, $\phi$ will have a prior concentrated on $a = b\cos u + b^{\dagger}\sin u$ and $-a$, so that the prior component of $-\phi$ along $b$ will be $-b\cos(u)$, again taking two opposite values with equal probability. As before, $E(x_a) = E(x_b) = 0$, where $x_a$ and $x_b$ are the two measured spin components, and we find

$$E(x_a x_b) = E(E(x_a x_b | \phi)) = \frac{1}{4} E(\theta_a \cdot (-\theta_a \cos(u))) = -\frac{1}{2} \cos(u),$$

(2)

so that the correlation between the spin component measurements is $-\cos(u)$, the same answer as given by ordinary quantum theory calculations.

The above calculation was made from the point of view of $A$, using his prior. It is easily seen that the same answer is found from the point of view of $B$. It is crucial that the argument requires no action at a distance, contrary to what seems to be claimed even in recent papers discussing quantum entanglement (e.g., Buchanan, 1998).

Similar conclusions were recently reached by Deutsch and Hayden (1999) using the Heisenberg picture to analyse quantum information. It may be of some interest to compare briefly these two approaches. Deutsch and Hayden make extensive use of the cnot operation which is a logical operation connected to the spin measurements. In particular, each of the two spin component measurements is associated with a cnot operator, which can be interpreted as described above. A crucial point, however, is that a third cnot operator is needed after the two measurements. This may be linked to a comparison of the result found by $A$ and the result found by $B$, and is thus directly connected to the correlation found. From a statistical point of view, the usual model for a measurement is an interaction between two players: nature and the observer. In the present case it may to some extent be fruitful to introduce a related concept to describe the interplay between two observers with different priors. At least this may seem to be a way to make the argument of Deutsch and Hayden (1999) less formal, and thus making it fairly close to the argument sketched above.

3 Framework and background.

The ordinary statistical framework for modelling experiments will be used: A sample space $\mathcal{X}$ listing the possible outcomes of the experiment, a $\sigma$-algebra (Boolean algebra) $\mathcal{F}$ of subsets of $\mathcal{X}$, giving the events to which probabilities can be assigned and a class $\{P^\theta\}$ of probability measures on $(\mathcal{X}, \mathcal{F})$. In statistics the parameter $\theta$ is usually thought of as the unknown quantity in the situation. The statistical concept of state of the system will then either be a fixed value of $\theta$ or a probability measure over the parameter space $\Theta$. 

7
We add the following feature, which we think will be of value both in statistical theory and in quantum theory: In a given situation we may have the choice between several potential experiments, sometimes mutually exclusive. Let \( a \in A \) index the possible experiments. Then the basic mathematical construct will be \( (\mathcal{X}_a, \mathcal{F}_a, \{P^{a}_a; \theta_a \in \Theta_a\}; a \in A) \). Though some may think that this is a more elaborate construct than the ordinary Hilbert space basis of quantum theory, it has the advantage that each element has a concrete interpretation, and that it is essentially the same framework that is used in the analysis of macroscopic experiments in a large range of empirical disciplines.

In this extended framework it may be natural to assume that each \( \theta_a \) is a function of some hyperparameter \( \phi \), varying over a space \( \Phi \), (in fact, by taking \( \Phi \) large enough, this is no essential restriction). Then the natural continuation of the statistician’s conceptual ideas would be to take the hyperparameter \( \phi \) as the new state concept. However, at this point a variant of the physicist’s important concept of complementarity comes in: There are important situations where as a matter of principle, it is completely impossible, by any conceivable experiment, to get full information about \( \phi \). The only parameters that are possible to estimate, are certain parametric functions \( \theta = \theta(\phi) \). By complementarity we then mean that if \( \theta_1(\cdot) \) is to be estimated accurately at a given point of time, then there exists no conceivable experiment which can estimate \( \theta_2(\cdot) \) accurately at the same point of time. This is a well known phenomenon in the quantum world; in my opinion one can find many such situations in the macroscopic world, too.

To be able to study closer the link between this framework and the usual quantum theory framework, we have chosen to make symmetry assumptions: Let \( G \) be a fixed symmetry group on \( \Phi \). It is assumed that the space \( \Phi \) is closed under the transformations in \( G \). Often the structure of the parameter space is such that there is a very natural choice of \( G \): If \( \Phi \) consists of only location and scale parameters, say, then the corresponding location and scale groups are used; if direction parameters are added, then this may make it natural to adjoin rotation group(s) and so on.

A group acting on some space \( \Phi \) is said to be exact if there is at most one group element \( g \) transforming any given point \( \phi_1 \) to any other point \( \phi_2 \). The group is transitive if there always is at least one such \( g \). Nontransitiveness implies several orbits in the space \( \Phi \); this corresponds to superselection rules, which can be tackled by taking each orbit separately. (This must not be confused with the nontransitivity of the subgroup \( G_0 \) which was encountered in the previous section, and will be further discussed later.) Non-exactness in an ordinary statistical setting was discussed in Helland (1998b).

Assume that the hyperparameterspace \( \Phi \) is locally compact, has a right Haar measure \( \nu \) with respect to \( G \) and satisfies weak requirements such that the space \( \mathcal{H} = L^2(\Phi, \nu) \) of complex square integrable functions on \( \Phi \) is a separable Hilbert
space. This will be the Hilbert space on which we will base our construction leading to quantum theory. One should have in mind, however, that all separable Hilbert spaces are equivalent, so one can as well think of a general, abstract Hilbert space.

As mentioned before, and following Helland (1998a,b) we define the parametric function \( \theta(\cdot) \) on \( \Phi \) to be \textit{permissible} with respect to the group \( G \) if we always have that \( \theta(\phi_1) = \theta(\phi_2) \) implies \( \theta(g\phi_1) = \theta(g\phi_2) \) for all \( g \in G \).

The first implication of this important concept is that a new group \( \tilde{G} \) can be introduced, a group of transformations \( \tilde{g} \) on the image space \( \Theta \) of the permissible parametric function, where \( \tilde{g} \) is defined by

\[
\tilde{g}(\theta(\phi)) = \theta(g\phi).
\]

This may often imply that it is possible to change focus from the large space \( \Phi \) with the big group \( G \) to the smaller space \( \Theta \) with the simpler group \( \tilde{G} \). Note that the mapping \( g \rightarrow \tilde{g} \) is a homomorphism: \( g_1 \rightarrow \tilde{g}_1, g_2 \rightarrow \tilde{g}_2 \) imply \( g_1^{-1} \rightarrow \tilde{g}_1^{-1} \) and \( g_1g_2 \rightarrow \tilde{g}_1\tilde{g}_2 \).

Secondly, this notion turns out to be important in statistical inference theory, as shown in Helland (1998b): Certain paradoxes in Bayesian inference are avoided when inference is confined to permissible parameters, best invariant estimators equal Bayesian estimators under Haar prior, and credibility sets (under Haar prior) and confidence sets coincide. Further results on Haar measures from the assumption of permissibility are derived in Section 7 below.

4 Subspaces and parametric functions.

A third class of permissibility-results will be of direct importance for the present paper: An order-preserving link is established between the permissible parametric functions on \( \Phi \) and the closed invariant subspaces of \( \mathcal{H} = L^2(\Phi, \nu) \) under the regular group-representation corresponding to the group \( G \).

This uses just the natural definitions. First, the regular unitary representation \( U_R(G) = \{ U_R(g) \} \) of \( G \) on \( \mathcal{H} \) is defined by

\[
U_R(g)f(\phi) = f(g^{-1}\phi).
\]

Here, unitarity and the representation property (cf. Appendix 1) are easily checked from the definition.

Next, the permissible functions on \( \Phi \) are ordered in the most natural way: Say that \( \theta_1(\cdot) \preceq \theta_2(\cdot) \) if \( \theta_1(\phi) = \psi(\theta_2(\phi)) \) for some function \( \psi \). If this ordering holds in
both directions, so that the two permissible parameters are one-to-one functions of each other, they are said to be equivalent.

Before proceeding it may be useful to look at the summary of group representation theory given in Appendix 1.

The first of the following results were also proved in Helland (1998a).

**Theorem 1.**

(a) For a given permissible parametric function $\theta(\cdot)$, the set $V = \{ f \in \mathcal{H} : f(\phi) = \hat{f}(\theta(\phi)) \}$ for some $\hat{f}$ is a closed subspace which is invariant under the regular representation $U_R$. The space $V$ depends only on the equivalence class of parametric functions $\theta(\cdot)$. The invariant subspaces of this form will in the following be called **parametric** invariant subspaces of $\mathcal{H}$.

(b) Every closed, invariant subspace $V$ of $\mathcal{H}$ is contained in a parametric invariant space with $\theta(\cdot) = (u_i; i = 1, \ldots)$, a basis for $V$.

(c) Ordering of the parametric invariant subspaces under inclusion correspond to ordering of parametric functions under $\preceq$. The space $V$ is irreducible within the class of parametric invariant spaces if and only if the corresponding permissible function $\theta(\cdot)$ is minimal with respect to the ordering $\preceq$. Nonequivalent minimal permissible functions correspond to irreducible parametric spaces with trivial intersection.

**Remark.**

It may seem that $V$ is typically trivially embedded into a very large parametric invariant space in (b). However, look at the case where $V$ is irreducibly invariant of dimension $d$. Then, by using the Clebsch-Gordan series for polynomial functions, and taking limits in general, one can show that the embedding space can be linearly spanned by the matrix elements of the representing matrix, thus has dimension at most $d^2$.

**Proof.**

(a) Let $\theta(\cdot)$ be permissible and let $V = \{ f \in \mathcal{H} : f(\phi) = \hat{f}(\theta(\phi)) \}$ for some $\hat{f}$. It is clear that this set of functions is closed under linear combinations, and also under infinite sums that converge in $L^2$-norm, so that it is a closed space. If $f \in V$, then

$$U_R(g)f(\phi) = f(g^{-1}\phi) = \hat{f}(g^{-1}(\theta(\phi))),$$

so $U_R(g)f \in V$. Hence the space is invariant.
It follows directly from the definitions that $\theta(\cdot) \preceq \theta_1(\cdot)$ iff the corresponding subspaces satisfy $V \subseteq V_1$. In particular, equivalence of parametric functions implies equality of subspaces.

(b) Let $\{u_i(\cdot); i = 1, 2, \ldots\}$ be a fixed orthonormal basis for $V$.

For fixed $\phi$, it is convenient to regard $u(\phi) = (u_i(\phi); i = 1, 2, \ldots)$ as a vector in $\mathbb{R}^i$. For each $i$ and $g$ we have that $u_i(g\phi) = U_i(g^{-1})u_i(\phi)$ as a function of $\phi$ is a vector in $V_i$, and can therefore be expressed in terms of basis functions: $u_i(g\phi) = \sum_j U_{ij}(g)u_j(\phi)$, or $u(g\phi) = U_R(g)[u(\phi)]$ for some operator $U_R(g)$ on $\mathbb{R}^i$. From this last representation it follows that the vector function $\theta(\phi) = (u_i(\phi); i = 1, 2, \ldots)$ is permissible.

