A classical extension of quantum mechanics

1995 J. Phys. A: Math. Gen. 28 3329

(http://iopscience.iop.org/0305-4470/28/12/007)

J. Phys. A: Math. Gen. 28 (1995) 3329–3343. Printed in the UK

A classical extension of quantum mechanics

E G Beltrametti† and S Bugajski‡

† Department of Physics, University of Genoa and Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy

‡ *Institute of Physics, University of Silesia, Uniwersytecka 4, PL 40 007 Katowice, Poland*

Received 4 November 1994, in final form 30 January 1995

Abstract. A physically natural generalization of the notion of observable that encompasses both the classical and the quantum ones is derived. Based on it, the idea of the classical extension of a theory is developed; the states of the extended theory being the probability measures on the pure states of the original one. It is shown that quantum theory admits such a classical extension, and that the qualifying features of quantum observables are preserved in the extended model.

1. Introduction

The question of possible relations between quantum and classical theories has stimulated intense research since the beginnings of quantum mechanics. Various embeddings of quantum theories into classical frameworks have been proposed without being contradicted by empirical facts; however, the recovery of classical features is in general at the cost of the abandonment of other properties that appear physically sound (for a short review of classical representations of quantum theories see, e.g., [1]).

In this paper we exploit a new kind of classical representation, or classical extension, of quantum theories. It carries into a classical context a surprisingly broad array of genuine quantum properties, suggesting that quantum theories can be viewed as specific subtheories of classical ones. This classical extension, called the delinearization approach in [1], rests on a physically natural notion of observable that encompasses both the classical and the usual quantum versions; thus the distinction between the classical and the quantum behaviour becomes coded solely in the convex structure of the set of states, which is a simplex in the first case, while in the second case we have the non-unique decomposition of mixtures into pure states. Our classical extension of a quantum theory involves the fact that starting from a quantum set of states, say S_Q, it is always possible to construct a new simplectic (hence classical) structure of states, consisting of the set of the probability measures on the pure elements of S_Q, and a mapping from it onto S_Q—already studied by Misra in 1974 [2]—that carries the classical extension under discussion. The usual quantum observables, when looked upon in this classical extension, appear as fuzzy classical ones.

If compared with the 'phase-space representation' of quantum theory (see [3]; for a concise review and recent results see [1, 4]), our approach shares the idea of a classical structure of states, but its relationship with the original quantum set of states S_Q is quite different, for in the phase-space representation we do not deal with a mapping of a classical structure of states onto S_Q but rather with a mapping of S_Q into the set of probability measures on the phase space.

0305-4470/95/013329+15$19.50

2. Observables

Let S be a set representing, in a given theoretical model, the set of all states of the physical system under consideration. We assume S to be convex, for this property translates the basic physical operation of forming statistical mixtures of states. If $\alpha_1, \alpha_2 \in S$ and $\lambda \in [0, 1]$ we write $\lambda\alpha_1 + (1 - \lambda)\alpha_2$ for the mixture of α_1 and α_2 with weights λ and $1 - \lambda$.

The intuitive physical notion of observable, or physical quantity, consists of the specification of the possible outcomes (e.g., real numbers in proper units for the most common case) and of their probability distribution for each state of the physical system. To give this notion mathematical clothing, let Ξ be a measurable space and write $\mathcal{B}(\Xi)$ for the Boolean σ-algebra of the measurable subsets of Ξ; we shall then define an observable, in the given theoretical model, as an affine map of S into the convex set $M_1^+(\Xi)$ of the probability measures on Ξ. Typically, we shall have for Ξ the real line $\mathbb{R}$, but also $\Xi=\mathbb{R}^n$ might occur when dealing, for instance, with vector-valued or with joint observables. All the singletons of Ξ will be assumed to be measurable.

Despite its naturalness, the above definition of observable is not frequently used in the literature; it was made explicit by Holevo [5] in studying probabilistic and statistical aspects of quantum theory, and one can find it applied in works on the phase space representation of quantum theories [1,3,4] or in studies on more general frameworks [5,6]. More specific and structured notions of observables are commonly used; as we shall see, they are encompassed by our definition.

Let $B : S \to M_1^+(\Xi)$ be an observable; to any pair $(\alpha, X), \alpha \in S, X \in \mathcal{B}(\Xi)$, B associates the real number $(B\alpha)(X) \in [0, 1]$, i.e. the value the measure $B\alpha$ takes at the set X. For fixed $X \in \mathcal{B}(\Xi)$ we get an affine function $E_{B,X}$ from S into [0,1], hence an *effect*, according to a common terminology.

Notice that the effects, namely the affine functions from S into [0,1], form a poset under the pointwise ordering; if a_1, a_2 are effects we say that $a_1 \leqslant a_2$ whenever $a_1(\alpha) \leqslant a_2(\alpha)$ for all $\alpha \in S$. We write 0_S for the least effect (the null function on S) and e_S for the greatest effect (the unit function on S); we denote by $[0_S, e_S]$ the set of all effects on S, and notice from now that it is naturally endowed with a convex structure, for the convex combination of two effects is obviously an effect.

The observables are thus associated to effect-valued measures on $\mathcal{B}(\Xi)$, and actually they can be identified as such, because any effect-valued measure on $\mathcal{B}(\Xi)$ clearly determines an affine map of S into $M_1^+(\Xi)$, hence an observable. We shall call semi-spectral resolution of the observable B its effect-valued measure E_B on $\mathcal{B}(\Xi)$, so transferring a terminology of functional analysis to the present more general context.

Let us now list a number of familiar notions that can be carried over by the above definition of observable.

(i) *Spectrum.* The spectrum of the observable $B : S \to M_1^+(\Xi)$, denoted $\mathrm{Sp}B$, can be defined if $\mathcal{B}(\Xi)$ is generated in the standard way by a topology on Ξ (as in the case $\Xi = \mathbb{R}$). Then $\mathrm{Sp}B$ is the smallest among the closed subsets of Ξ such that $E_{B,\mathrm{Sp}B} = e_S$. Intuitively, the spectrum of B is the smallest subset of Ξ that contains all possible values of B.

(ii) *Expectation value.* Assuming for Ξ some linear structure (take, e.g., $\Xi = \mathbb{R}$), let $\mu \in M_1^+(\Xi)$, and define as usual its expectation by

$$Exp(\mu) := \int_\Xi \xi \, \mathrm{d}\mu(\xi) \tag{1}$$

provided the integral exists (which is ensured if Ξ is bounded). If μ is the image of a state $\alpha \in S$ under the observable B, we call the quantity in (1) the expectation value (or mean value) of the observable B at the state α and we favour the notation $Exp(B, \alpha)$.

