International Journal of Theoretical Physics, Vol. 27, No. 5, 1988

Conditional Expectations, Conditional Distributions, and *A Posteriori* Ensembles in Generalized Probability Theory

Werner Stulpe[1]

Received January 12, 1988

A general probabilistic framework containing the essential mathematical structure of any statistical physical theory is reviewed and enlarged to enable the generalization of some concepts of classical probability theory. In particular, generalized conditional probabilities of effects and conditional distributions of observables are introduced and their interpretation is discussed in terms of successive measurements. The existence of generalized conditional distributions is proved, and the relation to M. Ozawa's *a posteriori* states is investigated. Examples concerning classical as well as quantum probability are discussed.

1. INTRODUCTION

Many attempts have been made to transfer the concept of conditional expectation in classical probability theory to the probabilistic scheme of quantum mechanics. Since the statistical interpretation of quantum theory is based on measurements and the classical conditional expectations can be interpreted in terms of two successive measurements with random outcomes, the notion of a quantum conditional expectation should not only be formally analogous to the classical one, it should also admit an interpretation in terms of successive measurements. In this paper, such a concept of conditional expectations is formulated in the context of a general framework for statistical theories containing classical and quantum probability as special cases. On that basis, conditional distributions of observables are introduced, and their existence is proved.

[1]Institut für Theoretische Physik, Technische Universität Berlin, 1000 Berlin 12, Germany. Present address: Department of Mathematics and Computer Science, University of Denver, Denver, Colorado 80208.

0020-7748/88/0500-0587$06.00/0

Conditional expectations in quantum mechanics were mostly defined and discussed in the context of von Neumann algebras (e.g., Umegaki, 1954, 1964; Nakamura and Umegaki, 1962; Gudder and Marchand, 1977). Davies and Lewis (1970) supposed the duals of instruments being quantum conditional expectations; however, Cycon and Hellwig (1977; Hellwig, 1981) pointed out that this concept is not completely analogous to the classical one. Our formulation of generalized conditional expectations was inspired by the latter two papers, and our development of generalized conditional distributions represents a further step toward a generalization of classical probability theory, which, in particular, may be useful in the investigation of stochastic processes in any statistical physical theory (Hellwig and Stulpe, 1983; Stulpe, 1986, 1987).

To make precise our conception of conditioning, let us briefly describe the relation between conditional expectations and successive measurements in classical probability. Given a probability space (Ω, Σ, μ), a measurable space (M, Ξ), the real numbers $\mathbf{R}$ with the σ-algebra $\Xi(\mathbf{R})$ of its Borel sets, and two random variables $X: \Omega \to M$ and $Y: \Omega \to \mathbf{R}$, then the *conditional expectation* $E^X(Y)$: $M \to \mathbf{R}$ *of* Y *given* X is defined by

$$\int_{X^{-1}(B)} Y\,d\mu = \int_B E^X(Y)\,d(\mu \circ X^{-1})$$

where Y is assumed to be μ-integrable and $B \in \Xi$ is arbitrary. The Ξ-measurable function $E^X(Y)$ is uniquely determined up to $\mu \circ X^{-1}$-equivalence. If there is a point $x \in M$ such that $\{x\} \in \Xi$ and $\mu(X^{-1}(\{x\})) \neq 0$, then

$$E^X(Y)(x) = \frac{1}{\mu(X^{-1}(\{x\}))} \int_{X^{-1}(\{x\})} Y\,d\mu = \int Y\,d\left(\frac{\mu(X^{-1}(\{x\}) \cap \cdot)}{\mu(X^{-1}(\{x\}))}\right)$$

holds true. Thus, $E^X(Y)(x)$ is the conditional expectation of Y given that $X(\omega) = x$, resp. the conditional expectation of Y given $X^{-1}(\{x\})$, which is just the expectation value of the observable Y in the subensemble of the ensemble μ obtained by selection of systems according to the outcome x of the observable X.