Since $f(\phi) = a^\dagger u(\phi)$, it follows that all vectors $f \in V$ are functions of $u(\cdot)$, which is permissible. Thus $V$ is contained in the invariant space corresponding to $\theta(\cdot)$.

(c) The first two statements follow directly from the definitions. Assume that two non-equivalent minimal permissible parameters $\theta_1$ and $\theta_2$ have spaces with nontrivial intersection $V_0$. Then $V_0$ must be an invariant space in the ordinary sense, and can without loss of generality be assumed to be irreducible (again in the ordinary sense). Let it have basis $u_0(\cdot)$, and let the space corresponding to $\theta_i(\cdot) (i = 1, 2)$ have basis $(u_0(\cdot), u_i(\cdot))$. Then by (b) and Lemma 2, we have $\theta_i(\cdot) = \psi_i(u_0(\cdot), u_i(\cdot))$ for some function $\psi_i$. By permissibility $u_i(g\phi) = \tilde{g}_i(u_i(\phi))$ for group elements in the three groups $G_0$, $G_1$ and $G_2$. Fix $\phi_0$ and take $\tilde{g}_i(\phi) = \psi_i(u_0(\phi), u_i(\phi_0))$. Then one can verify from the definition that these are permissible, and this contradicts the assumption that $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are minimal.

**Corollary 1.**

If $V_i (i = 1, \ldots, k)$ are parametric invariant subspaces of $H$ corresponding to the permissible parameters $\theta_i (i = 1, \ldots, k)$, then $V = \bigcap V_i$ is a parametric invariant subspace if and only if the minimum of $\{\theta_i; i = 1, \ldots, k\}$ with respect to the ordering $\preceq$, is a permissible parameter, and in the latter case this permissible parameter generates $V$.

**Corollary 2.**

Assume $f_0(\phi) = \tilde{f}(\theta(\phi))$ for some permissible parameter $\theta$. Then the closure of $\{U_R(g)f_0; g \in G\}$ is an invariant set in $H$ contained in and spanning the parametric invariant subspace corresponding to $\theta(\cdot)$.  

11
5 Hyperstates and hyperparameters.

While states of a system in statistical modelling are most naturally given by points \( \phi_1 \) in some parameter space \( \Phi \), in quantum theory they are given as rays in some Hilbert space. We will choose here to represent the Hilbert space by \( \mathcal{H} = L^2(\Phi, \nu) \).

The question is: Can one associate in a natural way a unique ray represented by a vector \( f_{\phi_1} \in \mathcal{H} \) to every \( \phi_1 \in \Phi \)? A simple informal solution might be to let \( f_{\phi_1} \) be \( \delta(\phi - \phi_1) \), but this is an improper function, and does certainly not belong to \( \mathcal{H} \).

A much better solution is to make use of the group structure associated with the Hilbert space. We will assume here that the group \( G \) is exact and transitive, so that for any pair \((\phi_0, \phi_1)\) there is exactly one \( g \in G \) such that \( \phi_1 = g\phi_0 \). A more general discussion of the same concepts may be found in Perelomov (1986). A brief comment to the general situation will be given at the end of the Section. The main results here are:

**Theorem 2.**

Assume that the group \( G \) is exact and transitive on a parameter space \( \Phi \), and let \( \{U_R(g)\} \) be its unitary regular representation on \( \mathcal{H} = L^2(\Phi, \nu) \).

(a) Fix \( f_0 \in \mathcal{H} \) and \( \phi_0 \in \Phi \). There is then a homomorphism from the parameters \( \phi_1 \in \Phi \) (with group elements \( g \)) to the set \( F \) of functions \( f_{\phi_1} \) (with group elements \( U(g) \)) which is given by \( f_{\phi_1}(\phi) = f_0(g_1^{-1}\phi) \) for \( \phi_1 = g_1\phi_0 \).

The set \( F \) is invariant under \( \{U(g)\} \), and \( F \) is contained in and spanning the least invariant subspace of \( \mathcal{H} \) containing \( f_0 \). The functions \( f_{\phi_1} \) are unit vectors in \( \mathcal{H} \). This homomorphism is an isomorphism if we assume that no nontrivial \( f \in \mathcal{H} \) is a strictly periodic function.

(b) Let \( V_1 \) be an irreducible invariant subspace of \( \mathcal{H} \) of dimension larger than 1 such that \( V_1 \cap F \) is non-trivial. Assume that no nontrivial subgroup of \( G \) has a representation of dimension 1 in \( V_1 \). Then the vectors in \( V_1 \cap F \) lie in different one-dimensional subspaces, and these span \( V_1 \).

(c) There is a larger space \( \Phi' \supset \Phi \) such that every one-dimensional subspace \( V_1 \) of \( \mathcal{H} \) is a pure state in the sense that there is a \( \phi' \in \Phi' \) that is mapped upon \( V_1 \) in the way described in (a). The group \( G \) can always be extended to a group \( G' \) such that \( \Phi \) and \( \Phi' \setminus \Phi \) are different (unions of) orbits for \( G' \). For any such extension, and corresponding extension of Haar measure \( \nu \) to \( \nu' \), we have that \( \mathcal{H} \), extended by putting functions equal to 0 on \( \Phi' \setminus \Phi \), is an invariant subspace of \( \mathcal{H}' = L^2(\Phi', \nu') \).

**Proof.**
(a) We have that
\[ U_R(g)f_{\phi_1}(\phi) = f_0(g^{-1}g^{-1}\phi) = f_0((gg_1)^{-1}\phi) = f_{g\phi_1}(\phi). \]
Assume that this homomorphism is not an isomorphism, so that \( U_R(g_1)f_0 = U_R(g_2)f_0 \) for some \( g_1 \neq g_2 \). Then \( U_R(g)f_0 = f_0 \) for \( g = g_1g_2^{-1} \), hence \( f_0(g^{-1}\phi) = f_0(\phi) \) for all \( \phi \). This is only possible if \( f_0 \) is strictly periodic. The rest is easy to prove. (Compare also Corollary 2.)

(b) The functions constructed above are of the form \( f_{\phi_1}(\phi) = f_0(g_1^{-1}\phi) \) if \( \phi_1 = g_1\phi \). Assume that two of these, say \( f_{\phi_1} \) and \( f_{\phi_2} \) lie in the same one-dimensional subspace. Then, by the linear independence of the basis vectors we get \( U_R(g_2)f_0 = cU_R(g_1)f_0 \) for some scalar \( c \neq 1 \), thus \( U_R(g_1^{-1}g_2)f_0 = cf_0 \). By taking norms, we find \( |c| = 1 \). Assume that \( c \) is a complex number different from 1. This implies that the cyclic group generated by \( g = g_1^{-1}g_2 \) has a one-dimensional irreducible representation on the vector-space spanned by \( v_1 \), which leads to a contradiction.

We note that as \( g_1 \) varies, \( U_R(g_1)f_0 = f_{g_1\phi_0} \) span \( V_1 \), since if it spanned a smaller space \( V_0 \), this would be invariant under the representation.

(c) Take \( \Phi' \setminus \Phi \) as the collection of unit vectors \( f(\cdot) \) not contained in \( F \), and let \( G' \) be any group on this set.

These results are essential for what we want to show in this paper: It is possible to establish a relationship between the concepts of theoretical statistics, based on unknown parameters in some parameter space, and the concepts of quantum theory, based on state vectors/rays in a Hilbert space. When it comes to interpretation, it seems that the parameter concept is easier to understand in most cases, even though physicists over the time have developed much intuition related to the Hilbert space paradigm.

The construction of this Section is closely related to the theory of coherent states, which go back to Schrödinger and von Neumann, and which has been developed and applied in great generality recently, see Perelomov (1986). The simplest application is to the Heisenberg-Weyl group related to the coordinate and moment operators. Another relatively simple application is to the rotation group of three-dimensional space; a special case of this has already been discussed in Section 2.

In the above discussion we have for simplicity assumed that the basic group \( G \) is exact. In general, non-exact groups can be handled by defining the isotropy group \( G_0 \) as the set of \( g' \)'s such that \( f_0(g\phi) = f_0(\phi) \) (modulo a phase factor), and then replacing \( G \) by \( \hat{G}/G_0 \) in the definitions above. Typically, this option will be needed when we go to concrete applications.
6 Experiments.

Consider first the traditional statistical setting where one concentrates on one single experiment. Let $x$ be some random variable of this experiment, and let $G^x$ be the group on the sample space corresponding to $G$ on the parameter space, so that $E^h(h(x)) = E^{g^x}(h(g^x x))$ for any function $h$. This gives a natural homomorphism from $G^x$ to $G$. Most of the physical literature concentrates on ideal experiments where one disregards experimental uncertainty, and thus the result $x$ of the experiment defines new values of the parameter, a new state of the system. For such systems we can thus disregard measurement values $x$ and concentrate on parameters and hyperparameters. An important special case is when the experiment just consists of measuring a parametric function $\theta(x)$. For the most part we can think of such highly informative experiments in this paper. To put things into a proper setting, however, we will briefly sketch some points from standard statistical theory, concentrating on the situation with an underlying symmetry group.

One very common purpose of an experiment is to estimate the unknown parameter value $\theta(x)$, say by some function of the data $\hat{\theta}(x)$. There is a considerable theory on how to choose $\hat{\theta}$ in some optimal way, see Lehmann (1983). The more formal approach to this problem is to define a loss function $l(\hat{\theta}(x); \theta(x))$, which gives the loss expressed in some way depending upon the situation when the estimator $\hat{\theta}$ is chosen, and then minimize it.

It was shown in Helland (1998b) that this problem was considerably simplified when $\theta(x)$ was a permissible parametric function, in particular certain classical paradoxes were avoided. Then it is also natural to concentrate on permissible estimators, satisfying the requirement that $\hat{\theta}(x_1) = \hat{\theta}(x_2)$ implies $\hat{\theta}(g^x x_1) = \hat{\theta}(g^x x_2)$ for all $g^x$.

Let $l$ be a loss function satisfying the invariance requirement

$$
l(\hat{\theta}(x); \theta(x)) = l(\hat{\theta}(g^x x); \theta(g x))$$

for all $g^x$. (This is equivalent to $l(\hat{g} a, \hat{g} b) = l(a, b)$, where $\hat{g}$ is the group element induced by $g$ on the image space of the permissible functions $\theta(\cdot)$ and $\hat{\theta}(\cdot)$.) Then, as in Helland (1998b), Corollary 9, we have under Haar prior:

$$
E^h(l(\hat{\theta}(x); \theta(x))) = \int l(\hat{\theta}(x); \theta(x))p^\theta(x)\nu(d\phi),
$$

(4)

where $\nu$ is Haar measure, and the probability density $p^\theta(x)$ is normalized so that $\int p^\theta(x)\nu(d\phi) = 1$. The right-hand side of (4) is independent of $x$. If the data group and/or the parameter is nontransitive (superselection variables), the equation is taken conditionally on each orbit.
The consequence of this identity is that one can use standard tools to minimize the right-hand side of (4). If \( l = \| \hat{\theta}(x) - \theta(\phi) \|^2 \) the solution is simply

\[
\hat{\theta}(x) = \int \theta(\phi) p^\phi(x) \nu(d\phi) = \int \theta q^\phi(x) \nu(d\phi),
\]

(5)

where \( p^\phi(x) = q^{\theta(\phi)}(x) \) and \( \nu \) is Haar measure on the image space of \( \theta \). This is thus the minimum loss solution, and it is also the so-called Bayes solution under Haar prior.