(iii) *Variance.* With the same assumption as in (ii), if $\mu \in M_1^+(\Xi)$ we define its variance as

$$Var(\mu) := \int_{\Xi} [\xi - Exp(\mu)]^2 \, d\mu(\xi) \tag{2}$$

provided the integral exists. Again, if μ is the image of $\alpha \in S$ under the observable B we call the quantity in (2) the variance of the observable B at the state α and denote it $Var(B, \alpha)$.

(iv) *Eigenstates and eigenvalues.* An eigenstate of the observable $B : S \to M_1^+(\Xi)$ is a state $\alpha \in S$ which is mapped by B into a probability measure concentrated at some $\lambda \in \Xi$. We write δ_λ for such a probability measure and call it a Dirac measure. λ is said to be an eigenvalue of B.

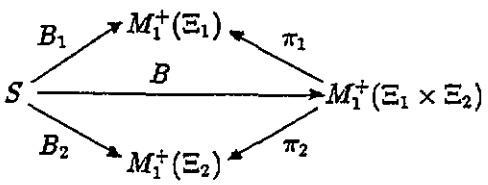
(v) *Sharpness.* It may happen that the semi-spectral decomposition of an observable B involves only effects that belong to the family $\partial[0_S, e_S]$ of the extremal elements of the convex set $[0_S, e_S]$. In this case the observable is said to be sharp.

(vi) *Uncertainty relations.* We say that two observables $B_1 : S \to M_1^+(\Xi_1)$, $B_2 : S \to M_1^+(\Xi_2)$ obey an uncertainty relation (or are complementary) if there is a positive h such that

$$Var(B_1, \alpha) Var(B_2, \alpha) \geqslant h \tag{3}$$

for all α's the two variances exist at.

(vii) *Comeasurability.* Two observables $B_1 : S \to M_1^+(\Xi_1)$, $B_2 : S \to M_1^+(\Xi_2)$ are called comeasurable if there exists a third observable $B : S \to M_1^+(\Xi_1 \times \Xi_2)$ such that $B_1 = \pi_1 \circ B$, $B_2 = \pi_2 \circ B$, where $\Xi_1 \times \Xi_2$ is the measurable-space product of Ξ_1, Ξ_2, while π_1 and π_2 are the marginal projections of $M_1^+(\Xi_1 \times \Xi_2)$ onto $M_1^+(\Xi_1)$ and $M_1^+(\Xi_2)$ respectively. Let us recall that a marginal projection, say π_1, of $M_1^+(\Xi_1 \times \Xi_2)$ is defined by $(\pi_1 \mu)(X) := \mu(X \times \Xi_2)$ for any $\mu \in M_1^+(\Xi_1 \times \Xi_2)$ and $X \in \mathcal{B}(\Xi_1)$. The observable B above is called the joint observable of B_1 and B_2, while $B\alpha$, $\alpha \in S$, is traditionally called the joint probability distribution of B_1 and B_2 at α. Pictorially, B_1, B_2 are comeasurable if there exists a B that makes commutative the diagram



Notice that the joint observable of B_1 and B_2 need not be unique, as one can check with specific examples. This non-uniqueness is related to the known fact that two (marginal) distributions do not determine a unique joint distribution.

We come now to the fact that our notion of observable accomodates both the quantum and the classical case.

The quantum case. We have to take for S the set S_Q of all density operators on a separable complex Hilbert space $\mathcal{H}$. It is a known result (see, e.g., [8, theorem VI.26]) that the effects on S_Q, i.e. the affine functions from S_Q into [0,1], are in one-to-one correspondence with the positive operators of $\mathcal{H}$ which have mean value at every state not bigger than one: explicitly, if $\mathcal{P}$ is such an operator and $D \in S_Q$, then the effect associated to $\mathcal{P}$ is the function $S_Q \to [0, 1]$ defined by $\mathrm{Tr}(D\mathcal{P})$. This fact is commonly referred to by saying that the observables are the POV-measures.

In the standard formulation of quantum mechanics one takes the self-adjoint operators on $\mathcal{H}$ as representatives of the 'observables'. Self-adjoint operators are known to correspond to projection-valued measures (in short, PV-measures) on $\mathcal{B}(\mathbb{R})$, and PV-measures are obviously a sub-class of POV-measures: thus the usual 'observables' of quantum mechanics are recovered as a particular case of the observables defined more generally as the affine functions from S_Q into $M_1^+(\mathbb{R})$. The usual 'observables' of quantum mechanics also admit a characterization in terms of the convex structure of the set $[0_{S_Q}, e_{S_Q}]$ of the effects on S_Q: in fact, the effects associated to projection operators of $\mathcal{H}$ are proved (see, e.g., [9, p 19]) to be the extremal elements of the convex set $[0_{S_Q}, e_{S_Q}]$. Summing up, we have that our definition of observables, when referred to the set S_Q of quantum states, gives—among others—the POV-measures on $\mathcal{B}(\mathbb{R})$; if we pick up the sharp observables, that is the ones whose effects are elements of $\partial[0_{S_Q}, e_{S_Q}]$, we get the observables of the standard formulation of quantum mechanics, i.e. the ones associated with self-adjoint operators.

Let us remark that the need to go beyond self-adjoint operators was discovered in the early seventies in the framework of the quantum theory of open systems, and has found a general assessment in the so-called operational approach [9–12].

The classical case. Now we have to take for S the simplex $M_1^+(\Omega)$ of all probability measures on some measurable space Ω, the 'phase space' of the physical system under discussion, whose elements can be thought of as the pure states (we assume the singletons of Ω to be measurable). Indeed, by taking $S = M_1^+(\Omega)$ we meet the most essential property of classical (as opposed to quantum) physical systems: the unique decomposability of mixed states into pure states. According to our definition, the observables are now the affine mappings of $M_1^+(\Omega)$ into $M_1^+(\Xi)$.

On the other hand, in classical statistical mechanics the physical quantities are commonly represented by measurable functions on Ω with values in Ξ. Let $f : \Omega \to \Xi$ be one of such functions and define $B_f : M_1^+(\Omega) \to M_1^+(\Xi)$ by

$$(B_f \nu)(X) := \nu(f^{-1}(X)) \tag{4}$$

where $\nu \in M_1^+(\Omega)$, $X \in \mathcal{B}(\Xi)$, and $f^{-1}(X)$ is the counter-image of X under f. It is easily seen that B_f is affine; hence it is an observable according to the more general definition here proposed. We come to the conclusion that the usual observables of classical statistical mechanics are recovered as a particular case.

