
A Basis for Statistical Theory and Quantum Theory

Inge S. Helland

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/53702>

1. Introduction

Compared to other physical theories, the foundation of quantum mechanics is very formal and abstract. The pure state of a system is defined as a complex vector (or ray) in some abstract vector space, the observables as Hermitian operators on this space. Even a modern textbook like Ballentine [1] starts by introducing two abstract postulates:

1. To each dynamical variable there corresponds a linear operator, and the possible values of the dynamical variable are the eigenvalues of the operator.
2. To each state there corresponds a unique operator ρ . The average value of a dynamical variable \mathbf{R} , represented by the operator R , in the state given by the operator ρ is given by

$$\langle \mathbf{R} \rangle = \frac{\text{Tr}(\rho R)}{\text{Tr}(\rho)}. \quad (1)$$

Here Tr is the trace operator. The discussion in [1] goes on by arguing that R must be Hermitian (have real eigenvalues) and that ρ ought to be positive with trace 1. An important special case is when ρ is one-dimensional: $\rho = |\psi\rangle\langle\psi|$ for a vector $|\psi\rangle$. Then the state is pure, and is equivalently specified by the vector $|\psi\rangle$. In general the formula (1) is a consequence of Born's formula: The probability of observing a pure state $|\phi\rangle$ when the system is prepared in a pure state $|\psi\rangle$ is given by $|\langle\phi|\psi\rangle|^2$.

From these two postulates a very rich theory is deduced, a theory which has proved to be in agreement with observations in each case where it has been tested. Still, the abstract nature of the basic postulates leaves one a little uneasy: Is it possible to find another basis which is more directly connected to what one observes in nature? The purpose of this chapter is to show that to a large extent one can give a positive answer to this question.

Another problem is that there are many interpretations of quantum mechanics. In this chapter I will choose an *epistemic* interpretation: Quantum mechanics is related to the knowledge we get about nature, not directly to how nature 'is'. The latter aspect - the ontological aspect of nature - is something we can talk about when all observers agree on the same information. Any knowledge about nature is found through an epistemic process - an experiment or an observational study. Typically we ask a question: What is θ ? And after the epistemic process is completed, nature gives an answer, in the simplest case: $\theta = u_k$, where u_k is one of several possible values. Here θ is what we will call an *epistemic conceptual variable* or *e-variable*, a variable defined by an observer or by a group of observers and defining the epistemic process.

In all empirical sciences, epistemic questions like this are posed to nature. It is well known that the answers are not always that simple. Typically we end up with a confidence interval (a frequentist concept) or a credibility interval (a Bayesian concept) for θ . This leads us into statistical science. In statistics, θ is most often called a parameter, and is often connected to a population of experimental units. But there are instances also in statistics where we want to predict a value for a single unit. The corresponding intervals are then called prediction intervals. In this chapter we will also use θ for an unknown variable for a single unit, which is a situation very often met in physics. This is the generalization we think about when we in general call θ an e-variable, not a parameter. Also, the notion of a parameter may have a different meaning in physics, so by this we will avoid confusion.

A more detailed discussion than what can be covered here, can be found in Helland [2].

2. A basis for statistics

Every experiment or observational study is made in a context. Part of the context may be physical, another part may be historical, including earlier experiments. Also, the status of the observer(s) may be seen as a part of the context, and another part of the context may be conceptual, including a goal for the study. In all our discussion, we assume that we have conditioned upon the context τ . We can imagine the context formulated as a set of propositions. But propositional calculus corresponds to set theory, as both are Boolean algebras. Therefore we can here in principle use the familiar concept of conditioning as developed in Kolmogorov's theory of probability, where it is defined as a Radon-Nikodym derivative. Readers unfamiliar to this mathematics may think of a more intuitive conditioning concept.

In addition, for every experiment, we have an e-variable of interest θ and we have data z . A basis for all statistical theory is the statistical model, the distribution of z as a function of θ . Conceptual variables which are not of interest, may be taken as part of the context τ . The density of the statistical model, seen as a function of θ , is called the likelihood. We will assume throughout:

- 1) The distribution of z , given τ , depends on an unknown e-variable θ .
- 2) If τ or part of τ has a distribution, this is independent of θ . The part of τ which does not have a distribution is functionally independent of θ .

A function of the data is called a statistic $t(z)$. Often it is of interest to reduce the data to a sufficient statistic, a concept due to R. A. Fisher.

Definition 1.

We say that $t = t(z)$ is a τ -sufficient for θ if the conditional distribution of z , given t, τ and θ is independent of θ .

The intuitive notion here is that if the distribution of z , given t is independent of θ , the distribution of the whole data set might as well be generated by the distribution of t , given θ together with some random mechanism which is totally independent of the conceptual variable of interest. This is the basis for

The sufficiency principle.

Consider an experiment in a context τ , let z be the data of this experiment, and let θ be the e -variable of interest. Let $t = t(z)$ be a τ -sufficient statistic for θ . Then, if $t(z_1) = t(z_2)$, the data z_1 and z_2 contain the same experimental evidence about θ in the context τ .

Here 'experimental evidence' is left undefined. The principle is regarded as intuitively obvious by most statisticians.

Another principle which is considered intuitively obvious by most statisticians, is

The conditionality principle 1.

Suppose that there are two experiments E_1 and E_2 with common conceptual variable of interest θ and with equivalent contexts τ . Consider a mixed experiment E^* , whereby $u = 1$ or $u = 2$ is observed, each having probability $1/2$ (independent of θ , the data of the experiments and the contexts), and the experiment E_u is then performed. Then the evidence about θ from E^* is just the evidence from the experiment actually performed.

Two contexts τ and τ' are defined to be equivalent if there is a one-to-one correspondence between them: $\tau' = f(\tau)$; $\tau = f^{-1}(\tau')$. The principle can be motivated by simple examples. From these examples one can also deduce

The conditionality principle 2.

In the situation of conditionality principle 1 one should in any statistical analysis condition upon the outcome of the coin toss.

It caused much discussion among statisticians when Birnbaum [3] proved that the sufficiency principle and the conditionality principle 1 together imply

The likelihood principle.

Consider two experiments with equivalent contexts τ , and assume that θ is the same full e -variable in both experiments. Suppose that the two observations z_1^* and z_2^* have proportional likelihoods in the two experiments. Then these two observations produce the same evidence on θ in this context.

It is crucial for the present chapter that these principles may be generalized from experiments to any epistemic processes involving data such that 1) and 2) are satisfied.

An important special case of the likelihood principle is when E_1 and E_2 are the same experiment and z_1^* and z_2^* have equal likelihoods. Then the likelihood principle says that any experimental evidence on θ must only depend on the likelihood (given the context). Without taking the context into account this is really controversial. It seems like common statistical methods like confidence intervals and test of hypotheses are excluded. But this is

saved when we can take confidence levels, alternative hypotheses, test levels etc. as part of the context.

A discussion of these common statistical methods will not be included here; the reader is referred to [2] for this. Also, a discussion of the important topic of model reduction in statistics will be omitted here. Sometimes a statistical model contains more structure than what has been assumed here; for instance group actions may be defined on the space of e-variables. Then any model reduction should be to an orbit or to a set of orbits for the group; for examples, see [2].

3. Inaccessible conceptual variables and quantum theory

An e-variable as it is used here is related to the question posed in an epistemic process: What is the value of θ ? Sometimes we can obtain an accurate answer to such a question, sometimes not. We call θ accessible if we in principle can devise an experiment such that θ can be assessed with arbitrary accuracy. If this in principle is impossible, we say that θ is inaccessible.

Consider a single medical patient which at time $t = 0$ can be given one out of two mutually exclusive treatments A or B. The time θ^A until recovery given treatment A can be measured accurately by giving this treatment and waiting a sufficiently long time, likewise the time θ^B until recovery given treatment B. But consider the vector $\phi = (\theta^A, \theta^B)$. This vector can not be assessed with arbitrary accuracy by any person neither before, during nor after treatment. The vector ϕ is inaccessible. A similar phenomenon occurs in all counterfactual situations.

Many more situations with inaccessible conceptual variables can be devised. Consider a fragile apparatus which is destroyed after a single measurement of some quantity θ_1 , and let θ_2 be another quantity which can only be measured by dismantling the apparatus. Then $\phi = (\theta_1, \theta_2)$ is inaccessible. Or consider two sensitive questions to be posed to a single person at some moment of time, where we expect that the order in which the questions are posed may be relevant for the answers. Let (θ_1, θ_2) be the answers when the questions are posed in one order, and let (θ_3, θ_4) be the answers when the questions are posed in the opposite order. Then the vector $\phi = (\theta_1, \theta_2, \theta_3, \theta_4)$ is inaccessible.

I will approach quantum mechanics by looking upon it as an epistemic science and pointing out the different inaccessible conceptual variables. First, by Heisenberg's uncertainty principle, the vector (ξ, π) is inaccessible, where ξ is the theoretical position and π is the theoretical momentum of a particle. This implies that $(\xi(t_1), \xi(t_2))$, the positions at two different times, is an inaccessible vector. Hence the trajectory of the particle is inaccessible. In the two-slit experiment (α, θ) is inaccessible, where α denotes the slit that the particle goes through, and θ is the phase of the particle's wave as it hits the screen.

In this chapter I will pay particular attention to a particle's spin/ angular momentum. The spin or angular momentum vector is inaccessible, but its component λ^a in any chosen direction a will be accessible.