In other cases, one is interested in conclusions to the effect that the unknown parameter \( \theta(\phi) \) belongs to a certain interval. When \( \theta(\cdot) \) is permissible, the two basic such interval types, confidence intervals and credibility intervals (again under Haar prior) coincide.

We will from now on assume that the experiment performed is ideal, in the sense that we arrive at a conclusion of the form \( \theta(\phi) \in C \), and that we consider this conclusion as certain. We may also consider statements of this form when they are not results of experiments, say, hypothetical statements in a discussion, or conclusions arrived at from other types of information.

7 Parametric functions and complementarity.

In classical statistics, there is no difference between hyperparameter and parameter. In quantum mechanics, however, the situation will be that it usually in principle is impossible to get full information about \( \phi \in \Phi \): The only thing one can do is to perform one out of several (ideal) experiments, each giving only partial information.

When \( \Phi \) and the group \( G \) are defined, the next step is to look at the parametric functions \( \theta_a(\cdot) \) from \( \Phi \) to the parameter space \( \Theta_a \) of the single experiments. The simplest case is when all these parametric functions are permissible. Then by Theorem 1 we have a total set of parameters \( \theta = (\theta_a; a \in A) \) such that each \( \theta_a \) is in one-to-one relation to an invariant subspace of \( \mathcal{H} = L^2(\Phi, \nu) \). These invariant spaces will in general be overlapping. The total parameter \( \theta \) will then also be permissible.

Now to a more quantummechanic-like situation: Again we will consider ideal experiments corresponding to parametric functions \( \theta_a(\cdot) \), but typically, these are not permissible with respect to the original group \( G \). As explained in Helland (1998a) such cases occur in the macroscopic world, too.

Again it is useful to consider the ordering of parametric functions: \( \theta^{(1)}(\cdot) < \theta^{(2)}(\cdot) \) if \( \theta^{(1)}(\phi) = \psi(\theta^{(2)}(\phi)) \) for some \( \psi \).
Definition 1.
A frame is defined as any collection of parametric functions \( \{ \theta_a(\cdot); a \in A \} \) such that the mapping \( \phi \mapsto (\theta_a(\phi); a \in A) \) can be inverted.

Since it will typically not be possible to perform experiments corresponding to all \( a \in A \), it does not follow from this that the hyperparameter \( \phi \) can be recovered from experiments.

Typically, in a frame, all (or some) \( \theta_a(\cdot) \) are nonpermissible with respect to \( G \). Thus we need to know how to tackle such functions. We choose to introduce the weaker concept of consistency.

Definition 2.
(a) The frame \( \{ \theta_a(\cdot); a \in A \} \) is consistent if each \( \theta_a(\cdot) \) is permissible with respect to a group \( G_a \) such that \( G \) is contained in the smallest group containing all \( G_a \).

(b) The hyperparameter space \( \Phi \) is consistent if it can be connected to an consistent frame.

In this paper, all frames and hyperparameter spaces under consideration will be assumed to be consistent. Furthermore, it will be assumed that the subgroups \( G_a \) have been chosen in a maximal way, which they can.

Proposition 1.
Let \( G_a \) be the set of \( g \in G \) such that for all \( \phi_1, \phi_2 \in \Phi \) we have that \( \theta_a(\phi_1) = \theta_a(\phi_2) \) if and only if \( \theta_a(g\phi_1) = \theta_a(g\phi_2) \). Then \( G_a \) is a subgroup of \( G \), and the maximal one such that \( \theta_a(\cdot) \) is permissible with respect to \( G_a \).

Proof.
\( G_a \) contains the identity. Using the definition with \( \phi_1, \phi_2 \) replaced by \( g_2\phi_1, g_2\phi_2 \), it follows that \( g_1g_2 \in G_a \) when \( g_1 \in G_a \) and \( g_2 \in G_a \). Using the definition with \( \phi_1, \phi_2 \) replaced by \( g^{-1}\phi_1, g^{-1}\phi_2 \), it is clear that it includes inverses. It follows from the construction that it is maximal.

Definition 3.
Given a frame \( \{ \theta_a(\cdot); a \in A \} \), a state space is defined as a set \( \Psi \) of parameters \( \psi = \psi(\phi) \), permissible with respect to \( G \), such that each \( \theta_a \) will be a permissible function of \( \psi \), and such that \( \psi(\cdot) \) is minimal with respect to this property under the ordering \( \preceq \).

The point with Definition 3 is just to reduce the parameter space as much as is possible. If \( \Psi \not= \Phi \), it is completely certain that the points \( \phi \in \Phi \) are impossible to recover from experiments; the most one can hope for, is information on \( \psi = \psi(\phi) \). So in this meaning, only the space \( \Psi \) makes sense, not \( \Phi \). The points \( \psi \in \Psi \) will be called (pure) states. The Hilbert space constructions in Section 5 can be carried out for \( \Psi \) as well. When experiments are performed, the state space may change, due to the minimality requirement in Definition 3.

The essence of all these definitions is that they make precise - in a setting with group symmetry and a set of parametric models - the quantum mechanical concept of complementarity. Different parametric functions may correspond to complementary aspects of reality. To get full information, we need to estimate all these parameters, which may be impossible with the same experimental units. As stressed already by Niels Bohr, this concept may be useful also in daily life and in a variety of sciences. Today, the concept is used for instance in psychology and in the social sciences, but its use in other natural sciences than physics is rather limited.

We develop first the theory for the simplest case, when all the parametric functions are assumed to be permissible with respect to the group \( G \).

8 An ideal experiment: measuring a permissible parameter.

In general, let \( \theta(\cdot) \) be a permissible parametric function on a (hyper)parameter space \( \Phi \), subject to a group \( G \). We may to begin with consider the specific values \( \phi_1 \in \Phi \) as the pure (hyper-)states of the given system; a general hyperstate is then a measure over \( \Phi \). When there is no information about the system, a natural choice of measure is the Haar measure, in other cases the Haar measure over an orbit or over a set of orbits. Note that the Haar measure over an orbit is unique (up to a multiplicative constant); when several orbits are involved, each can be given an arbitrary weight.

As in Section 2, a natural solution for this latter case can be to concentrate the Haar measure on teh orbit for which Fisher information is minimized.

By what has been discussed in Section 5, all this can be translated in a unique way to a set of rays (one-dimensional subspaces) in an invariant subspace of \( \mathcal{H} = \)
$L^2(\Phi, \nu)$. The hyperparameter $\phi_1$ corresponds to the subspace generated by $f_{\phi_1}() = f_0(g_1^{-1} \cdot )$ with $\phi_1 = g_1 \phi_0$. The group elements $g$ on $\Phi$ correspond to the unitary representation operators $U_R(g)$ on $\mathcal{H}$ by equation (3) in Section 4. The set of points in an orbit of $G$ in $\Phi$ correspond to the set of one-dimensional subspaces represented by $\{U_R(g)f_{\phi_1}(\cdot); g \in G\}$. The situation we will have in mind in this paper, is that $\phi_1$ is impossible to estimate directly; only certain functions $\theta(\phi_1)$ are available.

Let now $\theta(\cdot)$ describe the parameter of some particular experiment. Then, by Theorem 1, $\theta(\cdot)$ determines in a unique way a (parametric) invariant subspace $V$ of $\mathcal{H}$. In the same way as in Theorem 2(a), the different values of $\theta_1 = \theta(\phi_1)$ can be associated with different functions of $\theta$ that are unit vectors of $V$ when $V$ is considered as a Hilbert space on its own right. Note that the different Hilbert spaces of this kind are consistent with the original $\mathcal{H}$, since Haar measure $\nu$ on $\Phi$ induces a new Haar measure $\tilde{\nu}$ on the image of the permissible parameter $\theta(\cdot)$ in such a way that
\[
\int \hat{f}(\theta)\tilde{\nu}(d\theta) = \int \hat{f}(\theta(\phi))\nu(d\phi).
\]

We will start by discussing an ideal experiment in the parameter language, which is easiest conceptually, and then afterwards translate it to the vector space language.

Let the hyperstate before the experiment is performed be given by a probability measure $\pi$ on $\Phi$. Let $\theta(\cdot)$ be the target parameter of the experiment, assumed to be permissible.

Suppose that the ideal experiment results in some value $\theta_0$ for the target parameter. Let $\Phi_0$ be a collection of $\phi$-values such that $\theta(\phi) = \theta_0$. Then after the experiment, the basic group is changed to $G_0 = \{g : \theta(g\phi) = \theta(\phi), \forall \phi \in \Phi_0\}$, which is a subgroup of $G$ by Helland (1988b), Proposition 5. The cosets $gG_0$ may be identified with the elements $\tilde{g}$ of the group $\tilde{G}$ of transformation of $\theta$-values introduced earlier.

More generally, if $\Phi_0 = \{\phi : \theta(\phi) = \theta_0\}$ and $G_0 = \{h : \theta(h\phi) = \theta(\phi), \forall \phi \in \Phi_0\}$, with $\theta = \theta_0$, then $\Phi_0 = g\Phi_0$ and $G_0 = gG_0g^{-1}$ (which is the same for all $g$ that are mapped into the same $\tilde{g} \in \tilde{G}$). The sets $\Phi_0$ and also the cosets $gG_0$ will be disjoint for different $g$, but the groups $G_0$ will overlap. The intersection of all $G_0$ will be a normal subgroup $K$ of $G$.

An important observation from the above results is:

**Proposition 2.**

(a) Let the right and left Haar measure be equal. (For instance, the group could be compact.) Then the posteriori Haar measure $\nu'$ for $\Phi_0$ with respect to $G_0$ is independent of $\tilde{g}$, and therefore of the value of $\theta$. 

18
(b) As a consequence, with $\theta$ as above, such that $\phi = (\theta, \phi')$ with $\phi' \in \Phi_0$, we have, \[ \nu(d\phi) = \nu(d\theta)\nu'(d\phi') \] after suitable normalization of the Haar measure on $G$, $\tilde{G}$ and $G_0$, respectively.

Thus the setting for new measurements is independent of the value of previous measurements. This result is crucially dependent upon the assumption that the parametric function $\theta(\cdot)$ is permissible.