We are going to examine in some more detail the framework based on the set of states $M_1^+(\Omega)$, the reason being that it is precisely a framework like this that hosts the classical extension of quantum mechanics to be discussed in the coming sections. It is also useful for a better understanding of which place, inside that framework, is occupied by the usual classical statistical mechanics.

Since $a_{\chi_\emptyset}$ and a_{χ_Ω} are respectively the null and the unit functions on $M_1^+(\Omega)$, we shall use the simplified notation $[\chi_\emptyset, \chi_\Omega]$ for the set of all effects on $M_1^+(\Omega)$: this avoids the cumbersome replacement of S by $M_1^+(\Omega)$ in the standard notation $[0_S, e_S]$. A special role

is played by the subset of $[\chi_\emptyset, \chi_\Omega]$ formed by those effects that correspond to measurable functions from Ω into $[0, 1]$ according to the following prescription: given a measurable function $g : \Omega \to [0, 1]$, take the function $a_g : M_1^+(\Omega) \to [0, 1]$ defined by

$$a_g(\nu) := \int_\Omega g(\omega)\, d\nu(\omega) \tag{5}$$

and notice that it is an affine function, hence an effect. We call the effects so generated regular; correspondingly, we call an observable $B : M_1^+(\Omega) \to M_1^+(\Xi)$ regular when its effects $E_{B,X}$ are regular for every $X \in \mathcal{B}(\Xi)$.

It is easily seen that the regular effects form (under pointwise ordering) a distributive lattice with meet and join defined by

$$a_g \wedge a_h = a_i \qquad i(\omega) = \min\{g(\omega), h(\omega)\}$$

$$a_g \vee a_h = a_l \qquad l(\omega) = \max\{g(\omega), h(\omega)\}$$

for all $\omega \in \Omega$. This distributive lattice is, however, not a Boolean algebra since the mapping $a_g \mapsto e - a_g = a_{1-g}$ provides only a quasi-complementation (notice that $a_g \vee (e - a_g)$ can be less than e). The regular effects thus form a quasi-Boolean algebra. Incidentally, it can be mentioned that even the set $[\chi_\emptyset, \chi_\Omega]$ forms a distributive, not orthocomplemented, lattice [13].

The effects associated to the observables of the usual classical statistical mechanics are particular cases of regular effects. In fact the effect $E_{B_f,X} : M_1^+(\Omega) \to [0, 1]$ associated to the observable B_f—as defined by (4) above—and to $X \in M_1^+(\Xi)$ takes the form

$$E_{B_f,X}(\nu) = \int_\Omega \chi_{f^{-1}(X)}(\omega)\, d\nu(\omega) \tag{6}$$

where χ_Y is the characteristic function of $Y \in \mathcal{B}(\Xi)$. Thus $E_{B_f,X}$ is the regular effect that corresponds to the characteristic function of the counter-image of X under f.

Now we have the following theorem.

Theorem 1. A regular effect is an extremal element of the convex set $[\chi_\emptyset, \chi_\Omega]$ if and only if it comes from a characteristic function χ_Y for some $Y \in \mathcal{B}(\Omega)$.

Proof. Suppose $a_{\chi_Y} = \lambda a + (1 - \lambda)b$ for some $a, b \in [\chi_\emptyset, \chi_\Omega]$, $0 < \lambda < 1$. Then, for any $\nu \in M_1^+(\Omega)$, the equality $\nu(Y) = 1$ implies $a(\nu) = b(\nu) = 1$, while $\nu(Y) = 0$ implies $a(\nu) = b(\nu) = 0$. Since ν admits a unique decomposition $\nu = \nu(Y)\nu_1 + (1 - \nu(Y))\nu_2$, where ν_1 is concentrated at Y and ν_2 at $\Omega \backslash Y$, we have that $a(\nu_1) = b(\nu_1) = 1$ while $a(\nu_2) = b(\nu_2) = 0$. Then the affinity of the effects implies $a(\nu) = b(\nu) = \nu(Y)$ for every $\nu \in M_1^+(\Omega)$, which means that $a = b = a_{\chi_Y}$; thus a_{χ_Y} has to be an extreme effect. Conversely, let a_g be a regular extreme effect generated by a measurable function g on Ω according to equation (5). Assume that g is not a characteristic function: then there is a point ω_0 of Ω such that $\varepsilon := g(\omega_0) \neq 0, 1$. Now define two measurable functions, say g' and g'', on Ω which differ from g only at the point ω_0, where they take the values $\varepsilon', \varepsilon'' \in [0, 1]$, with $\varepsilon' < \varepsilon < \varepsilon''$, such that $\varepsilon = \lambda\varepsilon' + (1 - \lambda)\varepsilon''$. So the function g would be decomposed into a convex combination of g' and g'', which means that a_g is not extreme. □

In view of the above theorem we see that classical statistical mechanics makes use of a limited sub-class of all possible observables: the sub-class of the regular and sharp ones. We have also that the regular and sharp effects form a Boolean algebra isomorphic to $\mathcal{B}(\Omega)$, embedded in the quasi-Boolean algebra formed by all regular effects.

Notice that the regular effects need not exhaust, for uncountable Ω, all possible effects [4, 14]; correspondingly, the effects coming from characteristc functions need not exhaust $\partial[\chi_\emptyset, \chi_\Omega]$.

Let us now come to another property which will be relevant in the remainder.

Theorem 2. Any two regular observables $B_1 : M_1^+(\Omega) \to M_1^+(\Xi_1)$, $B_2 : M_1^+(\Omega) \to M_1^+(\Xi_2)$ are comeasurable.

Proof. Let Ξ denote the product $\Xi_1 \times \Xi_2$. If X_1 and X_2 are measurable subsets of Ξ_1 and Ξ_2 respectively, define a real-valued measurable function on Ω by $f_{X_1 \times X_2}(\omega) := (B_1\delta_\omega)(X_1)\,(B_2\delta_\omega)(X_2)$ and extend it over $M_1^+(\Omega)$ by integration. In this way any measure ν on Ω defines a function on the semi-ring of all rectangles $X_1 \times X_2$ which extends to a measure on Ξ by standard procedures (see, e.g., [15, theorem 11.3]). The map $B_1 \times B_2$ of $M_1^+(\Omega)$ into $M_1^+(\Xi)$ so obtained is affine, thus being an observable on $M_1^+(\Omega)$. It is easy to check that $B_1 \times B_2$ is a joint observable for regular B_1 and B_2. □

The generalization of classical statistical mechanics beyond the regular sharp observables is, to the authors' knowledge, a rather unexplored issue, in spite of the natural way it arises. In sections 4 and 5 we shall see that it is the framework within which a classical extension of a non-classical theory occurs. One might also guess, in analogy with the role of unsharp observables in the quantum context, that the above generalization of classical statistical mechanics could be appropriate to describe classical open systems. It might even embody the description of a lack of precision in the measurement: an observable that maps Dirac measures of $M_1^+(\Omega)$ into diffuse measures of $M_1^+(\Xi)$ is a natural tool for accounting for errors in measuring the corresponding physical quantity.