It will be crucial for my discussion that even though a vector is inaccessible, it can be seen upon as an abstract quantity taking values in some space and one can often act on it by group actions. Thus in the medical example which started this section, a change of time units will affect the whole vector ϕ , and a spin vector can be acted upon by rotations.

4. The maximal symmetrical epistemic setting

A general setting will be described, and then I will show that spin and angular momentum are special cases of this setting. This is called the maximal symmetrical epistemic setting.

Consider an inaccessible conceptual variable ϕ , and let there be accessible e-variables $\lambda^a(\phi)$ ($a \in \mathcal{A}$) indexed by some set \mathcal{A} . Thus for each a , one can ask the question: What is the value of λ^a ? and get some information from experiment. To begin with, assume that these are maximally accessible, more precisely maximal in the ordering where $\alpha < \beta$ when $\alpha = f(\beta)$ for some f . This can be assumed by Zorn's lemma, but it will later be relaxed. For $a \neq b$ let there be an invertible transformation g_{ab} such that $\lambda^b(\phi) = \lambda^a(g_{ab}(\phi))$.

In general, let a group H act on a conceptual variable ϕ . A function $\eta(\phi)$ is said to be permissible with respect to H if $\eta(\phi_1) = \eta(\phi_2)$ implies $\eta(h\phi_1) = \eta(h\phi_2)$ for all $h \in H$. Then one can define a corresponding group \tilde{H} acting upon η . For a given function $\eta(\phi)$ there is a maximal group with respect to which it is permissible.

Now fix $0 \in \mathcal{A}$ and let G^0 be the maximal group under which $\lambda^0(\phi)$ is permissible. Take $G^a = g_{a0}G^0g_{0a}$, and let G be the smallest group containing G^0 and all the transformations g_{a0} . It is then easy to see that G^a is the maximal group under which $\lambda^a(\phi)$ is permissible, and that G is the group generated by G^a ; $a \in \mathcal{A}$ and the transformations g_{ab} . Make the following assumptions about G :

- a) It is a locally compact topological group satisfying weak conditions such that an invariant measure ρ exists on the space Φ of ϕ 's.
- b) $\lambda^a(\phi)$ varies over an orbit or a set of orbits of the smaller group G^a . More precisely: λ^a varies over an orbit or a set of orbits of the corresponding group \tilde{G}^a on its range.
- c) G is generated by the product of elements of $G^a, G^b, \dots; a, b, \dots \in \mathcal{A}$.

As an important example, let ϕ be the spin vector or the angular momentum vector for a particle or a system of particles. Let G be the group of rotations of the vector ϕ , that is, the group which fixes the norm $\|\phi\|$. Next, choose a direction a i space, and focus upon the spin component in this direction: $\zeta^a = \|\phi\|\cos(\phi, a)$. The largest subgroup G^a with respect to which $\zeta^a(\phi)$ is permissible, is given by rotations around a together with a reflection in a plane perpendicular to a . However, the action of the corresponding group \tilde{G}^a on ζ^a is just a reflection together with the identity.

Finally introduce model reduction. As mentioned at the end of the previous section, such a model reduction for ζ^a should be to an orbit or to a set of orbits for the group \tilde{G}^a as acting on ζ^a . These orbits are given as two-point sets $\pm c$ together with the single point 0. To conform to the ordinary theory of spin/angular momentum, I will choose the set of orbits indexed by an integer or half-integer j and let the reduced set of orbits be $-j, -j + 1, \dots, j - 1, j$. Letting λ^a be the e-variable ζ^a reduced to this set of orbits of \tilde{G}^a , and assuming it to be a maximally accessible e-variable, we can prove the general assumptions of the maximal symmetrical epistemic setting (except for the case $j = 0$, where we must redefine G to be the trivial group). For instance, here is an indication of the proof leading to assumption c) above: given a and b , a transformation g_{ab} sending $\lambda_a(\phi)$ onto $\lambda_b(\phi)$ can be obtained by a reflection in a plane orthogonal to the two vectors a and b , a plane containing the midline between a and b .

The case with one orbit and $c = j = 1/2$ corresponds to electrons and other spin $1/2$ particles.

In general, assumption b) in the maximal symmetrical epistemic setting may be motivated in a similar manner: First, a conceptual variable ζ^a is introduced for each a through a chosen focusing, then define G^a as the maximal group under which $\zeta^a(\phi)$ is permissible, with \tilde{G}^a being the corresponding group acting on ζ^a . Finally define λ^a as the reduction of ζ^a to a set of orbits of \tilde{G}^a . The content of assumption b) is that it is *this* λ^a which is maximally accessible. This may be regarded as the quantum hypothesis.

5. Hilbert space, pure states and operators

Consider the maximal symmetrical epistemic setting. The crucial step towards the formalism of quantum mechanics is to define a Hilbert space, that is, a complete inner product space which can serve as a state space for the system.

By assumption a) there exists an invariant measure ρ for the group's action: $\rho(gA) = \rho(A)$ for all $g \in G$ and all Borel-measurable subsets A of the space Φ of inaccessible conceptual variables. If G is transitive on Φ , then ρ is unique up to a multiplicative constant. For compact groups ρ can be normalized, i.e., taken as a probability measure. For each a define

$$H^a = \{f \in L^2(\Phi, \rho) : f(\phi) = r(\lambda^a(\phi)) \text{ for some function } r.\}$$

Thus H^a is the set of L^2 -functions that are functions of $\lambda^a(\phi)$. Since H^a is a closed subspace of the Hilbert space $L^2(\Phi, \rho)$, it is itself a Hilbert space. To define our state space H , we now fix an arbitrary index $a = 0 \in \mathcal{A}$, and take

$$H = H^0.$$

First look at the case where the accessible e-variables take a finite, discrete set of values. Let $\{u_k\}$ be the set of possible values of λ^a . Since $\lambda^a(\cdot)$ is maximal, $\{u_k\}$ can be taken to be independent of a , see [2]. Now go back to the definition of an epistemic process: We start by choosing a , that is, ask an epistemic question: What is the value of λ^a ? After the process we get some information; I will here look upon the simple case where we get full knowledge: $\lambda^a = u_k$. I define this as a pure state of the system; it can be characterized by the indicator function $\mathbf{1}(\lambda^a(\phi) = u_k)$. This is a function in H^a , but I will show below that one can find an invertible operator V^a such that

$$f_k^a(\phi) = V^a \mathbf{1}(\lambda^a(\phi) = u_k) \tag{2}$$

is a unique function in $H = H^0$. Since H in this case is a K -dimensional vector space, where K is the number of values u_k , we can regard f_k^a as a K -dimensional vector. To conform to the ordinary quantum mechanical notation, I write this as a ket-vector $|a; k\rangle = f_k^a$. It is easy to see that $\{|0; k\rangle; k = 1, \dots, K\}$ is an orthonormal basis of H when ρ is normalized to be 1 for the whole space Φ . I will show below that $\{|a; k\rangle; k = 1, \dots, K\}$ has the same property. My main point is that $|a; k\rangle$ is characterized by and characterizes a question: What is λ^a ? together with an answer: $\lambda^a = u_k$. This is a pure state for the maximal symmetrical epistemic setting.

I will also introduce operators by

$$A^a = \sum_{k=1}^K u_k |a; k\rangle \langle a; k|,$$

where $\langle a; k|$ is the bra vector corresponding to $|a; k\rangle$. This is by definition the observator corresponding to the e-variable λ^a . Since λ^a is maximal, A^a will have non-degenerate eigenvalues u_k . Knowing A^a , we will have information of all possible values of λ^a together with information about all possible states connected to this variable.

The rest of this section will be devoted to proving (2) and showing the properties of the state vectors $|a; k\rangle$. To allow for future generalizations I now allow the accessible e-variables λ^a to take any set of values, continuous or discrete. The discussion will by necessity be a bit technical. First I define the (left) regular representation U for a group G . For given $f \in L^2(\Phi, \rho)$ and given $g \in G$ we define a new function $U(g)f$ by

$$U(g)f(\phi) = f(g^{-1}\phi). \tag{3}$$

Without proof I mention 5 properties of the set of operators $U(g)$:

- $U(g)$ is linear: $U(g)(a_1 f_1 + a_2 f_2) = a_1 U(g)f_1 + a_2 U(g)f_2$.
- $U(g)$ is unitary: $\langle U(g)f_1, f_2 \rangle = \langle f_1, U(g)^{-1}f_2 \rangle$ in $L^2(\Phi, \rho)$.
- $U(g)$ is bounded: $\sup_{f: \|f\|=1} \|U(g)f\| = 1 < \infty$.
- $U(\cdot)$ is continuous: If $\lim g_n = g_0$ in the group topology, then $\lim U(g_n) = U(g_0)$ (in the matrix norm in the finite-dimensional case, which is what I will focus on here, in general in the topology of bounded linear operators).
- $U(\cdot)$ is a homomorphism: $U(g_1 g_2) = U(g_1)U(g_2)$ for all g_1, g_2 and $U(e) = I$ for the unit element.

The concept of *homomorphism* will be crucial in this section. In general, a homomorphism is a mapping $k \rightarrow k'$ between groups K and K' such that $k_1 \rightarrow k'_1$ and $k_2 \rightarrow k'_2$ implies $k_1 k_2 \rightarrow k'_1 k'_2$ and such that $e \rightarrow e'$ for the identities. Then also $k^{-1} \rightarrow (k')^{-1}$ when $k \rightarrow k'$.