The fact that the analysis of consecutive measurements is relatively simple for permissible ideal measurements, can also be related to the following observation, already mentioned: If $\theta_1(\cdot)$ and $\theta_2(\cdot)$ both are permissible, then the compound parameter $\theta(\cdot) = (\theta_1(\cdot), \theta_2(\cdot))$ is also permissible. Thus results of the type above also apply for the compound parameter. The situation after the $\theta(\cdot)$-measurement is the same whether $\theta_1(\cdot)$ is measured first or $\theta_2(\cdot)$ is measured first. A similar observation holds for more than two measurements. However, the probability distribution of $\theta_1$ will in general not be the same whether $\theta_2$ has been measured or not. In the ordinary language of quantum mechanics, we may have non-commuting variables also in this case.

Here is a tentative definition of a state of a system: It contains two elements:

1. A permissible parametric function $\theta(\cdot)$ giving the focus of interest, or the quantity we choose to and are able to get information on.

2. A probability measure $\hat{\pi}$ on the range of $\theta(\cdot)$. A special case is the complete specification $\theta(\hat{\phi}) = \theta_1$.

Now we will try to translate this to the vector space language. This may seem like a rather meaningless exercise when we already have simple formulae for everything of interest. But the point is that we want to make a link to the formal apparatus of quantum theory. By doing that in the present simple situation, we will also be able to transform this link to more complicated situations.

As the choice of a basic Hilbert space we take $\mathcal{H} = L^2(\Phi, \nu)$.

We intend to focus on states, not on hyperstates. Then the relevant functions are those projected upon the space $V$ as in Theorem 1.

In particular, this means that the parameter space is changed from $\Phi$ to $\Theta$. Mathematically, it will be important now that the underlying function $\theta(\cdot)$ is permissible. It is this property which implies that is is possible to transfer a group symmetry from the first space to the second space.

The conditional expectation is given by the projection from $\mathcal{H} = L^2(\Phi, \nu)$ to the invariant space $V$ of functions $\hat{f}(\theta(\cdot))$ (cf. Section 4). This projection operator has the following mathematical description: It is the operator $P^\theta(\cdot)$ on $\mathcal{H}$ defined by

$$ P^\theta(\cdot) f(\hat{\phi}) = \hat{f}(\theta(\hat{\phi})), $$

19
where $\tilde{f}(\theta)$ is the Radon-Nikodym derivative of $f(\phi)\nu(d\phi)$ with respect to $\tilde{\nu}(d\theta)$, the Haar measure on the image space $\Theta$ of $\theta(\cdot)$. Explicitly, $\tilde{f}(\theta) = \int_{\Phi_0} f(\phi)\nu(\phi)\nu(d\phi)$, since $\nu(\phi) = \tilde{\nu}(d\theta)\nu(d\phi)$ (assuming $\nu(\phi) = 1$). Thus $P^\theta(\cdot)$ smooths $f$ over each single set $\Phi_0$ ($\theta \in \Theta$), and is the correct operator if we want to change the parameter from $\phi$ to $\theta$. In more statistical language $P^\theta(\cdot)$ is the conditional expectation of $f(\cdot)$, given $\theta(\cdot)$, looked upon as random variables.

**Theorem 3.**

Assume Haar measure on $\Phi$ such that $\nu(\Phi_0) = 1$. Then $P^\theta(\cdot)$ as described above is a projection operator, and projects upon the space $V$ of $H$ corresponding to the permissible parameter $\theta(\cdot)$.

**Proof.**

Let $f(\phi)\nu(d\phi) = \tilde{f}(\theta(\phi))\tilde{\nu}(d\theta)$. Then for any integrable function $c(\cdot)$ on $\Theta$ we have

$$\int c(\theta(\phi))f(\phi)\nu(d\phi) = \int c(\theta)\tilde{f}(\theta)\tilde{\nu}(d\theta).$$

Use this identity on $c(\theta) = g(\theta)^* - \tilde{f}(\theta)^*$ to find after some manipulation

$$\int |f(\phi) - g(\theta(\phi))|^2\nu(d\phi) - \int |f(\phi) - \tilde{f}(\theta(\phi))|^2\nu(d\phi) = \int |g(\theta) - \tilde{f}(\theta)|^2\tilde{\nu}(d\theta).$$

This shows that

$$\int |f(\phi) - \tilde{f}(\theta(\phi))|^2\nu(d\phi) \leq \int |g(\theta) - \tilde{f}(\theta)|^2\tilde{\nu}(d\theta)$$

for any $g$, so $P^\theta f(\cdot) = \tilde{f}(\theta(\cdot))$ is the projection of $f$ upon $V$.

Recall that we consider a situation where the hyperparameter $\phi$ is unobservable, while certain parametric functions $\theta(\cdot)$ are observable. Also recall that $f_{\theta_1}(\cdot) = f_{\phi_0}(g^{-1}\cdot)$ for a fixed $f_{\phi_0}$. For $\theta_0 = \theta(\phi_0)$ it is therefore natural to define

$$\tilde{f}_{\theta_0}(\theta(\phi)) = cf^\theta(\cdot)f_{\theta_0}(\phi)$$

with $P^\theta(\cdot)$ as in Theorem 3 and with $c$ such that $\tilde{f}_{\theta_0}^\dagger f_{\theta_0} = 1$. Let $U_R(\cdot)$ be the regular representation of $G$ induced by $U_R$, such that $U_R(g)\tilde{f}(\theta) = \tilde{f}(g^{-1}\theta)$. Taking $\tilde{f}_{\theta_0} = U_R(g)\tilde{f}_{\theta_0}$ for $\theta_0 = g\theta_0$ in a similar way as in (3), we get in a unique way $\tilde{f}_{\theta_1} = cP^\theta(\cdot)f_{\theta_1}$ for $\theta_1 = \theta(\phi_1)$.

Note that the mapping $g \rightarrow \tilde{g}$ is a homomorphism, so $g \rightarrow \tilde{U}(\tilde{g})$ also gives a representation (on $V$) of the basic group $G$. 

20
9 Towards a quantum theory for the simplest case: an ideal permissible experiment.

In accordance with ordinary quantum mechanics, it is natural to define the density 'matrix' \( \rho \) for the mixed state corresponding to a probability measure \( \hat{\pi}(\cdot) \) over \( \Theta \) as

\[
\rho = \int \hat{f}_{\theta_1} \hat{\tilde{f}}_{\theta_1} \hat{\pi}(d\theta_1), \quad \text{i.e.,} \quad \rho(\phi, \phi') = \int \hat{f}_{\theta_1}(\theta(\phi)) \hat{\tilde{f}}_{\theta_1}(\theta(\phi')) * \hat{\pi}(d\theta_1).
\]

This is consistent with the fact that \( \hat{f}_{\theta_1} \) is a pure state vector corresponding to the parameter value \( \theta_1 \). An objection might be that these vectors are not orthogonal in general, so the corresponding projectors do not usually commute. Nevertheless, the integral is well defined and gives a trace 1 positive definite operator \( \rho \) as it should. By choosing new basis vectors that are orthogonal, it can be expressed in a more ordinary form. Similar constructions are made in the theory of coherent states; see Perelomov (1986).

Even though there thus is a parallel to ordinary quantum theory here, and more parallels will turn up later, there is also a formal difference: The definition of \( \rho \) depends crucially upon the choice of the function \( \hat{f}_{\theta_0} \) such that \( \hat{f}_{\theta_1} = \hat{U}_R(\hat{g}_1) \hat{f}_{\theta_0} \), so this function must be held fixed. More precisely, for the formalism that we now are going to develop, one should fix once and for all a function \( f_0(\phi) = f_{\theta_0}(\phi) \), and then project it upon the various spaces corresponding to permissible parameters as described above.

The next step is to try to find an operator corresponding to a new parameter of interest \( \eta = \eta(\phi) \). A first attempt might be to define the 'diagonal' operator \( A_{\eta} \) by

\[
A_{\eta} \hat{f}_{\theta_1} = \mathbb{E}(\eta(\phi)|\theta = \theta_1) \hat{f}_{\theta_1}.
\]

Here a problem arises, however: In general, this will not be well defined, again since the \( \hat{f}_{\theta_0} \)'s are non-orthogonal. However, what we can do, is to define \( A_{\eta} \) such that

\[
A_{\eta} \hat{f}_{\theta_1} = \mathbb{E}(\eta(\phi)|\theta = \theta_1) \hat{f}_{\theta_1} + \epsilon(\theta) \hat{f}_{\theta_1} \epsilon(\theta) = 0,
\]

where the conditional expectation assumes Haar measure on each \( \Phi_0 \).

A far more general result is given by the following theorem, which probably also can be generalized beyond the case of a compact group.

Theorem 4.

(a) Let \( V \) be a subspace of \( L^2(\Phi, \nu) \). Fix \( q \in V \) and \( f_0 \in V \). Let \( U \) be a unitary representation of the compact group \( G \) on \( V \), so that \( V \) is invariant under \( U \). Then there is a unique operator \( A \) on \( V \) such that

\[
(U(g)f_0)^\dagger AU(g)f_0 = U(g)q
\]

(7)
for all $g \in G$.

(b) If $q = q^V$ and $f_0 = f_0^V$ on $V$ are taken as projections of basic functions, the same $A$ can be used for all $V$ that are invariant under the representation $U$.

Proof
(a) is proved (in the $L^2$ version) for irreducible $V$ in Appendix 2. Since $G$ is compact, every unitary representation is reducible. Use (a) on each irreducible part, and use a basis for which $U(g)$ is block diagonal. Then $A$ can also be taken as block diagonal, and we use the fact that all $V$ can be written as an orthogonal direct sum of irreducible spaces. This proves (b).

The obvious application of Theorem 4 is to the case where $V$ is the invariant space corresponding to a parametric function $\theta(\cdot)$, and where $U(g)f_0 = f_{\theta_1}$ with $\theta_1 = g\theta_0$. Let $\eta(\phi)$ be some general function of the hyperparameter. Then it follows from Theorem 4 that there exists an operator $A_\eta$ such that

$$f_{\theta_1} \hat{A}_\eta f_{\theta_1} = E(\eta(\phi)|\theta = \theta_1)$$

for all $\theta_1$, where the expectation on the right hand side is with respect to Haar measure over the remaining parameters.

For a state defined by the parametric function $\theta(\cdot)$ and a probability distribution $\hat{\pi}(\cdot)$ over $\theta$, hence by a density operator $\rho$, given by (6), we have

$$\int E(\eta(\theta)\hat{\pi}(d\theta) = \text{tr}(A_\eta \rho).$$

It was said above that the state vector $\rho$ was only defined uniquely modulo a fixed function. Looking at how the operator $Q$ depends upon the choice of $v$ in Theorem A1, we get a similar arbitrariness in the definition of $A_\eta$ in each irreducible space. The result is that $\text{tr}(A_\eta \rho)$ is uniquely defined, first proved if the space is irreducible, then in general.