Before closing this section let us emphasize a striking similarity between the cases of standard quantum mechanics and classical statistical mechanics: both theories work with a set of observables which is a narrow sample of the set of all potentially possible ones; in both cases the observables used are sharp.

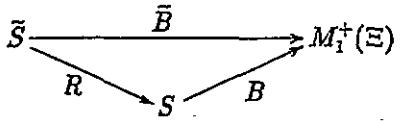
3. Extension of a state observable structure

Suppose that two theoretical models are given for the description of a physical system, both fitting the notion of observable discussed in the previous section. Let one model be based on the convex set S of states, and the other on the convex set $\widetilde{S}$. We say that the model based on $\widetilde{S}$ is an extension of the one based on S if there exists an affine surjective mapping $R : \widetilde{S} \to S$, to be called the reduction map.

As a familiar example consider, in the framework of standard quantum mechanics, a compound system and a subsystem of it. We can describe the subsystem either on the basis of the density operators of its own Hilbert space or in terms of the density operators of the Hilbert space pertaining to the compound system; the second descriptive model is an extension of the first one and the partial trace provides the reduction map. As another example think of a spin-$\frac{1}{2}$ particle and compare the description that accounts only for the spin coordinates with the description that also includes the position degrees of freedom; the

latter provides an extension of the former. Actually, the above definition of model extension fits the notion of 'coarse graining' [16]. Indeed, since R maps $\widetilde{S}$ onto S, it determines a partition of $\widetilde{S}$ into equivalence classes, in the sense that all elements of $\widetilde{S}$ having the same image in S form an equivalence class; in other words, the counter-image of R draws a coarse graining in $\widetilde{S}$.

Let us focus attention on the observables of the S-based model and on the ones of its $\widetilde{S}$-based extension. Loosely speaking, $\widetilde{S}$ being 'richer' than S, we expect to have 'more' observables on $\widetilde{S}$ than on S. Actually we have that every observable on S has a representative on $\widetilde{S}$ through the reduction map. In fact, if $B : S \to M_1^+(\Xi)$ is an observable, then the map composition $B \circ R : \widetilde{S} \to S \to M_1^+(\Xi)$ depicted in the following diagram



is an observable on $\widetilde{S}$, to be denoted $\widetilde{B}$. The observables on $\widetilde{S}$ which are representatives of the ones on S share a surprisingly wide array of properties of the original observables, in spite of the fact that they belong to another model. We shall outline the main invariants between corresponding observables.

(1) If $B\alpha$ is the probability measure on Ξ describing the statistical distribution of results of measurements of B at α, predicted by the S-based model, then $\widetilde{B}\widetilde{\alpha} = B\alpha$ for every $\widetilde{\alpha}$ in the counter-image of α under R; thus, the extended model based on $\widetilde{S}$ predicts the same statistical distribution of results of measurements of $\widetilde{B}$ at $\widetilde{\alpha}$. In short, $BS = \widetilde{B}\widetilde{S}$ for every $B : S \to M_1^+(\Xi)$. In particular, two corresponding observables B, $\widetilde{B}$ have the same spectrum, and the same eigenvalues (see (i), (iv) of section 2): intuitively, this means that both the original observable B on S and its representative $\widetilde{B}$ on $\widetilde{S}$ have the same set of possible outcomes. Moreover, if the observable B on S has expectation $Exp(B, \alpha)$ and variance $Var(B, \alpha)$ at the state $\alpha \in S$ (see (ii), (iii) of section 2), then

$$Exp(\widetilde{B}, \widetilde{\alpha}) = Exp(B, \alpha) \qquad Var(\widetilde{B}, \widetilde{\alpha}) = Var(B, \alpha) \tag{7}$$

for every $\widetilde{\alpha}$ in the counter-image of α under R.

(2) Let B_1, B_2 be two observables on S that satisfy the uncertainty relation

$$Var(B_1, \alpha)Var(B_2, \alpha) \geqslant h$$

for all $\alpha \in S$ the two variances exist at. In view of equation (7) and taking note that R maps $\widetilde{S}$ onto S, we have that $\widetilde{B}_1 := B_1 \circ R$, $\widetilde{B}_2 := B_2 \circ R$ satisfy the uncertainty relation

$$Var(\widetilde{B}_1, \widetilde{\alpha})Var(\widetilde{B}_2, \widetilde{\alpha}) \geqslant h \tag{8}$$

for all $\widetilde{\alpha} \in \widetilde{S}$ at which the two variances exist, with the same h. In other words, the uncertainty relations are invariant under the representation of the observables on S by observables on $\widetilde{S}$.

(3) If two observables B_1, B_2 on S are comeasurable (see (vii) of section 2), then also their representatives $\widetilde{B}_1 := B_1 \circ R$ and $\widetilde{B}_2 := B_2 \circ R$ on $\widetilde{S}$ are comeasurable. Indeed, if B is a joint observable of B_1 and B_2, then $\widetilde{B} := B \circ R$ is clearly a joint observable of $\widetilde{B}_1$ and $\widetilde{B}_2$. Notice, however, that this argument cannot be reversed; as we shall see in the next sections, it may happen that two observables on S which are not comeasurable have comeasurable

representatives on $\widetilde{S}$. This phenomenon shows that the extension procedure pushes toward 'less quantal', or 'more classical' models. Indeed, we shall see that any quantum-like set S of states admits an extension in which the set $\widetilde{S}$ is of classical nature, and all observables on it are comeasurable.

Let us finally remark that sharpness (see (v) of section 2) need not be preserved when we move from an observable B on S to its representative $\widetilde{B}$ on $\widetilde{S}$. Counter-examples will emerge in the remainder.