A *representation* of a group K is a continuous homomorphism from K into a group of invertible operators on some vector space. If the vector space is finite dimensional, the linear operators can be taken as matrices. There is a large and useful mathematical theory about operator (matrix) representations of groups; some of it is sketched in Appendix 3 of [2]. Equation (3) gives one such representation of the basic group G on the vector space $L^2(\Phi, \rho)$.

Proposition 1.

Let $U_a = U(g_{0a})$ with g_{ab} defined in the beginning of Section 4. Then

$$H^a = U_a^{-1} H \text{ through } r(\lambda^a(\phi)) = U_a^{-1} r(\lambda^0(\phi)).$$

Proof. If $f \in H^a$, then $f(\phi) = r(\lambda^a(\phi)) = r(\lambda^0(g_{0a}\phi)) = U(g_{0a})^{-1}r(\lambda^0(\phi)) = U_a^{-1}f_0(\phi)$, where $f_0 \in H = H^0$.

Since $a = 0$ is a fixed but arbitrary index, this gives in principle a unitary connection between the different choices of H , different representations of the ‘Hilbert space apparatus’. However this connection cannot be used directly in (2), since if $f_k^a = \mathbf{1}(\lambda^a = u_k)$ is the state function representing the question: What is λ^a ? together with the answer $\lambda^a = u_k$, then we have

$$U_a f_k^a = U(g_{0a})\mathbf{1}(\lambda^0(g_{0a}\phi) = u_k) = U(g_{0a})U(g_{0a})^{-1}\mathbf{1}(\lambda^0(\phi) = u_k) = f_k^0.$$

Thus by this simple transformation the indicator functions in H are not able to distinguish between the different questions asked.

Another reason why the simple solution is not satisfactory is that the regular representation U will not typically be a representation of the whole group G on the Hilbert space H . This can however be amended by the following theorem. Its proof and the resulting discussion below are where the Assumption c) of the maximal symmetrical epistemic setting is used. Recall that throughout, upper indices (G^a, g^a) are for the subgroups of G connected to the accessible variables λ^a , similarly (\tilde{G}^a, \tilde{g}^a) for the group (elements) acting upon λ^a . Lower indices (e.g., $U_a = U(g_{0a})$) are related to the transformations between these variables.

Theorem 1.

- (i) A representation (possibly multivalued) V of the whole group G on H can always be found.
- (ii) For $g^a \in G^a$ we have $V(g^a) = U_a U(g^a) U_a^\dagger$.

Proof. (i) For each a and for $g^a \in G^a$ define $V(g^a) = U(g_{0a})U(g^a)U(g_{a0})$. Then $V(g^a)$ is an operator on $H = H^0$, since it is equal to $U(g_{0a}g^a g_{a0})$, and $g_{0a}g^a g_{a0} \in G^0$ by the construction of G^a from G^0 . For a product $g^a g^b g^c$ with $g^a \in G^a, g^b \in G^b$ and $g^c \in G^c$ we define $V(g^a g^b g^c) = V(g^a)V(g^b)V(g^c)$, and similarly for all elements of G that can be written as a finite product of elements from different subgroups.

Let now g and h be any two elements in G such that g can be written as a product of elements from G^a, G^b and G^c , and similarly h (the proof is similar for other cases.) It follows that $V(gh) = V(g)V(h)$ on these elements, since the last factor of g and the first factor of h either must belong to the same subgroup or to different subgroups; in both cases the product can be reduced by the definition of the previous paragraph. In this way we see that V is a representation on the set of finite products, and since these generate G by Assumption c), and since U , hence by definition V , is continuous, it is a representation of G .

Since different representations of g as a product may give different solutions, we have to include the possibility that V may be multivalued.

- (ii) Directly from the proof of (i).

What is meant by a multivalued representation? As an example, consider the group $SU(2)$ of unitary 2×2 matrices. Many books in group theory will state that there is a homomorphism from $SU(2)$ to the group $SO(3)$ of real 3-dimensional rotations, where the kernel of the homomorphism is $\pm I$. This latter statement means that both $+I$ and $-I$ are mapped into the identity rotation by the homomorphism.

In this case there is no unique inverse $SO(3) \rightarrow SU(2)$, but nevertheless we may say informally that there is a multivalued homomorphism from $SO(3)$ to $SU(2)$. Here is a way to make this precise:

Extend $SU(2)$ to a new group with elements (g, k) , where $g \in SU(2)$ and k is an element of the group $K = \{\pm 1\}$ with the natural multiplication. The multiplication in this extended group is defined by $(g_1, k_1) \cdot (g_2, k_2) = (g_1 g_2, k_1 k_2)$, and the inverse by $(g, k)^{-1} = (g^{-1}, k^{-1})$. Then there is an invertible homomorphism between this extended group and $SO(3)$.

A similar construction can be made with the representation V of Theorem 1.

Theorem 2.

- (i) There is an extended group G' such that V is a univariate representation of G' on H .
- (ii) There is a unique mapping $G' \rightarrow G$, denoted by $g' \rightarrow g$, such that $V(g') = V(g)$. This mapping is a homomorphism.

Proof. (i) Assume as in Theorem 1 that we have a multivalued representation V of G . Define a larger group G' as follows: If $g^a g^b g^c = g^d g^e g^f$, say, with $g^k \in G^k$ for all k , we define $g'_1 = g^a g^b g^c$ and $g'_2 = g^d g^e g^f$. A similar definition of new group elements is done if we have equality of a limit of such products. Let G' be the collection of all such new elements that can be written as a formal product of elements $g^k \in G^k$ or as limits of such symbols. The product is defined in the natural way, and the inverse by for example $(g^a g^b g^c)^{-1} = (g^c)^{-1} (g^b)^{-1} (g^a)^{-1}$. By Assumption 2c), the group G' generated by this construction must be at least as large as G . It is clear from the proof of Theorem 1 that V also is a representation of the larger group G' on H , now a one-valued representation.

(ii) Again, if $g^a g^b g^c = g^d g^e g^f = g$, say, with $g^k \in G^k$ for all k , we define $g'_1 = g^a g^b g^c$ and $g'_2 = g^d g^e g^f$. There is a natural map $g'_1 \rightarrow g$ and $g'_2 \rightarrow g$, and the situation is similar for other products and limits of products. It is easily shown that this mapping is a homomorphism.

Note that while G is a group of transformations on Φ , the extended group G' must be considered as an abstract group.

Theorem 3.

- (i) For $g' \in G'$ there is a unique $g^0 \in G^0$ such that $V(g') = U(g^0)$. The mapping $g' \rightarrow g^0$ is a homomorphism.
- (ii) If $g' \rightarrow g^0$ by the homomorphism of (i), and $g' \neq e'$ in G' , then $g^0 \neq e$ in G^0 .

Proof. (i) Consider the case where $g' = g^a g^b g^c$ with $g^k \in G^k$. Then by the proof of Theorem 1:

$$\begin{aligned}
 V(g') &= U_a U(g^a) U_a^\dagger U_b U(g^b) U_b^\dagger U_c U(g^c) U_c^\dagger = U(g_{0a} g^a g_{a0} g_{0b} g^b g_{b0} g_{0c} g^c g_{c0}) \\
 &= U(g^0),
 \end{aligned}$$

where $g^0 \in G^0$. The group element g^0 is unique since the decomposition $g' = g^a g^b g^c$ is unique for $g' \in G'$. The proof is similar for other decompositions and limits of these. By the construction, the mapping $g' \rightarrow g^0$ is a homomorphism.

(ii) Assume that $g^0 = e$ and $g' \neq e'$. Since $U(g^0)\tilde{f}(\lambda^0(\phi)) = \tilde{f}(\lambda^0((g^0)^{-1}(\phi)))$, it follows from $g^0 = e$ that $U(g^0) = I$ on H . But then from (i), $V(g') = I$, and since V is a univariate representation, it follows that $g' = e'$, contrary to the assumption.

The theorems 1-3 are valid in any maximal symmetrical epistemic setting. I will now again specialize to the case where the accessible e-variables λ have a finite discrete range. This is often done in elementary quantum theory texts, in fact also in recent quantum foundation papers, and in our situation it has several advantages:

- It is easy to interpret the principle that λ can be estimated with any fixed accuracy.
- In particular, confidence regions and credibility regions for an accessible e-variable can be taken as single points if observations are accurate enough.
- The operators involved will be much simpler and are defined everywhere.
- The operators A^a can be understood directly from the epistemic setting; see above.

So look at the statement $\lambda^a(\phi) = u_k$. This means two things: 1) One has sought information about the value of the maximally accessible e-variable λ^a , that is, asked the question: What is the value of λ^a ? 2) One has obtained the answer $\lambda^a = u_k$. This information can be thought of as a perfect measurement, and it can be represented by the indicator function $\mathbf{1}(\lambda^a(\phi) = u_k)$, which is a function in H^a . From Proposition 1, this function can by a unitary transformation be represented in H , which now is a vector space with a discrete basis, a finite-dimensional vector space: $U_a f_k^a$. However, we have seen that this tentative state definition $U_a \mathbf{1}(\lambda^a(\phi) = u_k) = U(g_{0a})\mathbf{1}(\lambda^0(g_{0a}\phi) = u_k)$ led to ambiguities. These ambiguities can be removed by replacing the two g_{0a} 's here in effect by different elements g'_{0ai} of the extended group G' . Let g'_{0a1} and g'_{0a2} be two different such elements where both $g'_{0a1} \rightarrow g_{0a}$ and $g'_{0a2} \rightarrow g_{0a}$ according to Theorem 2 (ii). I will prove in a moment that this is in fact always possible when $g_{0a} \neq e$. Let $g'_a = (g'_{0a1})^{-1}g'_{0a2}$, and define

$$f_k^a(\phi) = V(g'_a)U_a \mathbf{1}(\lambda^a(\phi) = u_k) = V(g'_a)f_k^0(\phi).$$

This gives the relation (2).