This paper has essentially three aims. First, we formulate the general probabilistic framework on which our developments are based. According to Ludwig (1983, 1985), the essential mathematical structure of any statistical physical theory consists of a dual pair of a base-norm Banach space and an order-unit norm Banach space which is interpreted in terms of *ensembles* and *effects*. Following Werner (1983), we call such a dual pair a *statistical duality*. For technical reasons, we distinguish some statistical dualities we call *generalized probability spaces*. The order-unit norm space of a generalized probability space is closed under certain least upper bounds such that the ensembles are represented by normal positive functionals acting on the linear hull of the effects. Base-norm spaces in the context of statistical theories were also considered by Davies and Lewis (1970; Davies 1976), Edwards (1970, 1971), and Gudder (1979a,b). Moreover, Davies and Lewis introduced the concept of instruments to describe preparative measurements of observables. In Section 2 we establish statistical dualities and generalized probability spaces, and in Section 3 we recall the concepts of *operations*, *observables*, and *instruments* and put them into our setting.

The second aim is to introduce *conditional expectations* in *generalized probability theory* by means of instruments. Proceeding on the lines suggested above, we accomplish this in Section 4. It turns out that the generalized conditional expectations and distributions presented here are closely related to the corresponding notions in classical probability theory.

Our third aim concerns the introduction and investigation of *generalized conditional distributions*, which are treated in Section 5. In particular, we deal with their existence and uniqueness. In Section 6, finally, the relation of conditional distributions to the *a posteriori states* introduced in quantum mechanics by Ozawa (1985a,b) is investigated, and some examples are considered.

2. GENERALIZED PROBABILITY THEORY

As basic elements of a statistical physical theory we will comprehend the *statistical ensembles* of physical systems and the classes of statistically equivalent realistic measurements with the outcomes 0 and 1, which are called *effects*. The set K of ensembles and the set L of effects are connected by a probability functional

$$\mu:\quad K\times L\to[0,1]$$

$$(v,l)\to\mu(v,l)$$

which assigns to every $v\in K$ and every $l\in L$ the *probability for the outcome* 1 *of the effect* l *in the ensemble* v. In particular, we assume that there are two effects $0, e\in L$ yielding the probability 0, resp. 1, for outcome 1 in any ensemble. Corresponding to the physical fact that one can produce mixtures of ensembles as well as of effects, K and L should be, in a certain sense, convex sets, and the probability functional μ should be affine in both arguments.

As it was shown by Ludwig (1983, 1985), K and L can be embedded in a dual pair $\langle \mathbf{V}, \mathbf{W}\rangle$ of real vector spaces such that K and L linearly span $\mathbf{V}$ and $\mathbf{W}$, respectively, and μ coincides with the restriction of the bilinear functional $\langle\cdot,\cdot\rangle$ to $K\times L$. The pair $\langle \mathbf{V}, \mathbf{W}\rangle$ is uniquely determined up to isomorphism by K, L, and μ. Moreover, some slight mathematical idealizations are sufficient to equip $\mathbf{V}$ and $\mathbf{W}$ with further structures. It turns out that $\langle \mathbf{V}, \mathbf{W} \rangle$ is a dual pair of ordered Banach spaces where $(\mathbf{V}, K)$ is a base norm space, $(\mathbf{W}, e)$ an order-unit norm space, and $L = [0, e]$. The precise structure of this construction is given by the following definition (compare also Werner, 1983).

Definition 2.1. Let a complete base-norm space $(\mathbf{V}, K)$ with closed base $K \subset \mathbf{V}$, a complete order-unit norm space $(\mathbf{W}, e)$ with closed order-unit interval $L := [0, e] \subset \mathbf{W}$, and a nondegenerate bilinear functional $\langle \cdot, \cdot \rangle : \mathbf{V} \times \mathbf{W} \to \mathbf{R}$ be given. The pair $\langle \mathbf{V}, \mathbf{W} \rangle$ is called a *statistical duality* if $\langle \cdot, \cdot \rangle$ is compatible with the norms and orderings in $\mathbf{V}$ and $\mathbf{W}$, i.e., if for all $v \in \mathbf{V}$ and $l \in \mathrm{W}$ the following conditions hold:

(i) $\|v\| = \sup_{l \in [-e, e]} |\langle v, l \rangle|$
(ii) $\|l\| = \sup_{v \in K} |\langle v, l \rangle|$
(iii) $v \geq 0$ iff $\langle v, l \rangle \geq 0$ for all $l \geq 0$
(iv) $l \geq 0$ iff $\langle v, l \rangle \geq 0$ for all $v \geq 0$

Remark 2.1.

(a) Condition (i) means that the linear functionals $iv := \langle v, \cdot \rangle$ on $\mathbf{W}$ are bounded with $\|iv\| = \|v\|$, and (iii) means that iv is positive if and only if $v \geq 0$. The analogous statements are valid for the functionals $jl := \langle \cdot, l \rangle$.