4. The canonical classical extension

Suppose that a model based on the convex set S of states admits an extension based on the convex set $\widetilde{S}$. We say that the extension is classical if $\widetilde{S}$ is a simplex. We speak of the canonical classical extension if $\widetilde{S}$ consists of all the probability measures on the set ∂S of the pure states (i.e. the extremal elements) of S; in other words, the S-based model admits the canonical classical extension if there exists an affine surjective map

$$R_M : M_1^+(\partial S) \to S. \tag{9}$$

Of course, this extension is relevant only if S does not have a classical nature on its own, in which case (see the description of the classical case given in section 2) the existence of the canonical classical extension would become tautological.

We have attached the label M to the reduction map in (9) to remind us that this is the map studied by Misra [2] in the particular case in which S is the set S_Q of all density operators on a Hilbert space. The above notion of classical extension was already advanced by Holevo [5] in a more specific context.

Of course, to speak of $M_1^+(\partial S)$ we require ∂S to be a measurable space, but this does not imply additional assumptions on S. In fact, we can generate a Boolean σ-algebra $\mathcal{B}(\partial S)$ of measurable subsets of ∂S via the effects on S. If X is any measurable subset of $[0,1]$ and $a \in [0_S, e_S]$ is an effect on S, consider the counter-image of X under a and intersect it with ∂S; then define $\mathcal{B}(\partial S)$ as the smallest σ-algebra of subsets of ∂S containing the family

$$\left\{a^{-1}(X) \cap \partial S : a \in [0_S, e_S],\, X \in \mathcal{B}([0,1])\right\}. \tag{10}$$

When ∂S is equipped with this σ-algebra of measurable subsets the restriction of any effect $a : S \to [0,1]$ to ∂S is, by construction, a measurable function, say $f_a : \partial S \to [0,1]$.

We now come to the conditions on S that mirror the existence of the reduction map (9). Intuitively, a probability measure ν on ∂S corresponds to a convex combination of pure states; thus we are led to write

$$R_M(\nu) = \int_{\partial S} \alpha \, \mathrm{d}\nu(\alpha) \tag{11}$$

and recognize as the key condition for the surjectivity of R_M the fact that every (non-pure) state of S be expressible as a mixture of pure states. While the right-hand side of equation (11) is unambiguous when ν is concentrated at a finite number of points of ∂S, it needs some care in the general case. It would be natural to take it as the weak integral [17, p 10], i.e. as the function which attaches to any effect a on S the result of integrating f_a (as defined above) with respect to ν. But to reproduce all points of S we would need the effects to be uniquely defined by their restrictions on ∂S, a condition that implies [18] that S consists of countable convex combinations of its pure states (as it holds true in standard quantum mechanics). Indeed, we have the following theorem.

Theorem 3. If S is the set of all countable convex combinations of pure states then there exists an affine surjective map $R_M : M_1^+(\partial S) \to S$, which becomes a one-to-one correspondence when restricted to the extremal elements.

Proof. To $\nu \in M_1^+(\partial S)$ we associate the function $\widehat{\nu} : [0_S, e_S] \to [0, 1]$ defined by

$$\widehat{\nu} := \int_{\partial S} f_a(\alpha)\, \mathrm{d}\nu(\alpha)$$

where $a \in [0_S, e_S]$ and f_a is the restriction of a to ∂S. It is evident that $\widehat{\nu}$ is affine. Notice that to any $\alpha \in S$ we can similarly associate the function $\widehat{\alpha} : [0_S, e_S] \to [0, 1]$ defined by $\widehat{\alpha}(a) := a(\alpha)$, for all $a \in [0_S, e_S]$, and α is uniquely determined by $\widehat{\alpha}$ since $[0_S, e_S]$ is separating on S. From basic theorems of functional analysis ([19, theorem 1.18] [8, theorem IV.20]) it follows that every $\widehat{\nu}$ defines an element of S, say α_ν, uniquely determined by the equality $a(\alpha_\nu) = \widehat{\nu}(a)$ for all $a \in [0_S, e_S]$. We have in this way generated a map $R_M : M_1^+(\partial S) \to S$, which is easily seen to be affine. Denoting by δ_α the Dirac measure concentrated at the pure state $\alpha \in \partial S$, we get, as a special case, $R_M(\delta_\alpha) = \alpha$. Since R_M is affine, any $\nu \in M_1^+(\partial S)$ consisting of a convex combination of a finite number of Dirac measures is mapped in the element of S formed by the same convex combination of the pure states associated to those Dirac measures. This conclusion holds true even in the case of countably infinite convex combinations, provided they are naturally understood as the weak limit of the sequence of finite convex combinations. Since $M_1^+(\partial S)$ includes all countable convex combinations of Dirac measures, we see that the hypothesis of the theorem ensures that any $\alpha \in S$ is the image of some element of $M_1^+(\partial S)$, which implies the surjectivity of R_M. We have still to show that only Dirac measures are mapped into pure states by R_M. Let $\alpha \in \partial S$ and $\nu \in R_M^{-1}(\alpha)$ (with some abuse of notation we use the same symbol to denote an element of a set and the subset formed by that element alone), and notice that, should it be $\nu = \lambda\nu_1 + (1 - \lambda)\nu_2$ for some $\nu_1, \nu_2 \in M_1^+(\partial S)$ and $0 < \lambda < 1$, we would have $R_M(\nu) = \lambda R_M(\nu_1) + (1 - \lambda)R_M(\nu_2) = \alpha$, hence $R_M(\nu_1) = R_M(\nu_2) = \alpha$ because α is pure. We claim now that, for any given $a \in [0_S, e_S]$, the probability measure ν has to be concentrated at $f_a^{-1}(a(\alpha))$. In fact, if this were not the case ν would contain a convex component, say ν_2, concentrated at $f_a^{-1}([0, 1]\backslash a(\alpha))$, which contradicts the abovementioned property $R_M(\nu_2) = \alpha$, or $a \circ R_M(\nu_2) = a(\alpha)$. Thus ν is concentrated at $f_a^{-1}(a(\alpha))$, and since this conclusion must hold for every $a \in [0_S, e_S]$ we conclude that ν is concentrated at

$$\cap \left\{ f_a^{-1}(a(\alpha)) : a \in [0_S, e_S] \right\}$$

which equals α because $[0_S, e_S]$ separates the elements of S. Thus $\nu = \delta_\alpha$. □

It may happen that the convex structure of S is such that a non-pure state has a convex decomposition into pure states but this decomposition is non-unique: in these circumstances there are different convex combinations of pure states that correspond to the same element of S (as is well known, this is the case with S_Q), so the correspondence between $M_1^+(\partial S)$ and S becomes many-to-one.

Summing up, we have seen that a physical theory based on the convex set S of states admits the canonical classical extension if, loosely speaking, the non-pure states are mixtures of pure states. It is worth stressing that the construction of the canonical classical extension is uniquely defined by the fundamental convex structure of S and does not depend on any particular realization we might have for S.