The formula (8)-(9) are close to the basic ones in quantum mechanics, from which very many results can be derived, as demonstrated in several textbooks. Some of the most obvious consequences are:

1. In the pure state case, if $A_\eta$ can be made diagonal with respect to some countable orthogonal basis system, then the conditional expectation of $\eta(\phi)$, and of any function of $\eta(\phi)$ can be given as if the probability of some eigenvector was the squared norm of the component of the state vector along the corresponding eigenvector.
2. If $f_{\theta_1}$ is an eigenvector for $A_\eta$ corresponding to eigenvalue $\lambda$, then the conditional expectation of this state is $\lambda$, and the conditional variance is 0, so $\eta = \lambda$ almost surely in the state $f_{\theta_1}$. This must be taken quite literally, since all we know in this state is the value $\theta_1$; any knowledge of the underlying $\phi$ is assumed to be absolutely impossible.

The formula (9) can in general be derived from, and is hence equivalent to, a special case of it, where expectation here is replaced by probability. The corresponding quantum theory statement is considered to be an independent axiom of the theory, as recently pointed out by Deutsch (1999). Deutsch aims at deriving it from the other postulates of quantum theory, but his arguments are countered by Finkelstein (1999). In our approach this formula is automatically satisfied by the construction of the operator involved.

The same formula holds for expressing approximate expectation of random variables, and this variant of the formula seems to be potentially very useful for statistical applications, in particular when we need to approximate models with many parameters.

Corollary 4.

Define the approximate expectation for a random variable $X$, with a distribution depending upon $\phi \in \Phi$ when $\theta(\phi) = \theta_1$ as

$$\hat{E}_{\theta_1}(X) = \int_{f_{\theta_1}} F_\phi(X) \nu'(d\phi').$$

Then there exists an operator $A_X$, independent of all parameters, such that

$$\hat{E}_{\theta_1}(X) = f_{\theta_1}^\dagger A_X f_{\theta_1}$$

for all $\theta_1$; or, if there is a prior $\hat{\pi}(\cdot)$ on $\Theta$, then

$$\hat{E}(X) = \int F_\phi(X) \nu'(d\phi') \hat{\pi}(d\theta) = \text{tr}(A_X \rho),$$

where $\rho$ is given by equation (6). It is assumed that the measures $\nu'$ and $\hat{\pi}$ are normalized to 1.

We hope to discuss statistical applications of the formula (10) elsewhere.

Now we turn to the change of state by an (ideal) measurement. The simplest case is when this just means a change of measure $\hat{\pi}(d\theta_1)$ (from a priori to posteriori). Then one just changes $\rho$ in equation (6) accordingly, and (7) continues to hold.
This is the ordinary Bayesian solution in statistics. But the formalism of quantum mechanics points at another solution, which roughly can be described as follows: When a new parameter $\theta$ has been measured, then the values of all related parameters are forgotten, only the value $\theta$ is kept. In statistical terms, what we do, is the following: When a measurement of a parameter $\theta(\cdot)$ is being done, every Hilbert space vector in $L^2(\Phi, \nu)$ is replaced by its conditional expectation, given $\theta(\cdot)$. Furthermore, a particular conditional expectation is taken, namely the one where all previous information about the other parameters is ignored; their distribution is replaced by Haar measure. Physically and statistically this must mean that the measurement process somehow destroys completely this information.

In this Section we will concentrate on a transition from the hyperparameter space $\Phi$ to the parameter space $\Theta$ with $\theta = \theta(\phi)$. In the next section a more general situation is discussed. After the measurement we can now forget the old parameter space $\Phi$, and replace it by the new one, $\Theta$. In particular the old basic functions $f_{\phi_1}$ are transferred to

$$P_{\theta_1}(\phi) = \int_{\phi_0} f_{\phi_1}(\phi) \nu'(d\phi) = \int_{\phi_0} f_{\phi_1}(\phi) \nu'(d\phi_1) = \tilde{f}_{\theta_1}(\theta(\phi)), \quad (11)$$

where $\theta_1 = \theta(\phi_1)$. The equality between integrals here follows from the following result (cp. also Hora and Buehler, 1966 and Helland, 1998b, Theorem 7).

**Lemma 1.**

We have $f_{\phi_1}(g^{-1}\phi) = f_{g\phi_1}(\phi)$ for all $\phi_1, \phi, g$. When $\phi_1, \phi \in \Phi_{\theta_1}$, then we have $\int_{\phi_{\theta_1}} f_{\phi_1}(\phi) \nu'(d\phi) = \int_{\phi_{\theta_1}} f_{\phi_1}(\phi) \nu'(d\phi_1)$, where $\nu'$ is Haar measure on $\Phi_{\theta_1}$. In particular, this means that the integral is independent of $\phi_1$ when $\phi_1 \in \Phi_{\theta_1}$, and independent of $\phi$ when $\phi \in \Phi_{\theta_1}$.

**Proof.**

The identity $f_{\phi_1}(g^{-1}\phi) = f_{g\phi_1}(\phi)$ follows from equation (3) in Section 5. On compact sets, left and right Haar measures are equal, so

$$\int_{\phi_{\theta_1}} f_{\phi_1}(\phi) \nu'(d\phi) = \int_{G_{\theta_1}} f_{g^{-1}\phi_1}(g^{-1}\phi) \nu'(dg) = \int_{G_{\theta_1}} f_{g\phi_1}(g\phi) \nu'(dg) = \int_{\phi_{\theta_1}} f_{\phi_1}(\phi) \nu'(d\phi_1).$$

Let for fixed $\theta_1$, $P_{\theta_1} f$ be given by $E\{f(\phi)|\theta(\phi) = \theta_1\}$, where, as in Theorem 3 and Theorem 4, the probability measure on $\Phi_{\theta_1} = \{\phi: \theta(\phi) = \theta_1\}$ is given by Haar

24
measure. Similarly, for $C \subset \Theta$, let $P^C_\theta$ be the operator which according to Theorem 4 is given by $f^t_{\theta_1} P^C_\theta f_{\theta_1} = 1$ if $\theta(\phi_1) \in C$, otherwise 0.

**Theorem 5.**
Assume that Haar measure is used on each $\Phi_\theta$. Then $P[\theta \in C] = \text{tr}(\rho P^C_\theta)$ in the state given by (6), and also in the hyperstate given by $\rho = \int f_{\theta_1} f^t_{\theta_1} \pi(d\theta_1)$.

If $f$ is an integrable function, one can define $A_{f(\cdot)} = \int f(\phi) dP_\theta$. Then $E(f(\phi)) = \text{tr}(\rho A_{f(\cdot)})$ when the state is $\rho$. The operator $A_{f(\cdot)}$ will be hermitean if $f$ is real, unitary if $f(\cdot)^* f(\cdot) = 1$.

**Proof:**
We can write $P^C_\theta = Q^C_\theta P^{\theta(\cdot)}$, where $\int f_{\theta_1} Q^C_\theta f_{\theta_1} = 1$ if $\theta_1 \in C$, otherwise 0 (cp. Theorem 4). We then have

$$\text{tr}(\rho P^C_\theta) = \text{tr} \int f_{\theta_1} f^t_{\theta_1} P^C_\theta \pi(d\theta_1) = \int f_{\theta_1} Q^C_\theta f_{\theta_1} \pi(d\theta_1) = \pi(C).$$

The proof for the hyperstate case is similar.

The last part follows by a suitable limit argument.

**10 Several measurements/ observables.**

If a measurement of $\theta_a(\cdot)$ is made, say with result which can be described by a measure $\pi(d\theta_a)$ on $\theta_a(\cdot)$, then the new state $\rho_a$ is given by

$$\rho_a = \int f_{\theta_1}(\theta_a(\phi)) f^t_{\theta_1}(\theta_a(\phi')) \pi(d\theta_1).$$

We assume in the same way as before that $f_{\theta_1}(\theta_a(\phi)) = Q_{\theta_1} P^{\theta_a(\cdot)}$ for an operator $P^{\theta_a(\cdot)}$ in $\mathcal{H} = L^2(\Phi, \nu)$, and for $\theta_1 = \theta_a(\phi_1)$, but we do not assume that it is possible to estimate the base parameter $\phi_1$.

The question is now what happens if we measure the variable $\theta_b(\cdot)$ from this last state. Let $P^C_\theta$ be an operator such that $f^t_{\theta_1} P^C_\theta f_{\theta_1} = 1$ if $\theta_b(\phi_1) \in C$, otherwise 0. Furthermore, let $P^{\theta_b(\cdot)} f(\phi) = \hat{f}(\theta_b(\phi))$, so $P^C_\theta = Q^C_\theta P^{\theta_b(\cdot)}$. Then, from the usual postulates of quantum mechanics it follows that

(A) $P[\theta_b \in C] = \text{tr}(\rho_a P^C_\theta)$ in state $\rho_a$.
(B) After having observed $\theta_b$ from state $\rho_a$ with the result $\theta_b \in C$, the new state will be

$$
\rho'_b = \frac{P_{\theta_b}^C \rho_a P_{\theta_b}^C}{\text{tr}(P_{\theta_b}^C \rho_a)}.
$$

(13)

We will show that these formulae also follow from a parametric point of view if we look at the whole (hypothetical) sequence of observations: First $\theta_a(\cdot)$, then $\theta_b(\cdot)$, and by tracing the formula (12) for $\rho_a$ back to the hyperparameter $\phi_1$:

$$
\rho_a = P^{\phi_a(\cdot)} \rho P^{\phi_a(\cdot)} \text{ with } \rho = \int f_{\phi_1} f_{\theta_a} \pi(d\phi_1).
$$

This means that the formula for $P[\theta_b \in C]$ follows from Theorem 5 if we there take

$$
P_{\theta_b}^C = P_{\theta_b}^C P^{\phi_a(\cdot)} = Q_{\theta_b}^C P^{\phi_a(\cdot)}.
$$

Similarly we have

**Theorem 6.**

*Suppose that the only information we have after measurement of a permissible parametric function $\theta_b(\cdot)$ is that $\theta_b(\phi_1) \in C$ for some set $C$. Then the state is changed from $\rho_a$ to

$$
\rho'_b = \frac{P_{\theta_b}^C \rho_a P_{\theta_b}^C}{\text{tr}(P_{\theta_b}^C \rho_a)},
$$

where $P_{\theta_b}^C$ is defined above. In particular, if the measurement results in $\theta_b(\phi_1) = \theta_1$, then the state after measurement is $\rho'_b = P_{\theta_b} \rho_a P_{\theta_1}/\text{tr}(P_{\theta_1} \rho_a)$.*

**Proof.**

With the notation above,

$$
\frac{P_{\theta_b}^C \rho_a P_{\theta_b}^C}{\text{tr}(P_{\theta_b}^C \rho_a)} = P_{\theta_b} \rho P_{\theta_b}^C P[\theta_b \in C],
$$

which shows that the proof can be reduced to the case when $\rho$ is a hyperstate $\rho = \int f_{\phi_1} f_{\theta_a} \pi(d\phi_1)$, which we will assume from now on. Then

$$
\frac{P_{\theta_b}^C \rho P_{\theta_b}^C}{P[\theta_b \in C]} = \frac{\int f_{\phi_1(\cdot)} \rho_{\theta_a} P_{\theta_b}^C \pi_{\theta_b}(d\theta_b) \pi_{\theta_a}(d\phi_a)}{P[\theta_b \in C]} = \frac{\int f_{\phi_1(\cdot)} \rho_{\theta_b} P_{\theta_b}^C \pi_{\theta_b}(d\theta_b)}{P[\theta_b \in C]} = \rho_b,
$$

26
When an (ideal) observation is done, any state is changed into a pure state and \( \theta \) is completely determined. With permissible parameters, measurements on several different parameters can be done after each other, each leading to a determined parametric function. Typically then, after a while the complete original state/parameter \( \phi \) is determined. We shall see that the situation with non-permissible parameters may be completely different.