In order that the interpretation of f_k^a as a state $|a; k\rangle$ shall make sense, I need the following result. I assume that \tilde{G}^0 is non-trivial.

Theorem 4.

a) Assume that two vectors in H satisfy $|a; i\rangle = |b; j\rangle$, where $|a; i\rangle$ corresponds to $\lambda^a = u_i$ for one perfect measurement and $|b; j\rangle$ corresponds to $\lambda^b = u_j$ for another perfect measurement. Then there is a one-to-one function F such that $\lambda^b = F(\lambda^a)$ and $u_j = F(u_i)$. On the other hand, if $\lambda^b = F(\lambda^a)$ and $u_j = F(u_i)$ for such a function F , then $|a; i\rangle = |b; j\rangle$.

b) Each $|a; k\rangle$ corresponds to only one $\{\lambda^a, u_k\}$ pair except possibly for a simultaneous one-to-one transformation of this pair.

Proof. a) I prove the first statement; the second follows from the proof of the first statement. Without loss of generality consider a system where each e-variable λ takes only two values, say 0 and 1. Otherwise we can reduce to a degerate system with just these two values: The

statement $|a; i\rangle = |b; j\rangle$ involves, in addition to λ^a and λ^b , only the two values u_i and u_j . By considering a function of the maximally accessible e-variable (compare the next section), we can take one specific value equal to 1, and the others collected in 0. By doing this, we also arrange that both u_i and u_j are 1, so we are comparing the state given by $\lambda^a = 1$ with the state given by $\lambda^b = 1$.

By the definition, $|a; 1\rangle = |b; 1\rangle$ can be written

$$V(g'_a)U_a\mathbf{1}(\lambda^a(\phi) = 1) = V(g'_b)U_b\mathbf{1}(\lambda^b(\phi) = 1)$$

for group elements g'_a and g'_b in G' .

Use Theorem 3(i) and find g_a^0 and g_b^0 in G^0 such that $V(g'_a) = U(g_a^0)$ and $V(g'_b) = U(g_b^0)$. Therefore

$$U(g_a^0)U(g_{0a})\mathbf{1}(\lambda^a(\phi) = 1) = U(g_b^0)U(g_{0b})\mathbf{1}(\lambda^b(\phi) = 1);$$

$$\mathbf{1}(\lambda^a(\phi) = 1) = U(g^0)\mathbf{1}(\lambda^b(\phi) = 1) = \mathbf{1}(\lambda^b((g^0)^{-1}\phi) = 1),$$

for $g^0 = (g_{0a})^{-1}(g_a^0)^{-1}g_b^0g_{0b}$.

Both λ^a and λ^b take only the values 0 and 1. Since the set where $\lambda^b(\phi) = 1$ can be transformed into the set where $\lambda^a(\phi) = 1$, we must have $\lambda^a = F(\lambda^b)$ for some transformation F .

b) follows trivially from a).

Corollary.

The group G is properly contained in G' , so the representation V of Theorem 1 is really multivalued.

Proof. If we had $G' = G$, then $|a; k\rangle$ and $|b; k\rangle$ both reduce to $U_a\mathbf{1}(\lambda^a(\phi) = u_k) = U_b\mathbf{1}(\lambda^b(\phi) = u_k) = \mathbf{1}(\lambda^0 = u_k)$, so Theorem 4 and its proof could not be valid.

Theorem 4 and its corollary are also valid in the situation where we are interested in just two accessible variables λ^a and λ^b , which might as well be called λ^0 and λ^a . We can then provisionally let the group G be generated by g_{0a} , $g_{a0} = g_{0a}^{-1}$ and all elements g^0 and g^a . The earlier statement that it is always possible to find two *different* elements g'_{0a1} and g'_{0a2} in G' which are mapped onto g_{0a} follows.

Finally we have

Theorem 5.

For each $a \in \mathcal{A}$, the vectors $\{|a; k\rangle; k = 1, 2, \dots\}$ form an orthonormal basis for H .

Proof. Taking the invariant measure ρ on H as normalized to 1, the indicator functions $|0; k\rangle = \mathbf{1}(\lambda^0(\phi) = u_k)$ form an orthonormal basis for H . Since the mapping $|0; k\rangle \rightarrow |a; k\rangle$ is unitary, the Theorem follows.

So if $b \neq a$ and k is fixed, there are complex constants c_{ki} such that $|b; k\rangle = \sum_i c_{ki}|a; i\rangle$. This opens for the interference effects that one sees discussed in quantum mechanical texts. In

particular $|a; k\rangle = \sum_i d_{ki} |0; i\rangle$ for some constants d_{ki} . This is the first instance of something that we also will meet later in different situations: New states in H are found by taking linear combinations of a basic set of state vectors.

6. The general symmetrical epistemic setting

Go back to the definition of the maximal symmetrical epistemic setting. Let again ϕ be the inaccessible conceptual variable and let λ^a for $a \in \mathcal{A}$ be the maximal accessible conceptual variables, functions of ϕ . Let the corresponding induced groups G^a and G satisfy the assumptions a)-c). Finally, let t^a for each a be an arbitrary function on the range of λ^a , and assume that we observe $\theta^a = t^a(\lambda^a); a \in \mathcal{A}$. We will call this the symmetrical epistemic setting; it is no longer necessarily maximal with respect to the observations θ^a .

Consider first the quantum states $|a; k\rangle$. We are no longer interested in the full information on λ^a , but keep the Hilbert space as in Section 5, and now let $h_k^a(\phi) = \mathbf{1}(t^a(\lambda^a) = t^a(u_k)) = \mathbf{1}(\theta^a = u_k^a)$, where $u_k^a = t^a(u_k)$. We let again g'_{0a1} and g'_{0a2} be two distinct elements of G^a such that $g'_{0ai} \rightarrow g_{0ai}$, define $g'_a = (g'_{0a1})^{-1} g'_{0a2}$ and then

$$|a; k\rangle = V(g'_a) U_a h_k^a = V(g'_a) |0; k\rangle,$$

where $|0; k\rangle = h_k^0$.

Interpretation of the state vector $|a; k\rangle$:

1) The question: 'What is the value of θ^a ?' has been posed. 2) We have obtained the answer $\theta^a = u_k^a$. Both the question and the answer are contained in the state vector.

From this we may define the operator connected to the e-variable θ^a :

$$A^a = \sum_k u_k^a |a; k\rangle \langle a; k| = \sum_k t^a(u_k) |a; k\rangle \langle a; k|.$$

Then A^a is no longer necessarily an operator with distinct eigenvalues, but A^a is still Hermitian: $A^{a\dagger} = A^a$.

Interpretation of the operator A^a :

This gives all possible states and all possible values corresponding to the accessible e-variable θ^a .

The projectors $|a; k\rangle \langle a; k|$ and hence the ket vectors $|a; k\rangle$ are no longer uniquely determined by A^a : They can be transformed arbitrarily by unitary transformations in each space corresponding to one eigenvalue. In general I will redefine $|a; k\rangle$ by allowing it to be subject to such transformations. These transformed eigenvectors all still correspond to the same eigenvalue, that is, the same observed value of θ^a and they give the same operators A^a . In particular, in the maximal symmetric epistemic setting I will allow an arbitrary constant phase factor in the definition of the $|a; k\rangle$'s.

As an example of the general construction, assume that λ^a is a vector: $\lambda^a = (\theta^{a_1}, \dots, \theta^{a_m})$. Then one can write a state vector corresponding to λ^a as

$$|a; k\rangle = |a_1; k_1\rangle \otimes \dots \otimes |a_m; k_m\rangle$$

in an obvious notation, where $a = (a_1, \dots, a_m)$ and $k = (k_1, \dots, k_m)$. The different θ 's may be connected to different subsystems.

So far I have kept the same groups G^a and G when going from λ^a to $\theta^a = t^a(\lambda^a)$, that is from the maximal symmetrical epistemic setting to the general symmetrical epistemic setting. This implies that the (large) Hilbert space will be the same. A special case occurs if t^a is a reduction to an orbit of G^a . This is the kind of model reduction mentioned at the end of Section 2. Then the construction of the previous sections can also be carried with a smaller group action acting just upon an orbit, resulting then in a smaller Hilbert space. In the example of the previous paragraph it may be relevant to consider one Hilbert space for each subsystem. The large Hilbert space is however the correct space to use when the whole system is considered.

Connected to a general physical system, one may have many e-variables θ and corresponding operators A . In the ordinary quantum formalism, there is well-known theorem saying that, in my formulation, $\theta^1, \dots, \theta^n$ are compatible, that is, there exists an e-variable λ such that $\theta^i = t^i(\lambda)$ for some functions t^i if and only if the corresponding operators commute:

$$[A^i, A^j] \equiv A^i A^j - A^j A^i = 0 \text{ for all } i, j.$$

(See Holevo [4].) Compatible e-variables may in principle be estimated simultaneously with arbitrary accuracy.