Let us now shift attention to the observables of the S-based model and their representatives in the canonical classical extension. A first relevant fact is specified by the next theorem.

Theorem 4. If $B : S \to M_1^+(\Xi)$ is an observable of the original S-based model, then its representative $\widetilde{B} := B \circ R_M : M_1^+(\partial S) \to M_1^+(\Xi)$ is regular.

Proof. For any $X \in \mathcal{B}(\Xi)$ the effect $E_{\widetilde{B},X}$ acts on $\nu \in M_1^+(\partial S)$ according to

$$E_{\widetilde{B},X}(\nu) = E_{B,X} \circ R_M(\nu) = \int_{\partial S} E_{B,X} \circ R_M(\delta_\alpha)\, \mathrm{d}\nu(\alpha) = \int_{\partial S} E_{B,X}(\alpha)\, \mathrm{d}\nu(\alpha) \tag{12}$$

where the Dirac measure δ_α is the counter-image of $\alpha \in \partial S$ under R_M. Upon inspection of equation (5), and noticing that the restriction of $E_{B,X}$ to ∂S is a measurable function, we conclude that $E_{\widetilde{B},X}$ is regular. □

Combining the above result with the content of theorem 2 we come to the following corollary.

Corollary. The representatives, in the canonical classical extension, of any two observables of the original S-based model are comeasurable.

We come now to the problem of whether sharpness is preserved by the canonical classical extension. In view of theorem 4 we end up with regular observables on $M_1^+(\partial S)$, and we shall say that an observable is fuzzy when it is regular but non-sharp. We have now the following theorem.

Theorem 5. If there exists a pure state of S having dispersion on the observable $B : S \to M_1^+(\Xi)$ then the representative of B in the canonical classical extension is fuzzy.

Proof. First note that composing the effect $E_{B,X}$ on S with R_M we get an effect $E_{B,X} \circ R_M$ on $M_1^+(\partial S)$, and

$$E_{B,X}(\alpha) = E_{B,X} \circ R_M(\delta_\alpha) \tag{13}$$

for every $\alpha \in \partial S$. Should $\widetilde{B} := B \circ R_M$ be sharp, we would have that $E_{B,X} \circ R_M$ is an extreme effect for all $X \in \mathcal{B}(\Xi)$; but theorems 4 and 1 would imply that $E_{B,X} \circ R_M$ is generated by the characteristic function χ_Y for some measurable subset Y of ∂S. Hence, by equation (13), $E_{B,X}(\alpha) = 0, 1$ for any $X \in \mathcal{B}(\Xi)$ and $\alpha \in \partial S$; this means that the pure states of S would be dispersion-free on B, thus contradicting the hypothesis of the theorem. □

Theorem 4, together with the comments made in section 2 about the classical case, ensures that the poset $[0_S, e_S]$ of the effects of the original S-based model can be embedded in the quasi-Boolean algebra of the regular effects of the extended classical model based on $M_1^+(\partial S)$. However, when the premise of theorem 5 is met (as is the case for quantum mechanics) the possibility of embedding $[0_S, e_S]$ in the Boolean algebra of the regular extremal effects on $M_1^+(\partial S)$ is ruled out. Of course this possibility is ruled out even when we restrict to the set $\partial[0_S, e_S]$ of the extremal effects, which plays the role of the 'quantum logic' of the S-based model whenever the latter has a quantum nature [6, 9].

5. The extension of quantum theory

We now come to the specific case of quantum mechanics, based on the convex set S_Q of all density operators on a complex separable Hilbert space $\mathcal{H}$, the set ∂S_Q of the extremal elements consisting of the one-dimensional projectors.

The conditions for the existence of the canonical classical extension discussed in the previous section, more specifically the conditions of theorem 3, are met by S_Q. Thus we come to our main conclusion that there exists an affine surjective map

$$R_M : M_1^+(\partial S_Q) \to S_Q \tag{14}$$

which carries the canonical classical extension of quantum mechanics.

The decomposition of a density operator into a convex combination of one-dimensional projectors (the pure states of S_Q) is never unique: there are infinitely many distinct convex combinations of pure states that give rise to the same density operator (see, e.g., [20]). So, the coarse graining on $M_1^+(\partial S_Q)$ associated with the reduction map (14) corresponds to collapsing the family of all convex combinations of pure states that correspond to the same density operator into that density operator.

Besides the quoted papers of Misra [2] and Holevo [5], the use of probability measures on ∂S_Q to represent states was considered by Ghirardi, Rimini, and Weber [21] and Hudson [22, 23]. Recently it was applied by Amann [24] as a basis for his 'individual-stochastic interpretation' of quantum mechanics.

Going through the results of sections 3 and 4 we can summarize a number of facts about observables. For short we call Q-observables the ones on S_Q; as seen in section 2 they include the usual observables described by self-adjoint operators on $\mathcal{H}$, as well as the unsharp ones associated with POV-measures. We call C-representatives their counterparts in the canonical classical extension, namely their representatives on $M_1^+(\partial S_Q)$.

(1) The statistical distribution of results of measurements of a Q-observable is the same as the statistical distribution of results of measurements of its C-representative; in short, $BS_Q = \widetilde{B}M_1^+(\partial S_Q)$ for any Q-observable B. In particular, a Q-observable and its C-representative have the same spectrum and the same eigenvalues; moreover, expectation values and variances are unchanged going from Q-observables to their C-representatives.

(2) Whenever two Q-observables obey an uncertainty relation their C-representatives do the same (with the same uncertainty limit).

(3) The C-representatives of the Q-observables are regular and fuzzy.

(4) Any two Q-observables have comeasurable C-representatives.

To visualize some aspects of the canonical classical extension of quantum mechanics, consider the description of the spin-$\frac{1}{2}$. The convex set S_Q can be viewed as a unit sphere in three dimensions (see, e.g., [20]): any two diametrically opposed points on the surface represent the 'up' and 'down' polarization (pure) states along some direction in ordinary space. The non-unique decomposability of non-pure states into pure states is geometrically apparent: for instance, the mixture with equal weights of the 'spin-up' and 'spin-down' states along a given axis represents the unpolarized state, but the choice of the axis is completely arbitrary, so that the (degenerate) density operator of the unpolarized state admits infinitely many convex decompositions into pure states.