(b) For $v \geq 0$ it follows that $\|v\| = \langle v, e \rangle$. Namely, $v \geq 0$ and $-e \leq l \leq e$ imply $-\langle v, e \rangle \leq \langle v, l \rangle \leq \langle v, e \rangle$, resp. $|\langle v, l \rangle| \leq \langle v, e \rangle$. Hence, $\|v\| = \sup_{l \in [-e, e]} |\langle v, l \rangle| = \langle v, e \rangle$ In particular, if $v \in K$, then $\|v\| = \langle v, e \rangle = 1$.

(c) The dual space $\mathbf{V}'$ together with the canonical positive linear functional $\tilde{e}$ defined by $\tilde{e}(v) := 1$ for all $v \in K$ is a complete order-unit norm space with closed order-unit interval $\tilde{L} := [0, \tilde{e}] \subset \mathbf{V}'$, and by $\tilde{K} := \{\Lambda \in \mathbf{W}' | \Lambda \geq 0 \text{ and } \Lambda(e) = 1\}$, $\mathbf{W}'$ is a complete base-norm space with closed base $\tilde{K} \subset \mathbf{W}'$ (see, e.g., Nagel, 1974).

(d) The canonical embeddings $i: \mathbf{V} \to \mathbf{W}'$ and $j: \mathbf{W} \to \mathbf{V}'$ are linear, isometric, and positive with positive inverses $i^{-1}: i\mathbf{V} \to \mathbf{V}$ and $j^{-1}: j\mathbf{W} \to \mathbf{W}$. Moreover, we have $iK \subseteq \tilde{K}$, $je = \tilde{e}$, and $jL \subseteq \tilde{L}$.

(e) $\langle \mathbf{V}, \mathbf{V}' \rangle$ and $\langle \mathbf{W}', \mathbf{W} \rangle$ are statistical dualities by the canonical bilinear forms of the constituting spaces.

The physical interpretation of statistical dualities is, by construction, as follows. Any sort of physical system is associated with suitable $(\mathbf{V}, K)$, $(\mathbf{W}, e)$, and $\langle \mathbf{V}, \mathbf{W} \rangle$. The elements of the convex sets K and L are mathematical pictures of the ensembles and effects, respectively. The probability for the outcome 1 of an effect $l \in L$ in an ensemble $v \in K$ is given by

$$0 = \langle v, 0 \rangle \leq \langle v, l \rangle \leq \langle v, e \rangle = 1$$

We mention that the partial order in **W**, resp. in L, can be interpreted physically. The condition $l_1 \leq l_2$ for two effects $l_1, l_2 \in L$ is equivalent to $\langle v, l_1\rangle \leq \langle v, l_2\rangle$ for all $v \in K$, that is, a measuring apparatus representing l_2 is more sensitive than an apparatus for l_1. We call the extreme points of K, if any, *pure ensembles* and the extreme points of L *decision effects.*

The weak topology $\sigma(\mathbf{V}, \mathbf{W})$ is the coarsest topology in **V** such that all linear functionals jl, $l \in \mathbf{W}$, are continuous. A neighborhood base of $v \in \mathbf{V}$ is given by the sets

$$U(v; l_1, \ldots, l_n; \varepsilon) := \{\tilde{v} \in \mathbf{V} \,|\, |\langle \tilde{v}, l_i\rangle - \langle v, l_i\rangle| < \varepsilon \text{ for } l_i \in \mathbf{W}, i = 1, 2, \ldots, n\}$$

An ensemble $v \in K$ is physically approximated by $\tilde{v} \in K$ if for many (but finitely many!) effects $l_1, l_2, \ldots, l_n \in L$ the probabilities $\langle \tilde{v}, l_i\rangle$ differ from $\langle v, l_i\rangle$ by an amount less than a small $\varepsilon > 0$. This statement can be tested experimentally and can mathematically be characterized by $\tilde{v} \in U(v; l_1, \ldots, l_n; \varepsilon)$. Hence, $\sigma(\mathbf{V}, \mathbf{W})$, resp. $\sigma(\mathbf{V}, \mathbf{W}) \cap K = \sigma(\mathbf{V}, L) \cap K$, is the "topology of physical approximation of ensembles" (cf. Ludwig, 1983, 1985; Werner, 1983; Haag and Kastler, 1964). Analogously, the weak topologies $\sigma(\mathbf{W}, \mathbf{V})$ and $\sigma(\mathbf{W}, \mathbf{V}) \cap L = \sigma(\mathbf{W}, K) \cap L$, respectively, describe the physical approximation of effects.