The way I have defined pure state, the only state vectors that are allowed, are those which are eigenvectors of some physically meaningful operator. This is hardly a limitation in the spin/angular momentum case where operators corresponding to all directions are included. Nevertheless it is an open question to find general conditions under which all unit vectors in H correspond to states $|a; k\rangle$ the way I have defined them. It is shown in [5] that this holds under no further conditions for the spin 1/2 case.

7. Link to statistical inference

Assume now the symmetrical epistemic setting. We can think of a spin component in a fixed direction to be assessed. To assume a state $|a; k\rangle$ is to assume perfect knowledge of the e-variable θ^a : $\theta^a = u_k^a$. Such perfect knowledge is rarely available. In practice we have data z^a about the system, and use these data to obtain knowledge about θ^a . Let us start with Bayesian inference. This assumes prior probabilities π_k^a on the values u_k^a , and after the inference we have posterior probabilities $\pi_k^a(z^a)$. In either case we summarize this information in the density operator:

$$\sigma^a = \sum_k \pi_k^a |a; k\rangle \langle a; k|.$$

Interpretation of the density operator σ^a :

1) We have posed the question 'What is the value of θ^a ?' 2) We have specified a prior or posterior probability distribution π_k^a over the possible answers. The probability for all possible answers to the question, formulated in terms of state vectors, can be recovered from the density operator.

A third possibility for the probability specifications is a relatively new, but important concept of a confidence distribution ([6], [7]). This is a frequentist alternative to the distribution connected to a parameter (here: e-variable). The idea is that one looks at a one-sided confidence interval for any value of the confidence coefficient γ . Let the data be z , and let $(-\infty, \beta(\gamma, z)]$ be such an interval. Then $\beta(\gamma) = \beta(\gamma, z)$ is an increasing function. We define $H(\cdot) = \beta^{-1}(\cdot)$ as the confidence distribution for θ . This H is a cumulative distribution function, and in the continuous case it is characterized with the property that $H(\beta(\gamma, z))$ has a uniform distribution over $[0, 1]$ under the model. For discrete θ^a the confidence distribution function H^a is connected to a discrete distribution, which gives the probabilities π_k^a . Extending the argument in [7] to this situation, this should not be looked upon as a distribution of θ^a , but a distribution for θ^a , to be used in the epistemic process.

Since the sum of the probabilities is 1, the trace (sum of eigenvalues) of any density operator is 1. In the quantum mechanical literature, a density operator is any positive operator with trace 1.

Note that specification of the accessible e-variables θ^a is equivalent to specifying $t(\theta^a)$ for any one-to-one function t . The operator $t(A^a)$ has then distinct eigenvalues if and only if the operator A^a has distinct eigenvalues. Hence it is enough in order to specify the question 1) to give the set of orthonormal vectors $|a; k\rangle$.

Given the question a , the e-variable θ^a plays the role similar to a parameter in statistical inference, even though it may be connected to a single unit. Inference can be done by preparing many independent units in the same state. Inference is then from data z^a , a part of the total data z that nature can provide us with. All inference theory that one finds in standard texts like [8] applies. In particular, the concepts of unbiasedness, equivariance, minimaxity and admissibility apply. None of these concepts are much discussed in the physical literature, first because measurements there are often considered as perfect, at least in elementary texts, secondly because, when measurements are considered in the physical literature, they are discussed in terms of the more abstract concept of an operator-valued measure; see below.

Whatever kind of inference we make on θ^a , we can take as a point of departure the statistical model and the likelihood principle of Section 2. Hence after an experiment is done, and given some context τ , all evidence on θ^a is contained in the likelihood $p(z^a | \tau, \theta^a)$, where z^a is the portion of the data relevant for inference on θ^a , also assumed discrete. This is summarized in the likelihood effect:

$$E(z^a, \tau) = \sum_k p(z^a | \tau, \theta^a = u_k^a) |a; k\rangle \langle a; k|.$$

Interpretation of the likelihood effect $E(z^a, \tau)$:

1) We have posed some inference question on the accessible e-variable θ^a . 2) We have specified the relevant likelihood for the data. The likelihood for all possible answers of the question, formulated in terms of state vectors, can be recovered from the likelihood effect.

Since the focused question assumes discrete data, each likelihood is in the range $0 \leq p \leq 1$. In the quantum mechanical literature, an effect is any operator with eigenvalues in the range $[0, 1]$.

Return now to the likelihood principle of Section 2. The following principle follows.

The focused likelihood principle (FLP)

Consider two potential experiments in the symmetrical epistemic setting with equivalent contexts τ , and assume that the inaccessible conceptual variable ϕ is the same in both experiments. Suppose that the observations z_1^ and z_2^* have proportional likelihood effects in the two experiments, with a constant of proportionality independent of the conceptual variable. Then the questions posed in the two experiments are equivalent, that is, there is an e-variable θ^a which can be considered to be the same in the two experiments, and the two observations produce the same evidence on θ^a in this context.*

In many examples the two observations will have equal, not only proportional, likelihood effects. Then the FLP says simply that the experimental evidence is a function of the likelihood effect.

In the FLP we have the freedom to redefine the e-variable in the case of coinciding eigenvalues in the likelihood effect, that is, if $p(z^a|\tau, \theta^a = u_k) = p(z^a|\tau, \theta^a = u_l)$ for some k, l . An extreme case is the likelihood effect $E(z^a, \tau) = I$, where all the likelihoods are 1, that is, the probability of z is 1 under any considered model. Then any accessible e-variable θ^a will serve our purpose.

We are now ready to define the operator-valued measure in this discrete case:

$$M^a(B|\tau) = \sum_{z^a \in B} E(z^a, \tau)$$

for any Borel set in the sample space for experiment a . Its usefulness will be seen after we have discussed Born's formula. Then we will also have background for reading much of [9], a survey over quantum statistical inference.

8. Rationality and experimental evidence

Throughout this section I will consider a fixed context τ and a fixed epistemic setting in this context. The inaccessible e-variable is ϕ , and I assume that the accessible e-variables θ^a take a discrete set of values. Let the data behind the potential experiment be z^a , also assumed to take a discrete set of values.

Let first a single experimentalist A be in this situation, and let all conceptual variables be attached to A , although he also has the possibility to receiving information from others through part of the context τ . He has the choice of doing different experiments a , and he also has the choice of choosing different models for his experiment through his likelihood $p_A(z^a|\tau, \theta^a)$. The experiment and the model, hence the likelihood, should be chosen before the data are obtained. All these choices are summarized in the likelihood effect E , a function

of the at present unknown data z^a . For use after the experiment, he should also choose a good estimator/predictor $\hat{\theta}^a$, and he may also have to choose some loss function, but the principles behind these latter choices will be considered as part of the context τ . If he chooses to do a Bayesian analysis, the estimator should be based on a prior $\pi(\theta^a|\tau)$. We assume that A is trying to be as rational as possible in all his choices, and that this rationality is connected to his loss function or to other criteria.

What should be meant by experimental evidence, and how should it be measured? As a natural choice, let the experimental evidence that we are seeking, be the marginal probability of the obtained data for a fixed experiment and for a given likelihood function. From the experimentalist A 's point of view this is given by:

$$p_A^a(z^a|\tau) = \sum_k p_A(z^a|\tau, \theta^a = u_k) \pi_A(\theta^a = u_k|\tau),$$

assuming the likelihood chosen by A and A 's prior π_A for θ^a . In a non-Bayesian analysis, we can let $p_A^a(z^a|\tau)$ be the probability given the true value u_k^0 of the e-variable: $p_A^a(z^a|\tau) = p_A(z^a|\tau, \theta^a = u_k^0)$. In general, take $p_A^a(z^a|\tau)$ as the probability of the part of the data z^a which A assesses in connection to his inference on θ^a . By the FLP - specialized to the case of one experiment and equal likelihoods - this experimental evidence must be a function of the likelihood effect: $p_A^a(z^a|\tau) = q_A(E(z^a)|\tau)$.

We have to make precise in some way what is meant by the rationality of the experimentalist A . He has to make many difficult choices on the basis of uncertain knowledge. His actions can partly be based on intuition, partly on experience from similar situations, partly on a common scientific culture and partly on advices from other persons. These other persons will in turn have their intuition, their experience and their scientific education. Often A will have certain explicitly formulated principles on which to base his decisions, but sometimes he has to dispense with the principles. In the latter case, he has to rely on some 'inner voice', a conviction which tells him what to do.

We will formalize all this by introducing a perfectly rational superior actor D , to which all these principles, experiences and convictions can be related. We also assume that D can observe everything that is going on, in particular A , and that he on this background can have some influence on A 's decisions. The real experimental evidence will then be defined as *the probability of the data z^a from D 's point of view, which we assume also to give the real objective probabilities*. By the FLP this must again be a function of the likelihood effect E , where the likelihood now may be seen as the objectively correct model.

$$p^a(z^a|\tau) = q(E(z^a)|\tau) \tag{4}$$

As said, we assume that D is perfectly rational. This can be formalized mathematically by considering a hypothetical betting situation for D against a bookie, nature N . A similar discussion was recently done in [10] using a more abstract language. Note the difference to the ordinary Bayesian assumption, where A himself is assumed to be perfectly rational. This difference is crucial to me. I do not see any human scientist, including myself, as being

perfectly rational. We can try to be as rational as possible, but we have to rely on some underlying rational principles that partly determine our actions.