Consider the usual spin observables σ_y, σ_z along the y, z axes. As depicted in figure 1, both σ_y and σ_z map the sphere of states into the probability measures on a two-element set, hence into a segment whose extremal points are the Dirac measures corresponding to the two eigenvalues -1 and 1. The spin-up state in the (θ, ϕ) direction, whose wavefunction

can be written as

$$\psi(\theta,\phi)=\begin{pmatrix}\cos\frac{1}{2}\theta\\ e^{i\phi}\sin\frac{1}{2}\theta\end{pmatrix}$$

corresponds to the weights

$$w(\sigma_y=\pm1)=\frac{1\pm\sin\theta\sin\phi}{2}\qquad w(\sigma_z=\pm1)=\frac{1\pm\cos\theta}{2}. \tag{15}$$

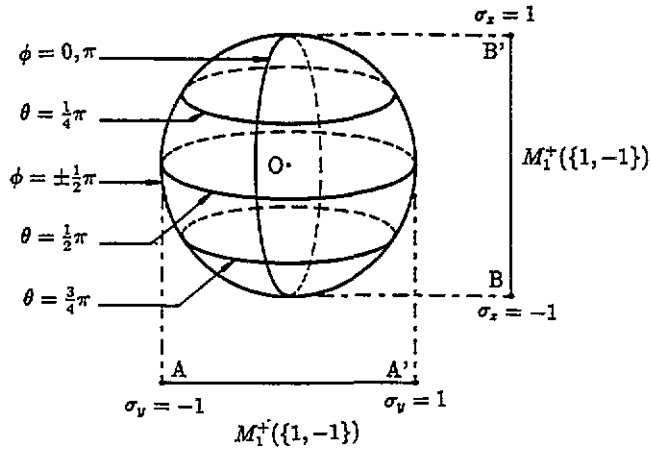


Figure 1. The sphere represents the set of polarization states of a spin-$\frac{1}{2}$ system. The mapping of the sphere associated with σ_y (resp., σ_z) can be viewed as the orthogonal projection on the $\phi=\pm\frac{1}{2}\pi$ plane followed by the orthogonal projection on the segment AA' (resp., BB').

The range of any joint observable of the C-representatives of σ_y and σ_z would then be the set of the probability measures on a four element set, namely the tetrahedron whose vertices correspond to the ordered pairs (1,1), (−1, 1), (1, −1), (−1, −1) of values of σ_y,σ_z (or of their C-representatives). The weights in (15) provide, for pure states, the two marginal distributions of any joint distribution, which then become forced to take the form

$$w(\sigma_y=1,\sigma_z=1)=v(\theta,\phi)$$

$$w(\sigma_y=-1,\sigma_z=1)=\frac{1+\cos\theta}{2}-v(\theta,\phi)$$

$$w(\sigma_y=1,\sigma_z=-1)=\frac{1+\sin\theta\sin\phi}{2}-v(\theta,\phi)$$

$$w(\sigma_y=-1,\sigma_z=-1)=v(\theta,\phi)-\frac{\sin\theta\sin\phi}{2}-\frac{\cos\theta}{2}$$

where $v(\theta,\phi)$ is any function that makes non-negative the four weights above. Different choices of $v(\theta,\phi)$ correspond to different joint distributions which, however, have the same marginal distributions. Each choice of $v(\theta,\phi)$ determines a correspondence between the

surface of the sphere of figure 1, that represents the set of pure states, and a surface inside the tetrahedron that we denote $s(\theta, \phi)$: in figure 2 we give an example for a particular choice of $v(\theta, \phi)$. One can check that for all possible choices of $v(\theta, \phi)$ the surface $s(\theta, \phi)$ does not lie in a plane. We can now visualize what a joint observable of the C-representatives of σ_y and σ_z should be: it is an affine mapping from the set of all the probability measures on the surface of the sphere of figure 1 onto the convex hull of $s(\theta, \phi)$. Notice that the probability measures concentrated, with equal weights, at diametrically opposed points on the sphere are now mapped by this affine mapping into the points of a segment inside the convex hull of $s(\theta, \phi)$, as shown in figure 2. This is why it is impossible to have a joint observable of σ_y and σ_z in standard quantum mechanics, where the above probability measures are all represented by the same density operator, i.e. by the centre of the sphere. An affine mapping of the sphere onto the convex hull of $s(\theta, \phi)$ is thus impossible.

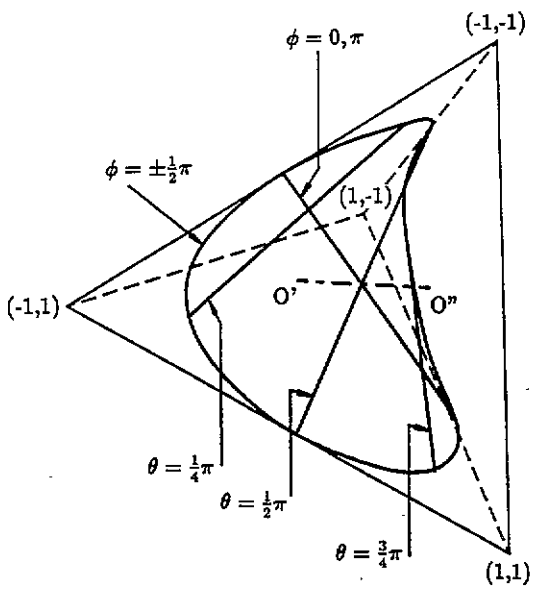


Figure 2. The curved surface inside the tetrahedron is the image of the surface of the sphere of figure 1 under the joint observable of the C-representatives of σ_y and σ_z, for the particular choice $v(\theta, \phi) = \frac{1}{4}(1 + \cos\theta)(1 + \sin\theta \sin\phi)$. The segment O'O" is the image of the mixtures, with equal weights, of opposed points on the surface of the sphere.

Let us now leave the spin-$\frac{1}{2}$ example and return to more general aspects.

The fact that quantum theory admits the canonical classical extension has, we believe, an interest on the epistemological side. This classical extension adds something to the debate on possible connections and compatibilities between quantum theory and classical frameworks. Notice that, in view of the remark at the end of section 4, no conflict can arise with the known 'no-go' theorems for hidden variable models of quantum mechanics which prevent the embedding of the projection lattice of a Hilbert space into Boolean algebras [20]. Our classical extension of quantum theory calls into play a richer set of states—not just S_Q but all the probability measures on ∂S_Q—and a richer set of observables, only some of them being C-representatives of Q-observables. In this sense we might pictorially say that the canonical classical extension is somewhat like a *hidden observable* generalization of quantum theory.