11 Non-permissible parameters; some preliminaries.

The setting is as before: There is a hyperparameter space \( \Phi \) representing all unknown quantities in a given system, a group \( G \) acting on \( \Phi \), a Haar measure \( \nu \) on \( \Phi \) and the basic Hilbert space \( \mathcal{H} = L^2(\Phi, \nu) \). We will allow the hyperparameter space to have a rather arbitrary structure, and then, if needed, introduce the necessary Hilbert spaces through the symmetry of the situation and the implied group representation theory.

The general natural procedure when \( \theta(\cdot) \) is non-permissible, is to reduce the group. It follows from Proposition 1 in Section 7 that this can always be done in such a way that \( \theta(\phi) \) becomes permissible. However, when several parameters are involved, the order in which we do this, matters. Physically (or statistically) this implies that the order in which experiments are performed, matter. (It should be remarked, however, that also for permissible parametric functions the measuring process may fail to commute.)

The following technical result will give the background needed for a proper discussion of non-permissible parameters.

**Lemma 2.**

Let \( \{U(g) : g \in G\} \) be a unitary representation of \( G \) on a vector space \( V \subseteq \mathcal{H} \). Let \( V_1 \) be a subspace of \( V \), and define \( G_1 = \{g \in G : U(g)v \in V_1 \text{ when } v \in V_1\} \). Then \( G_1 \) is a subgroup of \( G \), and \( \{U(g) : g \in G_1\} \) is a representation of \( G_1 \) on \( V_1 \).

**Proof.**

It is obvious that \( g_1g_2 \in G_1 \) when \( g_1 \in G_1 \) and \( g_2 \in G_1 \), since \( U(g_1g_2) = U(g_1)U(g_2) \). Let \( g \in G_1 \). Note that \( V_1 \) is an invariant space for the set of unitary operators \( \{U(g) : g \in G_1\} \), and so we also have that \( V_1^\perp \) is an invariant subspace. Suppose that \( U(g^{-1}) \) has a component \( v_2 \neq 0 \) in \( V_1^\perp \) for some \( v \in V_1 \). Then
\[ U(g)^{-1}v = v_1 + v_2 \text{ with } v_1 \in V_1 \text{ and } v_2 \in V_1^\perp, \text{ and } v = U(g)v_1 + U(g)v_2. \] This last equation implies \( U(g)v_2 \in V_1 \), which gives a contradiction to the assumption that \( v_2 \neq 0 \) since \( V_1^\perp \) is invariant. Therefore: \( U(g^{-1})v \in V_1 \) for \( v \in V_1 \), and \( g^{-1} \in G_1 \) for \( g \in G_1 \). Thus \( G_1 \) is a group.

It is clear that \( \{U(g) ; g \in G_1\} \) gives a representation of \( G_1 \) on \( V_1 \).

**Proposition 3.**

Let \( \{U(g) ; g \in G_1\} \) be a unitary representation of \( G \) on a vector space \( V \subseteq \mathcal{H} \).

Let \( \dim(V) = k \) (possibly infinity).

Let \( W_1 \oplus W_2 \oplus \ldots \) be an arbitrary decomposition of \( V \) in orthogonal subspaces. Put \( V_j = W_1 \oplus \ldots \oplus W_j \) for \( 1 \leq j < k \), and let \( G_j = \{g \in G : U(g)v \in V_j \text{ when } v \in V_j\} \).

Then:

(a) \( G_1 \subseteq G_2 \subseteq \ldots \subseteq G \), and all \( G_j \) are subgroups of \( G \).
(b) \( \{U(g) ; g \in G_j\} \) is a representation of \( G_j \) upon \( V_j \).
(c) If \( V_j \) is a parametric invariant space \( (j = 1, 2, \ldots) \), corresponding to the permissible (with respect to \( G_j \)) parametric function \( \theta_j(\cdot) \), then \( \theta_1 \preceq \theta_2 \preceq \ldots \preceq \phi \).

**Proof.**

Follows from Lemma 2 and Theorem 1.

This Proposition gives us a way to handle non-permissible parametric functions in concrete cases: Find a subgroup \( G_\alpha \) of \( G \) so that the parameter is permissible with respect to \( G_\alpha \). This is always possible, and in many cases it is easy to find such a \( G_\alpha \). In general we can let \( G_\alpha \) be the maximal group such that \( \theta_\alpha(\cdot) \) is permissible with respect to \( G_\alpha \); cf. Proposition 1.

The procedure is to do all calculations as if this was the relevant group. Let \( V_\alpha \) be the invariant space with respect to \( G_\alpha \) corresponding to the parameter of interest. Use Theorems 3, 4 and 5 with the reduced group.

If another parameter \( \theta_\beta \) is of interest, it might be necessary to choose another group \( G_\beta \). Typically, these two measurement processes will not commute.
12 A possible starting point for a conceptually simpler quantum theory.

We have now collected enough technical tools to explain the relation between what seems to be an intuitively simple theory for quantum phenomena, based on symmetry and statistical modelling concepts, and the traditional one, based on an abstract Hilbert space formulation.

For clarity we repeat again the assumptions made: We let \( \Phi \) be some general hyperparameter space, assumed in principle only to be locally compact, even though compactness was assumed for some of the technical results above. Let \( G \) be a (locally compact) group acting on \( \Phi \). A parametric function \( \theta(\cdot) \) on \( \Phi \) is called permissible if \( \theta(g\phi_1) = \theta(g\phi_2) \) implies \( \theta(g\phi_1) = \theta(g\phi_2) \). Any permissible parametric function induces a group \( G \) on the image space \( \Theta \), i.e., \( \tilde{\theta}(g\phi) = \theta(g\phi) \).

We will assume here that the group \( G \) is transitive; superselection rules can be tackled by taking orbits separately.

Any statistical experiment \( a \) involves a class of measures indexed by some parameter \( \theta_a \). We will assume here that there is some well-defined setting where any one in a fixed class of experiments can be performed, so that each \( \theta_a \) is a function on \( \Phi \).

Basically, only two simple assumptions are needed now:

1. \( \theta_a(\cdot) \) is permissible with respect to some subgroup \( G_a \) of \( G \).

2. During measurement of a parameter \( \theta_a(\cdot) \) the distribution of the remaining parameters is taken as Haar measure with respect to \( G/G_a \).

The simplest case is when the group \( G_a \) is transitive on the sets \( \Phi_{a,\omega} = \{ \phi : \theta_a(\phi) = \theta_0 \} \). If this is not the case, we choose orbit by minimizing Fisher information as in Section 2.

This framework is in fact rich enough to give a very interesting theory for experiments. The theory for each single experiment is contained in existing statistical theory. The theory of a collection of possible experiments like this, seems to be new.

In addition to the assumptions above, it may be convenient to assume, as in Definition 1, that \( \phi \) can be recovered from all the potential values \( \theta_a(\phi) \), and, as in Definition 2, that the basic group \( G \) is contained in the smallest group containing all \( G_a \).

For the technical results in this paper, we have made the further rather strong assumption that \( G \), in addition to being transitive, is exact and compact. These assumptions can certainly be partly removed by going deeper into the mathematics; a start of that process was made in Helland (1998b), using among other things results of Varadarajan (1985). However, the point here is mainly to show that a connection
can be made, at least in the simplest case. Note that the new theory in itself does not need these strong assumptions to make sense, but assumptions are needed to do calculations and to recover the ordinary quantum formulations.

The basic results now are very similar to the results for the permissible case. The main difference is that the group \( G \) must be changed to \( G'_a \), depending upon which measurement is made. Some care must be made in the case of several parameters. If measurements are made in a sequence, then every point in that sequence must be connected to a group for which the involved parametric function is permissible.

Use definitions as in Section 6, but allow the measurement parameter \( \theta(\cdot) \) to be non-permissible. Then:

**Theorem 7.**

Under the above assumptions, let \( \nu \) be the Haar measure of the group \( G \), and let \( \nu'_a \) be right Haar measure of \( G'_a \) on \( \Phi'_\theta_0 = \{ \phi_1 : \theta_a(\phi_1) = \theta_0 \} \) or on the orbit of \( G'_a \) chosen from this set. Then \( \nu'_a \) can be chosen to be independent of the value \( \theta_0 \), and will be normalized so that \( \nu'_a(\Phi'_\theta_0) = 1 \). Furthermore, the Haar measure for \( G'_a \) between orbits (different \( \theta_0 \)) can be chosen so that it is absolutely continuous with respect to the Haar measure for \( G \).

For \( f \in \mathcal{H} \) let \( P^{\theta_0}(t) = \int_{\Phi'_{\theta_0}} f(\phi)\nu'_a(d\phi) \), the Radon-Nikodym derivative of \( f(\phi)\nu(\phi d\phi) \) with respect to the Haar measure \( \nu(\phi d\phi) \) induced by the group \( G'_a \) on the space \( \Theta_a \). Then \( P^{\theta_0}(t) \) is a projection operator, and projects upon the space \( V_a \) in \( \mathcal{H} \) corresponding to the parameter \( \theta_a(\cdot) \) (assumed to be permissible with respect to the group \( G'_a \)).

**Proof.**

The group \( G \) was at the outset assumed to be transitive and exact on \( \Phi \), but if \( G'_a \) then is a proper subgroup of \( G \), it cannot be transitive. Let \( \Phi_a \) be a fixed orbit of \( G_a \). Then using Proposition 1 on \( \Phi_a \), we have

\[ \nu_a(\phi d\phi) = \nu_a(d\theta_a)\nu'_a(d\phi') \]  

(14)

for the respective Haar measures.