So let the hypothetical odds of a given bet for D be $(1 - q)/q$ to 1, where q is the probability as defined by (4). This odds specification is a way to make precise that, given the context τ and given the question a , the bettor's probability that the experimental result takes some value is given by q : For a given utility measured by x , the bettor D pays in an amount qx - the stake - to the bookie. After the experiment the bookie pays out an amount x - the payoff - to the bettor if the result of the experiment takes the value z^a , otherwise nothing is payed.

The rationality of D is formulated in terms of

The Dutch book principle.

No choice of payoffs in a series of bets shall lead to a sure loss for the bettor.

For a related use of the same principle, see [11].

Assumption D.

Consider in some context τ a maximal symmetrical epistemic setting where the FLP is satisfied, and the whole situation is observed and acted upon by a superior actor D as described above. Assume that D 's probabilities q given by (4) are taken as the experimental evidence, and that D acts rationally in agreement with the Dutch book principle.

A situation where all the Assumption D holds together with the assumptions of a symmetric epistemic setting will be called a *rational epistemic setting*.

Theorem 6.

Assume a rational epistemic setting. Let E_1 and E_2 be two likelihood effects in this setting, and assume that $E_1 + E_2$ also is a likelihood effect. Then the experimental evidences, taken as the probabilities of the corresponding data, satisfy

$$q(E_1 + E_2|\tau) = q(E_1|\tau) + q(E_2|\tau).$$

Proof. The result of the theorem is obvious, without making Assumption D, if E_1 and E_2 are likelihood effects connected to experiments on the same e-variable θ^a . We will prove it in general. Consider then any finite number of potential experiments including the two with likelihood effects E_1 and E_2 . Let $q_1 = q(E_1|\tau)$ be equal to (4) for the first experiment, and let $q_2 = q(E_2|\tau)$ be equal to the same quantity for the second experiment. Consider in addition the following randomized experiment: Throw an unbiased coin. If head, choose the experiment with likelihood effect E_1 ; if tail, choose the experiment with likelihood effect E_2 . This is a valid experiment. The likelihood effect when the coin shows head is $\frac{1}{2}E_1$, when it shows tail $\frac{1}{2}E_2$, so that the likelihood effect of this experiment is $E_0 = \frac{1}{2}(E_1 + E_2)$. Define $q_0 = q(E_0)$. Let the bettor bet on the results of all these 3 experiments: Payoff x_1 for experiment 1, payoff x_2 for experiment 2 and payoff x_0 for experiment 0.

I will divide into 3 possible outcomes: Either the likelihood effect from the data z is E_1 or it is E_2 or it is none of these. The randomization in the choice of E_0 is considered separately from the result of the bet. (Technically this can be done by repeating the whole series of

experiments many times with the same randomization. This is also consistent with the conditionality principle.) Thus if E_1 occurs, the payoff for experiment 0 is replaced by the expected payoff $x_0/2$, similarly if E_2 occurs. The net expected amount the bettor receives is then

$$x_1 + \frac{1}{2}x_0 - q_1x_1 - q_2x_2 - q_0x_0 = (1 - q_1)x_1 - q_2x_2 - (1 - 2q_0)\frac{1}{2}x_0 \text{ if } E_1,$$

$$x_2 + \frac{1}{2}x_0 - q_1x_1 - q_2x_2 - q_0x_0 = -q_1x_1 - (1 - q_2)x_2 - (1 - 2q_0)\frac{1}{2}x_0 \text{ if } E_2,$$

$$-q_1x_1 - q_2x_2 - 2q_0 \cdot \frac{1}{2}x_0 \text{ otherwise.}$$

The payoffs (x_1, x_2, x_0) can be chosen by nature N in such a way that it leads to sure loss for the bettor D if not the determinant of this system is zero:

$$0 = \begin{vmatrix} 1 - q_1 & -q_2 & 1 - 2q_0 \\ -q_1 & 1 - q_2 & 1 - 2q_0 \\ -q_1 & -q_2 & -2q_0 \end{vmatrix} = q_1 + q_2 - 2q_0.$$

Thus we must have

$$q\left(\frac{1}{2}(E_1 + E_2)\right|\tau) = \frac{1}{2}(q(E_1|\tau) + q(E_2|\tau)).$$

If $E_1 + E_2$ is an effect, the common factor $\frac{1}{2}$ can be removed by changing the likelihoods, and the result follows.

Corollary.

Assume a rational epistemic setting. Let E_1, E_2, \dots be likelihood effects in this setting, and assume that $E_1 + E_2 + \dots$ also is a likelihood effect. Then

$$q(E_1 + E_2 + \dots|\tau) = q(E_1|\tau) + q(E_2|\tau) + \dots$$

Proof. The finite case follows immediately from Theorem 6. Then the infinite case follows from monotone convergence.

The result of this section is quite general. In particular the loss function and any other criterion for the success of the experiments are arbitrary. So far I have assumed that the choice of experiment a is fixed, which implies that it is the same for A and for D . However, the result also applies to the following more general situation: Let A have some definite purpose of his experiment, and to achieve that purpose, he has to choose the question a in a clever manner, as rationally as he can. Assume that this rationality is formalized through the actor D , who has the ideal likelihood effect E and the experimental evidence $p(z|\tau) = q(E|\tau)$. If two such questions shall be chosen, the result of Theorem 6 holds, with essentially the same proof.

9. The Born formula

9.1. The basic formula

Born's formula is the basis for all probability calculations in quantum mechanics. In textbooks it is usually stated as a separate axiom, but it has also been argued for by using various sets of assumptions; see [12] for some references. Here I will base the discussion upon the result of Section 8.

I begin with a recent result by Busch [13], giving a new version of a classical mathematical theorem by Gleason. Busch's version has the advantage that it is valid for a Hilbert space of dimension 2, which Gleason's original theorem is not, and it also has a simpler proof. For a proof for the finite-dimensional case, see Appendix 5 of [2].

Let in general H be any Hilbert space. Recall that an effect E is any operator on the Hilbert space with eigenvalues in the range $[0, 1]$. A generalized probability measure μ is a function on the effects with the properties

- (1) $0 \leq \mu(E) \leq 1$ for all E ,
- (2) $\mu(I) = 1$,
- (3) $\mu(E_1 + E_2 + \dots) = \mu(E_1) + \mu(E_2) + \dots$ whenever $E_1 + E_2 + \dots \leq I$.

Theorem 7. (Busch, 2003).

Any generalized probability measure μ is of the form $\mu(E) = \text{Tr}(\sigma E)$ for some density operator σ .

It is now easy to see that $q(E|\tau) = p(z|\tau)$ on the ideal likelihood effects of Section 8 is a generalized probability measure if Assumption D holds: (1) follows since q is a probability; (2) since $E = I$ implies that the likelihood is 1 for all values of the e-variable, hence $p(z) = 1$; finally (3) is a consequence of the corollary of Theorem 6. Hence there is a density operator $\sigma = \sigma(\tau)$ such that $p(z|\tau) = \text{Tr}(\sigma(\tau)E)$ for all ideal likelihood effects $E = E(z)$.

Define now a *perfect experiment* as one where the measurement uncertainty can be disregarded. The quantum mechanical literature operates very much with perfect experiments which give well-defined states $|k\rangle$. From the point of view of statistics, if, say the 99% confidence or credibility region of θ^b is the single point u_k^b , we can infer approximately that a perfect experiment has given the result $\theta^b = u_k^b$.

In our symmetric epistemic setting then: We have asked the question: 'What is the value of the accessible e-variable θ^b ?', and are interested in finding the probability of the answer $\theta^b = u_j^b$ though a perfect experiment. This is the probability of the state $|b; j\rangle$. Assume now that this probability is sought in a context $\tau = \tau^{a,k}$ defined as follows: We have previous knowledge of the answer $\theta^a = u_k^a$ of another accessible question: What is the value of θ^a ? That is, we know the state $|a; k\rangle$. If θ^a is maximally accessible, this is the maximal knowledge about the system that τ may contain; in general we assume that the context τ does not contain more information about this system. It can contain irrelevant information, however.

Theorem 8. (Born's formula)

Assume a rational epistemic setting. In the above situation we have:

$$P(\theta^b = u_j^b | \theta^a = u_k^a) = |\langle a; k | b; j \rangle|^2.$$

Proof. Fix j and k , let $|v\rangle$ be either $|a; k\rangle$ or $|b; j\rangle$, and consider likelihood effects of the form $E = |v\rangle\langle v|$. This corresponds in both cases to a perfect measurement of a maximally accessible parameter with a definite result. By Theorem 7 there exists a density operator $\sigma^{a,k} = \sum_i \pi_i(\tau^{a,k}) |i\rangle\langle i|$ such that $q(E | \tau^{a,k}) = \langle v | \sigma^{a,k} | v \rangle$, where $\pi_i(\tau^{a,k})$ are non-negative constants adding to 1. Consider first $|v\rangle = |a; k\rangle$. For this case one must have $\sum_i \pi_i(\tau^{a,k}) |\langle i | a; k \rangle|^2 = 1$ and thus $\sum_i \pi_i(\tau^{a,k})(1 - |\langle i | a; k \rangle|^2) = 0$. This implies for each i that either $\pi_i(\tau^{a,k}) = 0$ or $|\langle i | a; k \rangle| = 1$. Since the last condition implies $|i\rangle = |a; k\rangle$ (modulus an irrelevant phase factor), and this is a condition which can only be true for one i , it follows that $\pi_i(\tau^{a,k}) = 0$ for all other i than this one, and that $\pi_i(\tau^{a,k}) = 1$ for this particular i . Summarizing this, we get $\sigma^{a,k} = |a; k\rangle\langle a; k|$, and setting $|v\rangle = |b; j\rangle$, Born's formula follows, since $q(E | \tau^{a,k})$ in this case is equal to the probability of the perfect result $\theta^b = u_j^b$.