From the physical point of view, it is precisely this richer set of observables that makes an interesting challenge. The familiar Q-observables do not separate different convex combinations of pure states that correspond to the same density operator; in other words there are distinct preparation procedures of statistical ensembles that are not distinguished by the Q-observables. On the contrary, the new observables entering the canonical classical extension do that separation.

One might wonder what kind of hypothetical phenomenon could effect the above separation of convex combinations of pure states. It was observed in [22] that multiparticle correlations might do this. Let us here outline that nonlinear phenomena appear to be natural candidates for that job, as pointed out years ago by Mielnik [25] and further discussed in [26]. (For recent surveys on possible nonlinear generalizations of quantum mechanics see, e.g., [27, 28].)

Let us again take the spin-$\frac{1}{2}$ example to give a hint in that direction. Suppose that a particle in the (pure) state $\psi(\theta, \phi)$ interacts with some device and let $f(\theta, \phi)$ be the response function. The fact, characteristic of standard quantum mechanics, that the unpolarized state can be thought of as a mixture with equal weights of the spin-up states along the opposed directions (θ, ϕ) and $(\pi - \theta, \pi + \phi)$, together with the complete arbitrariness of the choice (θ, ϕ), can be translated by saying that the response function must meet the constraint [20]

$$f(\theta, \phi) + f(\pi - \theta, \pi + \phi) = \text{constant}. \tag{16}$$

Now, if $f(\theta, \phi)$ has the common form

$$f(\theta, \phi) = (\psi, A\psi) \tag{17}$$

for some self-adjoint matrix with elements a_{ij}, $i, j = 1, 2$, we immediately get

$$f(\theta, \phi) = \frac{a_{11} + a_{22}}{2} + \frac{a_{11} - a_{22}}{2} \cos\theta + (\operatorname{Re} a_{12} \cos\phi - \operatorname{Im} a_{12} \sin\phi) \sin\theta$$

and the constraint (16) is obviously met. But if A in (17) is replaced by some nonlinear (i.e. depending on $\psi(\theta, \phi)$) operator then we have clearly to expect that the response function will escape the constraint (16), which precisely corresponds to the separation of different mixtures of pure states associated to the same density operator.

As a parallel example, consider the polarization states of a photon whose quantum description corresponds to the Poincaré sphere. The non-unique decomposability of quantum mixtures is again self-evident from the shape of that convex set. Let us focus attention on the linear polarization states, i.e. on the equatorial section of the Poincaré sphere: the mixture with equal weights of two pure states polarized along perpendicular directions, say along θ and $\frac{1}{2}\pi + \theta$, leaves no memory of θ in any measurement associated with the ordinary (quantum) observables. Hence the response function $f(\theta)$ of any device interacting with a photon linearly polarized along θ must meet the constraint [20]

$$f(\theta) + f(\tfrac{1}{2}\pi + \theta) = \text{constant}. \tag{18}$$

A hypothetical (nonlinear) device transforming, as an example, the polarization along θ into the one along 2θ would then give rise, if followed by a linear polarimeter, to a response function $f(\theta) = \cos^2(2\theta)$ that escapes the constraint (18). In other words, such a hypothetical device would separate different mixing procedures of pure states that are collapsed into the same density operator by the quantum description.

Besides the hints advanced by the above hypothetical examples, it is worth noting that the observables described by nonlinear operators find an appropriate representation as observables related to the corresponding canonical classical extension [26].

Acknowledgments

One of the authors (S Bugajski) acknowledges the support provided by the Science Research Committee (KBN, Warsaw) under grant no. 2/2420/92/03.

References

[1] Bugajski S 1993 *Int. J. Theor. Phys.* **32** 969–77
[2] Misra B 1974 *Physical Reality and Mathematical Description* ed C P Enz and J Mehra (Dordrech: Reidel) pp 455–76
[3] Ali S T and Prugovecki E 1977 *J. Math. Phys.* **18** 219–28
[4] Singer M and Stulpe W 1992 *J. Math. Phys.* **33** 131–42
[5] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[6] Beltrametti E G and Bugajski S 1993 *Int. J. Theor. Phys.* **32** 2235–44
[7] Cassinelli G and Lahti P J 1993 *J. Math. Phys.* **34** 5468–75
[8] Reed M and Simon B 1972 *Methods of Modern Mathematical Physics* vol 1 (New York: Academic)
[9] Davies E B 1976 *Quantum Theory of Open Systems* (London: Academic)
[10] Ludwig G 1970 *Deutung des Begriffs „Physikalische Theories" und Axiomatische Grundlagung der Hilbertraumstruktur der Quantenmechanik* Lecture Notes in Physics 4 (Berlin: Springer)
[11] Busch P, Grabowski M and Lahti P J 1995 *Operational Quantum Physics* Lecture Notes in Physics 2 (Berlin: Springer)
[12] Foulis D J and Schroeck F E 1990 *Found. Phys.* **20** 823–58
[13] Bugajski S 1993 *Int. J. Theor. Phys.* **32** 389–98
[14] Semandeni Z 1971 *Banach Spaces of Continuous Functions vol 1 (Monografie Matematyczne* **55***)* (Warsaw: PWN)
[15] Billingsley P 1979 *Probability and Measure* (New York: Wiley)
[16] Busch P and Quadt R 1993 *Int. J. Theor. Phys.* **32** 2261–9
[17] Alfsen E M 1971 *Compact Convex Sets and Boundary Integrals* (Berlin: Springer)
[18] Bugajski S 1994 *Rep. Math. Phys.* **34** 235–40
[19] Rudin W 1973 *Functional Analysis* (New York: McGraw-Hill)
[20] Beltrametti E G and Cassinelli G 1981 *The Logic of Quantum Mechanics* (Reading, MA: Addison-Wesley)
[21] Ghirardi G C, Rimini A and Weber T 1976 *Nuovo Cimento* **36** 97–118
[22] Hudson R L 1981 *Found. Phys.* **11** 805–8
[23] Hudson R L and Moody G R 1976 *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **33** 343
[24] Amann A 1993 The quantum mechanical measurement process in the thermodynamical formalism *Symp. Foundations of Modern Physics (1993)* ed P Busch, P Lahti and P Mittelstaedt (Singapore: World Scientific)
[25] Mielnik B 1974 *Commun. Math. Phys.* **37** 221–56; 1985 *Commun. Math. Phys.* **101** 323–39
[26] Bugajski S 1992 *Int. J. Theor. Phys.* **30** 961–71
[27] Weinberg S 1989 *Ann. Phys., NY* **194** 336–86
[28] Jones K R W 1994 *Ann. Phys., NY* **233** 295–316