By assumption each \( \Phi'_\theta_0 \) is included in some orbit or a fixed orbit is chosen for each \( \theta_0 \). But between orbits the weight of the Haar measure can be selected freely. Furthermore, since \( G_a \) is a subgroup of \( G \), the two Haar measures must be equal on each orbit; and for the \( G \)-Haar measures we have \( \nu(\phi d\phi) = \nu(d\theta)\nu'(d\phi') \) by Proposition 2(b). Therefore, the \( \nu'_a \) must be proportional on different orbits (different \( \theta_0 \)). Hence weights can be adjusted so that (14) holds throughout with \( \nu'_a \) independent of \( \theta_0 \).
Choose any \( f \in \mathcal{H} \). Since \( \Phi_{\theta_0}^f \) belongs to a fixed orbit \( \Phi_{\theta_1}^f \), then \( P_{\theta_0}^\mathcal{H}(\cdot) f \) can be defined by using Theorem 3 on \( (\Phi_{\theta_1}^f, G_a) \). The properties of this projection operator follows also from Theorem 3. As before, the state \( \rho \) is given by

\[
\rho = \int f_{\theta_0} \int f_{\theta_1} \tilde{\pi}^a (d\theta_1).
\]

The important Theorem 4 is also valid in this setting:

**Theorem 8.**

Assume that the group \( \tilde{G}_a \) induced by \( G_a \) on \( \Theta \) is compact. For any integrable function \( \eta(\cdot) \) on \( \Theta \), we can define an operator \( A_\eta \) such that

\[
\int f_{\theta_1}^\eta A_{\theta_1} f_{\theta_1} = \mathbb{E}(\eta(\cdot)|\theta = \theta_1)
\]

(assuming Haar measure with respect to \( G_a \) on \( \Phi_a \) = \{ \phi : \theta(\phi) = \theta \} \) or the selected orbit) for all parametric functions \( \theta_a(\cdot) \) and every \( \theta_1 = \theta_a(\phi_1) \). Therefore, for any state, defined by a parametric function \( \theta_a(\cdot) \) and a probability distribution \( \tilde{\pi}(\cdot) \) over \( \Theta_a \), hence by a density operator \( \rho \), as given by (15), we have \( \mathbb{E}(\eta(\cdot)|\theta) \tilde{\pi} (d\theta) = \text{tr}(A_\eta \rho) \).

The corresponding Corollary 3 is also true as it stands, and finally, Theorem 5 and Theorem 6 are also valid as long as the group is changed from \( G \) to \( G_a \), respectively \( G_b \).

Thus the basic new thing here compared to the permissible case is that the original group \( G \) is replaced by a new group \( G_a \), the maximal group for which \( \theta_a(\cdot) \) is permissible. This means that each measurement may involve a new group, implying that different measurements as a rule will not commute, and also that information from a sequence of ideal measurements usually will not converge towards full information about the hyperparameter \( \phi_1 \).

This implies that one seems to be closer to the ordinary quantummechanical situation in the non-permissible case than in the permissible case. Further light upon this is shed by the simple example which follows. Note that, in this example it is essential that a state, as given by a parametric function \( \theta(\cdot) \) and a probability measure over its values, implies that further knowledge of the hyperparameter \( \phi_1 \) is excluded.

Note: All this theory has been built upon a one-to-one correspondence between parametric functions \( \theta(\cdot) \) and values \( \theta_1 \) of such functions on the one hand and Hilbert state vectors \( f_{\theta_1}(\theta(\cdot)) \) on the other hand. This means that a pure state is simply a pair \( (\theta(\cdot), \theta_1) \). If knowledge of the hyperparameter \( \phi_1 \) was a possibility, then from a quantummechanical point of view a rather trivial theory would result. On the
other hand, it is this situation which is closest to practically all current statistical applications. One can easily imagine large scale situation, too, where the notion of complementarity is relevant, but it remains to see how useful this way of thinking may be in concrete statistical applications.

In quantum mechanical applications, the parametric functions \( \theta(\cdot) \) seem to be very close to what Bohr and Ulfbeck (1995) call symmetry variables, while saying that these variables constitute the elements of physical reality.

13 A simple large-scale ‘quantum mechanical’ example.

Since we have shown that our approach gives a formalism that is essentially equivalent to the ordinary quantummechanically formalism, any standard example from quantum theory could have been used to illustrate our theory. Instead we will give a large scale example, using a group which is as simple as possible. Examples of more direct physical relevance, using a related symmetry approach, can be found in Bohr and Ulfbeck (1995).

One of the simplest non-commutative groups is the group \( S_3 \) of permutations of 3 objects. Visually, it is the group of possible positions of a solid triangle under the constraint that the corners should touch 3 fixed points on a table, say. It has a two-dimensional representation discussed in many books in group theory and several books in quantum theory. The quantum theory book by Wolbarst (1977) is largely based on this group as a pedagogical example. The famous controversies around Marilyn vos Savant’s example with doors, goats and a car (see Morgan et al., 1991, and references there) can also be coupled to the non-commutativeness of the group \( S_3 \).

We will identify the hyperparameter space \( \Phi \) with the 6 elements of \( G = S_3 \): 1=ABC, 2=CAB, 3=BAC, 4=ACB, 5=CBA and 6=BAC, where we can picture the letters as marking the corners of the triangle, read counter-clockwise with the first letter at the lower lefthand corner. We still think of a solid, physical triangle, and we may assume that it is painted white on one side, say, corresponding to the rotation permutations 1, 2 and 3, and black on the other side, corresponding to the reflections 4, 5 and 6.

In this case the two-dimensional representation of the group is not only a homomorphism, but an isomorphism (Wolbarst, 1977). Hence the permissible parametric function corresponding to the two-dimensional representation (cp. Theorem 1) is just the identity in this case. There are 4 parametric functions of interest: \( \theta_a(\cdot), \theta_b(\cdot), \theta_c(\cdot) \), whose values are the upper case letter at the corners \( a, b \) and \( c \),

32
respectively, and \( \theta_0(\cdot) \), whose value is the colour of the triangle.

**Proposition 4.**

\( \theta_0(\cdot) \) is a permissible parametric function, while \( \theta_a(\cdot) \), \( \theta_b(\cdot) \) and \( \theta_c(\cdot) \) each are non-permissible.

**Proof.**

Let the group element which takes the ground position 1 to the position \( i \) as listed in the hyperparameter list above, be called \( g_i (i = 1, \ldots, 6) \). Assume \( \theta_0(\phi_1) = \theta_0(\phi_2) \), say black, for two hyperparameter values \( \phi_1 \) and \( \phi_2 = 1, 2, \ldots, 6 \). Then by simple inspection \( \theta_0(g_i\phi_1) = \theta_0(g_i\phi_2) = \text{‘black’} \) for \( i = 1, 2, 3 \) and ‘white’ for \( i = 4, 5, 6 \). Hence \( \theta_0 \) is permissible.

For the other functions it is enough to produce a counterexample. Here is one for \( \theta_a \): We have \( \theta_a(1) = \theta_a(4) = A \), but \( \theta_a(g_51) = \theta_a(5) = C \) and \( \theta_a(g_54) = \theta_a(3) = B \).

The maximal reduced group corresponding to \( \theta_a \) will consist of the identity together with a single reflection: \( G_a = \{g_1, g_4\} \), which is transitive on \( \Phi_a = \{1, 4\} \), the set of elements with \( \theta_a = A \). With respect to this group the parametric function \( \theta_a \) will be permissible. The results for \( \theta_b \) and \( \theta_c \) are similar.

So far the terms ‘parameter’ and ‘hyperparameter’ have only been words, but now we intend to construct a large scale quantum system based on this solid triangle. To this end, place the triangle within a hollow nontransparent sphere in such a way that it can rotate freely around its center point, placed at the centre of the sphere. Let there first be 4 windows in the sphere, one at the north pole, where the colour facing up can be observed, and three equidistant windows along the equator, where single corners of the triangle can be observed. We could also let the observations in each window be uncertain for some reason, but for simplicity we will again stick to ideal measurements in each window.

To make this correspond as well as possible to a quantummechanical system, we have to make some perfections on the sphere with its windows and solid triangle inside. We assume that we have some mechanisms to the following ends:

Before any measurements are made, the triangle rotates completely freely around its center point. Once an equatorial window is opened, there is some mechanism which draws the nearest corner to this window. After the measurement at this window is done, this particular corner is fixed, and the triangle rotates around the implied axis until a new equatorial measurement is done. The second measurement implies that the hold in the first corner loosens.
Once the top window is opened, the triangle flattens out to (the nearest) horizontal position, and it continues to rotate in this position until a new measurement is done.

We assume that the windows are ideal in the sense that they never provide more information than they are supposed to. We finally assume that it is impossible to open the sphere.

Note that each 'observable' \( \theta_\alpha, \theta_\beta, \theta_\gamma \) has three different values. In the state where \( \theta_\alpha = A \) has been observed, each of the observables \( \theta_\beta \) and \( \theta_\gamma \) then has a probability of 1/2 on each of the values B and C. If a \( \theta_\theta \)-measurement has been done, the next observation in an equatorial window will be essentially unaffected. Therefore the state of the system after two such observations, one in the top window and one in the horizontal window, is completely determined.

Note also the following feature related to the general theory: If only observation of the permissible observator \( \theta_\theta \) is allowed, this will be a rather trivial system, where every future observation is fixed after just one measurement.

A 'true' quantum mechanical system will first be obtained if we close and seal the top window. Then the triangle is never at rest. Each measurement puts the system into a new state. In particular, the question of retrieving \( \phi \), the 'true' position of the triangle, from a measurement, is out of question.

Standard quantummechanical calculations can be carried out for this example, e.g., by choosing a three-dimensional basis where \( \theta_\alpha(\cdot) \) is diagonal in the sense that it can be associated with a diagonal 3 × 3 matrix \( A_\alpha \). In this basis neither \( \theta_\beta \) nor \( \theta_\gamma \) are diagonal. However, ordinary probability calculations are of course much easier to carry out in this case.

14 Concluding remarks.

In this paper we have confined the discussion to the parameter space of the experiment in question. Taking up again the idea from Section 2 of a set of potential experiments \( \{ \mathcal{X}_\alpha, \mathcal{F}_\alpha, \{ \mathcal{P}_\alpha^\theta, \theta \in \Theta_\alpha \}; a \in \mathcal{A} \} \), a related discussion would result if we introduce a symmetry group on \( \mathcal{A} \). This discussion seems to be related to randomization theory, where group representations and invariant spaces have already proved to be of fundamental value (Bailey, 1991; Bailey and Rowley, 1990). An interesting question is if it is possible to combine these two approaches.

It is of course too early to say anything final about the value of the mathematical results discussed in this paper. However, personally, I feel that there is a possibility that results of this kind can help starting a process which eventually may shed new light upon several aspects in the interpretation of quantum physics and also perhaps
upon more general issues related to modelling, observing and interpreting nature. If I am right in these assumptions, there seems to be room for and need for much further work in several directions, and I think that some of this work may call for cooperation between several disciplines, physicists, mathematicians, statisticians, philosophers, and perhaps also other specialists.

At least on an intuitive level, I feel that there are less philosophical difficulties implied if one should be able to develop a complete parameter interpretation of quantum physics than what is implied by the conventional interpretation. Issues like Schrödinger’s cat become less problematic when everything is governed by models with hypothetical parameters instead of formalisms involving states that are assumed to be ‘real’ in some sense. Similar remarks may help to explain why the Kochen-Specker theorem (Isham, 1995) does not seem to represent a problem here.