9.2. Consequences

Here are three easy consequences of Born's formula:

- (1) If the context of the system is given by the state $|a; k\rangle$, and A^b is the operator corresponding to the e-variable θ^b , then the expected value of a perfect measurement of θ^b is $\langle a; k | A^b | a; k \rangle$.
- (2) If the context is given by a density operator σ , and A is the operator corresponding to the e-variable θ , then the expected value of a perfect measurement of θ is $\text{Tr}(\sigma A)$.
- (3) In the same situation the expected value of a perfect measurement of $f(\theta)$ is $\text{Tr}(\sigma f(A))$.

Proof of (1):

$$\begin{aligned} E(\theta^b | \theta^a = u_k^a) &= \sum_i u_i^b P(\theta^b = u_i^b | \theta^a = u_k^a) \\ &= \sum_i u_i^b \langle a; k | b; i \rangle \langle b; i | a; k \rangle = \langle a; k | A^b | a; k \rangle. \end{aligned}$$

These results give an extended interpretation of the operator A compared to what I gave in Section 5: There is a simple formula for all expectations in terms of the operator. On the other hand, the set of such expectations determine the state of the system. Also on the other hand: If A is specialized to an indicator function, we get back Born's formula, so the consequences are equivalent to this formula.

As an application of Born's formula, we give the transition probabilities for electron spin. I will, for a given direction a , define the e-variable θ^a as +1 if the measured spin component by a perfect measurement for the electron is $+\hbar/2$ in this direction, $\theta^a = -1$ if the component is $-\hbar/2$. Assume that a and b are two directions in which the spin component can be measured.

Proposition 2.

For electron spin we have

$$P(\theta^b = \pm 1 | \theta^a = +1) = \frac{1}{2}(1 \pm \cos(a \cdot b)).$$

This is proved in several textbooks, for instance [4], from Born's formula. A similar proof using the Pauli spin matrices is also given in [5].

Finally, using Born's formula, we deduce the basic formula for quantum measurement. Let the state of a system be given by a density matrix $\rho = \rho(\eta)$, where η is an unknown statistical parameter, and let the measurements be determined by an operator-valued measure $M(\cdot)$ as defined in Section 7. Then the probability distribution of the observations is given by

$$P(B; \eta) = \text{Tr}(\rho(\eta)M(B)).$$

This, together with an assumption on the state after measurement, is the basis for [9].

10. Entanglement, EPR and the Bell theorem

The total spin components in different directions for a system of two spin 1/2 particles satisfy the assumptions of a maximal symmetric epistemic setting. Assume that we have such a system where $j = 0$, that is, the state is such that the total spin is zero. By ordinary quantum mechanical calculations, this state can be explicitly written as

$$|0\rangle = \frac{1}{\sqrt{2}}(|1, +\rangle \otimes |2, -\rangle - |1, -\rangle \otimes |2, +\rangle), \quad (5)$$

where $|1, +\rangle \otimes |2, -\rangle$ is a state where particle 1 has a spin component $+\hbar/2$ and particle 2 has a spin component $-\hbar/2$ along the z -axis, and *vice versa* for $|1, -\rangle \otimes |2, +\rangle$. This is what is called an entangled state, that is, a state which is not a direct product of the component state vectors. I will follow my own programme, however, and stick to the e -variable description.

Assume further that the two particles separate, the spin component of particle 1 is measured in some direction by an observer Alice, and the spin component of particle 2 is measured by an observer Bob. Before the experiment, the two observers agree both either to measure spin in some fixed direction a or in another fixed direction b , orthogonal to a , both measurements assumed for simplicity to be perfect. As a final assumption, let the positions of the two observers at the time of measurement be spacelike, that is, the distance between them is so large that no signal can reach from one to the other at this time, taking into account that signals cannot go faster than the speed of light by the theory of relativity.

This is Bohm's version of the situation behind the argument against the completeness of quantum mechanics as posed by Einstein et al. [14] and countered by Bohr [15], [16]. This discussion is still sometimes taken up today, although most physicists now support Bohr. So will I, but I will go a step further. The main thesis in [14] was as follows: *If, without in any*

way disturbing a system, we can predict the value of a physical quantity, then there exists an element of physical reality corresponding to this physical reality. Bohr answered by introducing a strict interpretation of this criterion: To ascribe reality to P , the measurement of an observable whose outcome allows for the prediction of P , must actually be performed, or one must give a description of how it can be performed. Several authors have argued that Einstein’s criterion of reality lead an assumption of non-locality: Signals between observers with a spacelike separation must travel faster than light. Recently, it has been shown in [17] that the possibility is open to interpret the non-locality theorems in the physical literature as arguments supporting the strict criterion of reality, rather than a violation of locality. I agree with this last interpretation.

I will be very brief on this discussion here. Let λ be the spin component in units of $\hbar/2$ as measured by Alice, and let η be the spin component in the same units as measured by Bob. Alice has a free choice between measuring in the the directions a and in the direction b . In both cases, her probability is $1/2$ for each of $\lambda = \pm 1$. If she measures $\lambda^a = +1$, say, she will predict $\eta^a = -1$ for the corresponding component measured by Bob. According to Einstein et al. [14] there should then be an element of reality corresponding to this prediction, but if we adapt the strict interpretation of Bohr here, there is no way in which Alice can predict Bob’s actual real measurement at this point of time. Bob on his side has also a free choice of measurement direction a or b , and in both cases he has the probability $1/2$ for each of $\eta = \pm 1$. The variables λ and η are conceptual, the first one connected to Alice and the second one connected to Bob. As long as the two are not able to communicate, there is no sense in which we can make statements like $\eta = -\lambda$ meaningful.

The situation changes. however, if Alice and Bob meet at some time after the measurement. If Alice then says ‘I chose to make a measurement in the direction a and got the result u ’ and Bob happens to say ‘I also chose to make a measurement in the direction a , and then I got the result v ’, then these two statements must be consistent: $v = -u$. This seems to be a necessary requirement for the consistency of the theory. There is a subtle distinction here. The clue is that the choices of measurement direction both for Alice and for Bob are free and independent. The directions are either equal or different. If they should happen to be different, there is no consistency requirement after the measurement, due to the assumed orthogonality of a and b . Note again that we have an epistemic interpretation of quantum mechanics. At the time of measurement, nothing exists except the observations by the two observers.

Let us then look at the more complicated situation where a and b are not necessarily orthogonal, where Alice tosses a coin and measures in the direction a if head and b if tail, while Bob tosses an independent coin and measures in some direction c if head and in another direction d if tail. Then there is an algebraic inequality

$$\lambda^a \eta^c + \lambda^b \eta^c + \lambda^b \eta^d - \lambda^a \eta^d \leq 2. \tag{6}$$

Since all the conceptual variables take values ± 1 , this inequality follows from

$$(\lambda^a + \lambda^b) \eta^c + (\lambda^b - \lambda^a) \eta^d = \pm 2 \leq 2.$$

Now replace the conceptual variables here with actual measurements. Taking then formal expectations from (6), assumes that the products here have meaning as random variables;

in the physical literature this is stated as an assumption of realism and locality. This leads formally to

$$E(\widehat{\lambda^a \eta^c}) + E(\widehat{\lambda^b \eta^c}) + E(\widehat{\lambda^b \eta^d}) - E(\widehat{\lambda^a \eta^d}) \leq 2 \quad (7)$$

This is one of Bell's inequalities, called the CHSH inequality.

On the other hand, using quantum-mechanical calculations, that is Born's formula, from the basic state (5), shows that a, b, c and d can be chosen such that Bell's inequality (7) is violated. This is also confirmed by numerous experiments with electrons and photons.

From our point of view the transition from (6) to (7) is not valid. One can not take the expectation term by term in equation (6). The λ 's and η 's are conceptual variables belonging to different observers. Any valid statistical expectation must take one of these observers as a point of departure. Look at (6) from Alice's point of view, for instance. She starts by tossing a coin. The outcome of this toss leads to some e-variable λ being measured in one of the directions a or b . This measurement is an epistemic process, and any prediction based upon this measurement is a new epistemic process. During these processes she must obey Conditionality principle 2 of Section 2. By this conditionality principle she should condition upon the outcome of the coin toss. So in any prediction she should condition upon the choice a or b . It is crucial for this argument that the prediction of an e-variable is an epistemic process, not a process where ordinary probability calculations can be immediately used.

By doing predictions from her measurement result, she can use Born's formula. Suppose that she measures λ^a and finds $\lambda^a = +1$, for instance. Then she can predict the value of λ^c and hence $\eta^c = -\lambda^c$. Thus she can (given the outcome a of the coin toss) compute the expectation of the first term (6). Similarly, she can compute the expectation of the last term in (6). But there is no way in which she simultaneously can predict λ^b and η^d . Hence the expectation of the second term (and also, similarly the third term) in (6) is for her meaningless. A similar conclusion is reached if the outcome of the coin toss gives b . And of course a similar conclusion is valid if we take Bob's point of view. Therefore the transition from (6) to (7) is not valid, not by non-locality, but by a simple use of the conditionality principle. This can also in some sense be called lack of realism: In this situation is it not meaningful to take expectation from the point of view of an impartial observer. By necessity one must see the situation from the point of view of one of the observers Alice or Bob.

Entanglement is very important in modern applications of quantum mechanics, not least in quantum information theory, including quantum computation. It is also an important ingredient in the theory of decoherence [18], which explains why ordinary quantum effects are not usually visible on a larger scale. Decoherence theory shows the importance of the entanglement of each system with its environment. In particular, it leads in effect to the conclusion that all observers share common observations after decoherence between the system and its environment, and this can then be identified with the 'objective' aspects of the world; which is also what the superior actor D of Section 8 would find.

11. Position as an e-variable and the Schrödinger equation

So far I have looked at e-variables taking a finite discrete set of values, but the concept of an e-variable carries over to the continuous case. Consider the motion of a non-relativistic

one-dimensional particle. Its position ζ at some time t can in principle be determined by arbitrary accuracy, resulting in an arbitrarily short confidence interval. But momentum and hence velocity cannot be determined simultaneously with arbitrary accuracy, hence the vector $(\zeta(s), \dot{\zeta}(t))$ for positions at two different time points is inaccessible. Now fix some time t . An observer i may predict $\zeta(t)$ by conditioning on some σ -algebra \mathcal{P}_i of information from the past. This may be information from some time point $s_i < t$, but it can also take other forms. We must think of different observers as hypothetical; only one of them can be realized. Nevertheless one can imagine that all this information, subject to the choice of observer later, is collected in an inaccessible σ -algebra \mathcal{P}_t , the past of $\zeta(t)$. The distribution of $\zeta(t)$, given the past \mathcal{P}_t , for each t , can then be represented as a stochastic process.

In the simplest case one can then imagine $\{\zeta(s); s \geq 0\}$ as an inaccessible Markov process: The future is independent of the past, given the present. Under suitable regularity conditions, a continuous Markov process will be a diffusion process, i.e., a solution of a stochastic differential equation of the type

$$d\bar{\zeta}(t) = b(\bar{\zeta}(t), t)dt + \sigma(\bar{\zeta}(t), t)dw(t). \tag{8}$$

Here $b(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are continuous functions, also assumed differentiable, and $\{w(t); t \geq 0\}$ is a Wiener process. The Wiener process is a stochastic process with continuous paths, independent increments $w(t) - w(s)$, $w(0) = 0$ and $E((w(t) - w(s))^2) = t - s$. Many properties of the Wiener process have been studied, including the fact that its paths are nowhere differentiable. The stochastic differential equation (8) must therefore be defined in a particular way; for an introduction to Itô calculus or Stochastic calculus; see for instance [19].

So far we have considered observers making predictions of the present value $\zeta(t)$, given the past \mathcal{P}_t . There is another type of epistemic processes which can be described as follows: Imagine an actor A which considers some future event for the particle, lying in a σ -algebra \mathcal{F}_j . He asks himself in which position he should place the particle at time t as well as possible in order to have this event fulfilled. In other words, he can adjust $\zeta(t)$ for this purpose. Again one can collect the σ -algebras for the different potential actors in one big inaccessible σ -algebra \mathcal{F}_t , the future after t . The conditioning of the present, given the future, defines $\{\bar{\zeta}(t); t \geq 0\}$ as a new inaccessible stochastic process, with now t running backwards in time. In the simplest case this is a Markov process, and can be described by a stochastic differential equation

$$d\bar{\zeta}(t) = b_*(\bar{\zeta}(t), t)dt + \sigma_*(\bar{\zeta}(t), t)dw_*(t), \tag{9}$$

where again $w_*(t)$ is a Wiener process.

Without having much previous knowledge about modern stochastic analysis and without knowing anything about epistemic processes, Nelson [20] formulated his stochastic mechanics, which serves our purpose perfectly. Nelson considered the multidimensional case, but for simplicity, I will here only discuss a one-dimensional particle. Everything can be generalized.

Nelson discussed what corresponds to the stochastic differential equations (8) and (9) with σ and σ_* constant in space and time. Since heavy particles fluctuate less than light particles,

he assumed that these quantities vary inversely with mass m , that is, $\sigma^2 = \sigma_*^2 = \hbar/m$. The constant \hbar has dimension action, and turns out to be equal to Planck's constant divided by 2π . This assumes that $\sigma^2 = \sigma_*^2$, a fact that Nelson actually proved in addition to proving that

$$b_* = b - \sigma^2(\ln\rho)_x,$$

where $\rho = \rho(x, t)$ is the probability density of $\zeta(t)$.

Introduce $u = (b - b_*)/2$ and $v = (b + b_*)/2$. Then $R = \frac{1}{2}\ln\rho(x, t)$ satisfies $R_x = mu/\hbar$. Let S be defined up to an additive constant by $S_x = mv/\hbar$ and define the wave function of the particle by $f = \exp(R + iS)$. Then $|f(x, t)|^2 = \rho(x, t)$ as it should. By defining the acceleration of the particle in a proper way and using Newton's second law, a set of partial differential equations for u and v can be found, and by choosing the additive constant in S properly, one deduces from these equations

$$i\hbar \frac{\partial}{\partial t} f(x, t) = \left[\frac{1}{2m} \left(-i\hbar \frac{\partial}{\partial x} \right)^2 + V(x) \right] f(x, t), \quad (10)$$

where $V(x)$ is the potential energy. The details of these derivations can be found in [2] with more details in [20]. Identifying $-i\hbar \frac{\partial}{\partial x}$ as the operator for momentum, we see that (10) is the Schrödinger equation for the particle.

12. Conclusion

Even though the mathematics here is more involved, the approach of the present chapter (expressed in more detail in [2]) should serve to take some of the mystery off the ordinary formal introduction to quantum theory. A challenge for the future will be to develop the corresponding relativistic theory, by using representations of the Poincaré group together with an argument like that in Section 11. Also, one should seek a link to elementary particle physics using the relevant Lie group theory. Group theory is an important part of physics, and it should come as no surprise that this also is relevant to the foundation of quantum mechanics.

Author details

Inge S. Helland

* Address all correspondence to: ingeh@math.uio.no

Department of Mathematics, University of Oslo, Blindern, Oslo, Norway

13. References

- [1] Ballentine, L. E. (1998). *Quantum Mechanics. A Modern Development*. Singapore: World Scientific.
- [2] Helland, I. S. (2012). *A Unified Scientific Basis for Inference*. Submitted as a Springer Brief; arXiv:1206.5075.

- [3] Birnbaum, A. (1962). On the foundation of statistical inference. *Journal of the American Statistical Association* 57, 269-326.
- [4] Holevo A.S. (2001). *Statistical Structure of Quantum Theory*. Berlin: Springer-Verlag.
- [5] Helland, I.S. (2010). *Steps Towards a Unified Basis for Scientific Models and Methods*. Singapore: World Scientific.
- [6] Schweder, T. and Hjort, N.L. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics* 29, 309-332.
- [7] Xie, M. and Singh, K. (2012). Confidence distributions, the frequentist distribution estimator of a parameter - a review. To appear in *International Statistical Review*
- [8] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. New York: Springer.
- [9] Barndorff-Nielsen, O.E., Gill, R.D. and Jupp, P.E. (2003). On quantum statistical inference. *Journal of the Royal Statistical Society B* 65, 775-816.
- [10] Hammond, P.J. (2011). Laboratory games and quantum behaviour. The normal form with a separable state space. Working paper. University of Warwick, Dept of Economics.
- [11] Caves, C.M., Fuchs, C.A. and Schack, R. (2002). Quantum probabilities as Bayesian probabilities. *Physical Review* A65, 022305.
- [12] Helland, I.S. (2008). Quantum mechanics from focusing and symmetry. *Foundations of Physics* 38, 818-842.
- [13] Busch, P. (2003). Quantum states and generalized observables: A simple proof of Gleason's Theorem. *Physical Review Letters* 91 (12), 120403.
- [14] Einstein, A., Podolsky, B. and Rosen, N. (1935). Can quantum mechanical description of physical reality be considered complete? *Physical Review* 47, 777-780.
- [15] Bohr, N. (1935a). Quantum mechanics and physical reality. *Nature* 136, 65.
- [16] Bohr, N. (1935b). Can quantum mechanical description of physical reality be considered complete? *Physical Review* 48, 696-702.
- [17] Nisticó, G. and Sestito, A. (2011). Quantum mechanics, can it be consistent with reality? *Foundations of Physics* 41, 1263-1278.
- [18] Schlosshauer, M. (2007). *Decoherence and the Quantum-to Classical Transition*. Berlin: Springer-Verlag.
- [19] Klebaner, F.C. (1998). *Introduction to Stochastic Calculus with Applications*. London: Imperial College Press.
- [20] Nelson, E. (1967). *Dynamic Theories of Brownian Motion*. Princeton: Princeton University Press.