However, the discussion around hidden variables is very relevant to the model discussed in this paper. If the present models shall describe something like the real quantum world, then it must be impossible by any means to get full information about $\phi_1$. Only certain derived parameters can be estimated. One question then might be if models involving action at a distance on the hyperparameters can be permitted, since noe information is ever retrieved about these parameters. The discussion of Section 2 is relevant here.

Also, several other problems do exist, for instance: What models should be allowed in quantum physics? What models are reasonable? Are there ways to tackle ambiguity in the model choice? Again there are many details to be worked out, but it is quite obvious that symmetry considerations must be very important in governing model choice in this field.

In this paper I have not said anything about time development and the Hamiltonian. The issue was briefly touched upon in Helland (1998a).

Modelling involves information. It would be very interesting if some connections could be found to classical themes like Maxwell’s demon, to blackbody-radiation or to thermodynamics in general.

At first sight the models of this paper may seem too simple to explain the complicated world of quantum physics. A natural question, for instance, is: Where do we find the superposition of state vectors in these models? In fact, we do find it, for instance in equation (11) above, or more generally, whenever focus is changed from one parameter to another.

Griffiths’(1984) history approach is an important issue in current development of quantum physics. Closer connections here would definitely be of interest to study. Of special interest may be the connection between decoherence in quantum physics and randomization and other ways of deliberately ignoring information in statistics.

A theme that is not touched upon at all here, is generalization to relativistic systems. There are undoubtedly problems here demanding the skills of both physi-
ists and mathematicians. But knowing that large parts of high energy physics is based upon group theory, even the most superficial thinking around this may point at possibilities for some kind of parametric modelling based on symmetry.

Calculation is another problem area that seems to lead to new possibilities if the theories in this paper are accepted. In some future one can think of both statistical techniques (Markov Chain Monte Carlo for instance) that may turn out to be useful in quantummechanical calculations, or one can investigate if in the large resource of highly developed quantummechanical calculating techniques there might be something that is useful for instance in Bayesian statistics.

Finally, again if the ideas here are accepted, there seems to be room for thinking through once again some concepts in statistical theory. But that is not a theme here.

As repeatedly said above, all these remarks assume that the main idea of this paper should be accepted, namely that there is a possibility of a common foundation, at least partly, for the two areas which today seem so far apart: Theoretical statistics and quantum theory. My own feeling is definitely that there seems to be such a possibility, and that the technical arguments in the main body of this paper may give some of the tools needed to find that possibility. Another very recent development which seems to point in the same direction, is the Fisher information approach (Frieden, 1995; see also Matthews, 1999) treating among other things the relation between Cramer-Rao's inequality in statistics and Heisenberg's inequality in quantum physics. The book by Frieden discusses largely problems that have not been touched upon here, however; on the other hand, no discussion of quantum theory is given there except for the treatment of the Schrödinger equation and its relativistic counterparts.

Acknowledgements
I am grateful to Chris Islam and Carsten Lütken for discussions, to Phil Dawid, Tore Schweder and Mervyn Stone for comments and to Lennart Bondesson, Richard Gill and Dag Tjøstheim for encouragements during the work with this paper.

References


**Appendix 1: Some group representation theory for compact groups.**

In general, a matrix representation of a group is defined as a function $D$ from the group to the set of (here complex) matrices satisfying $D(gh) = D(g)D(h)$ for all $g$ and $h$. Any representation $D$ and any fixed nonsingular matrix $K$ of the same size can be used to construct another representation $S(g) = KD(g)K^{-1}$. If the group is compact (and also in some other cases), we can always find such an $S$ of minimal block diagonal form, and at the same time we can take $S$ to be unitary $(S(g))^{*}S(g) = I$. If (and only if) the group is Abelian, each minimal block will be one-dimensional.

An important aspect of this reduction appears if we look upon the matrices as operators on a vector space. Then each collection of blocks gives an invariant vector space, and each single minimal block gives an irreducible invariant vector space. For compact groups, the irreducible invariant vector spaces will be finite-dimensional. The minimal matrices in the blocks are called irreducible representations of the group.

More generally, a class of operators $\{U(g); g \in G\}$ (where $G$ is a group) on a, possibly infinite-dimensional, vector space is a representation if $U(gh) = U(g)U(h)$ for all $g, h$. A representation of a compact group has always a complete reduction in minimal matrix representations as described above. The theory of general groups is more complicated, but one can always talk about (irreducible) invariant spaces and irreducible subrepresentations. In particular, this holds for the unitary regular representation defined on a Hilbert space $H = L^{2}(\Phi, \nu)$ by $U(g)f(\phi) = f(g^{-1}\phi)$. Here $\nu$ is the right Haar measure for $G$ on $\Phi$ (defined by $\nu(Bg) = \nu(B)$ for all $B, g$).

Two useful results are Schur’s lemmas:

1. If $D$ and $D'$ are irreducible representations of different dimensions, and $A$ is such that $D(g)A = AD'(g)$ for all $g$, then necessarily, $A = 0$.

2. If $D$ and $D'$ are irreducible representations of the same dimension, and $A$ is such that $D(g)A = AD'(g)$ for all $g$, then either $D$ and $D'$ are isomorphic or $A = 0$. If $D(g)A = AD(g)$ for all $g$, then necessarily $A = \lambda I$ for some scalar $\lambda$.

The character of a group representation is defined as $\chi(g) = \text{tr}(D(g))$. The characters of the minimal blocks $D_{t}(g)$ are the irreducible characters $\chi_{t}(g)$. These
satisfy the orthogonality relations
\[ \int \chi_r(g)^* \chi_{r'}(g) \nu(dg) = \delta_{r,r'}, \]
where \( \nu \) is the Haar measure of the group (which can be uniquely defined as a probability measure for compact groups).

Let now \( V \) be an invariant vector space of any representation \( D(g) \). The orthogonal projection upon the irreducible part \( V_r \) of \( V \) corresponding to \( D_r(g) \) (dimension \( d_r \)) can then be computed as
\[ P_r = d_r \int \chi_r(g)^* D(g) \nu(dg). \]

A Fourier transform corresponding to the representation \( D(g) \) can be defined by
\[ \hat{f}(D_r) = \int f(g) D_r(g) \nu(dg), \]
and for compact groups it can be inverted by
\[ f(g) = \sum_r d_r \text{tr}(D_r(g^{-1}) \hat{f}(D_r)). \]

Appendix 2: The construction of an operator.

Below we prove Theorem 4(a) in its $l^2$ version. With the easily proved connection between $L^2$ and $l^2$ given by $f(x) = v^tu(x)$, and correspondingly $Qf(x) = (Qv)^tu(x)$ for operators, where $u(x) = (u_1(x), u_2(x), \ldots)$ is a basis for $L^2$, it is a straightforward exercise to show that the following result in $L^2$-version follows from its version in $l^2$.

**Theorem A1.**

Let $\hat{U}(\cdot)$ be a unitary irreducible representation on a vector space $V \subset l^2$ of a compact group $G$, let $v \in V$ ($v \neq 0$), and let $c(\cdot)$ be a function on $G$ which is integrable with respect to Haar measure. Then there exists a unique operator $Q$ such that

$$v^\dagger \hat{U}(g)^\dagger Q\hat{U}(g)v = c(g) \text{ for all } g.$$  \hspace{1cm} (16)

**Proof.**

With Haar measure $\nu$, the Fourier transform of $c(\cdot)$ is

$$\hat{c}(\hat{U}) = \int c(g)\hat{U}(g)\nu(dg),$$

which for a given $Q$ must be

$$I\hat{c}(\hat{U}) = v^\dagger \int \hat{U}(g)^\dagger Q\hat{U}(g)v\hat{U}(g)\nu(dg).$$

This implies

$$\hat{c}(\hat{U}) \cong \lim_{n \to \infty} \int \hat{U}(g)\nu(dg) \cong \hat{U}(g)^\dagger v\hat{U}(g)$$

hence

$$\hat{c}(\hat{U}) = \sum_{i,j} Q_{ij} D_{ij}, \hspace{1cm} (17)$$

where, for each $(i,j)$ the operator $D_{ij}$ is defined by

$$D_{ij} = \int (\hat{U}(g)v)_{i,j}(\hat{U}(g)v)_{i,j}^\dagger \hat{U}(g)\nu(dg).$$

We want to show that this equation has a unique solution. Once this is proved, then, since Fourier transforms over irreducible $\hat{U}(\cdot)$ can be uniquely inverted, we also know that we will get a unique solution of (16). To show existence and uniqueness of (17), the following observation simplifies matter a great deal: For compact groups, every irreducible representation is finite. Hence, taking components and using double
indices, (17) is a matrix equation of the form $T^*x = c$, which has a unique solution for all $c$ if and only if the adjoint equation $T^\dagger x = c$ has a unique solution for all $c$, and this happens if and only if $T^\dagger y = 0$ implies $y = 0$.

Assume then that (17) does not have a unique solution. Then there exists a nontrivial set of constants $\{c_{kl}\}$ such that
\[ \int f(g)\hat{U}(g)vv^\dagger\hat{U}(g)^\dagger\nu(dg) = 0 \text{ for } f(g) = \sum_{k,l} c_{kl}\hat{U}_{kl}(g). \] (18)

From (18) for a fixed $v \in V$ it follows that
\[ \int f(g)\hat{U}(g)ab^\dagger\hat{U}(g)^\dagger\nu(dg) = 0, \] (19)

first for $a = \hat{U}(g_a)v$, $b = \hat{U}(g_b)v$ by multiplying (18) by suitable $\hat{U}$-matrices on both sides, then for linear combinations of such terms, and hence for all $a, b \in V$, since these linear combinations span $V$ by Theorem 2(b).

Then, by letting $a$ and $b$ be basis vectors in $V$ and multiplying both sides by another set of basis vectors, it follows that
\[ \int f(g)\hat{U}_{ij}(g)\hat{U}_{kl}(g)\nu(dg) = 0 \]
for all $i, j, k, l$. We now again use the fact that the Fourier transform can be uniquely inverted. This first gives $f(g)\hat{U}_{ij}(g) = 0$ for all $i, j$, from which it follows that $f(g) = 0$. Now $f(g^{-1}) = \text{tr}(C\hat{U}(g^{-1}))$ for $C = \{c_{ij}\}$, which is just the formula for the inverse Fourier transform. Thus $c_{ij} \equiv 0$. This proves the existence and the uniqueness of the solution of (16) for integrable $c(\cdot)$.

**Corollary**

If $c(g) \geq 0$ for all $g$, then $Q$ is nonnegative definite. If $\{c_i(\cdot)\}$ are functions such that $\sum c_i(\cdot) = 1$ for all $g$, then the corresponding operators $Q_i$ satisfy $\sum Q_i = I \cdot \text{const.}$, where the constant will be 1 if we normalize: $v^\dagger v = 1$.

**Proof**

Straightforward verification.