The topologies $\sigma(\mathbf{V}, \mathbf{W})$ and $\sigma(\mathbf{W}, \mathbf{V})$ correspond to $\sigma(\mathbf{W}', \mathbf{W}) \cap i\mathbf{V}$ and $\sigma(\mathbf{V}', \mathbf{V}) \cap j\mathbf{W}$. The K, L, $\tilde{K}$, $\tilde{L}$, and the corresponding positive cones are closed in the respective topologies $\sigma(\mathbf{V}, \mathbf{W})$, $\sigma(\mathbf{W}, \mathbf{V})$, $\sigma(\mathbf{W}', \mathbf{W})$, and $\sigma(\mathbf{V}', \mathbf{V})$. Hence, $\tilde{K}$ and $\tilde{L}$ are even compact by the Banach–Alaoglu theorem. The statement **V** separates the points of **W** is equivalent to $i\mathbf{V} \subseteq \mathbf{W}'$ being $\sigma(\mathbf{W}', \mathbf{W})$-dense in $\mathbf{W}'$; the analog holds for $j\mathbf{W} \subseteq \mathbf{V}'$. In particular, iK is a $\sigma(\mathbf{W}', \mathbf{W})$-dense subset of $\tilde{K}$ and jL a $\sigma(\mathbf{V}', \mathbf{V})$-dense subset of $\tilde{L}$.

If $\{v_n\}_{n\in\mathbf{N}}$ is an increasing norm-bounded sequence in **V** (i.e., $v_n \leq v_{n+1}$, $\|v_n\| \leq c$, $c \geq 0$), then the least upper bound $\sup_{n\in\mathbf{N}} v_n$ exists and is equal to $\|\cdot\|$-$\lim_{n\to\infty} v_n$ (Ludwig, 1983, 1985). An analogous result for **W** does not hold in general. However, for increasing norm-bounded sequences $\{\lambda_n\}_{n\in\mathbf{N}}$ in $\mathbf{V}'$ (i.e., $-c\tilde{e} \leq \lambda_n \leq \lambda_{n+1} \leq c\tilde{e}$), $\sup_{n\in\mathbf{N}} \lambda_n = \sigma(\mathbf{V}', \mathbf{V})$-$\lim_{n\to\infty} \lambda_n$ holds, which can be generalized to increasing norm-bounded nets $\{\lambda_\alpha\}_{\alpha\in A}$ in $\mathbf{V}'$. This assertion follows from the $\sigma(\mathbf{V}', \mathbf{V})$-compactness of the closed unit ball in $\mathbf{V}'$, but it can also be proved elementarily. Let $\{\lambda_n\}_{n\in\mathbf{N}}$ be an increasing norm-bounded sequence in $\mathbf{V}'$; then for all $v \geq 0$, $\lim_{n\to\infty} \lambda_n(v) = \sup_{n\in\mathbf{N}} \lambda_n(v)$ exists. Because the positive cone in **V** is generating, the limit even exists for all $v \in \mathbf{V}$. Moreover, an element $\lambda \in \mathbf{V}'$ is defined by $\lambda(v) := \lim_{n\to\infty} \lambda_n(v)$. It is obvious that

$$\lambda = \sigma(\mathbf{V}', \mathbf{V})\text{-}\lim_{n\to\infty} \lambda_n = \sup_{n\in\mathbf{N}} \lambda_n$$

holds, where the latter least upper bound is understood with respect to the partial order in $\mathbf{V}'$.

For technical reasons it is useful now to distinguish those statistical dualities $\langle \mathbf{V}, \mathbf{W} \rangle$ whose order-unit norm spaces $\mathbf{W}$ are closed under the least upper bounds of increasing sequences (Stulpe, 1986).

Definition 2.2. We call a statistical duality $\langle \mathbf{V}, \mathbf{W} \rangle$ a *generalized probability space* if all increasing norm-bounded sequences $\{l_n\}_{n\in\mathbf{N}}$ in $\mathbf{W}$ converge in the $\sigma(\mathbf{W}, \mathbf{V})$-topology.

Remark 2.2.

(a) For a statistical duality $\langle \mathbf{V}, \mathbf{W} \rangle$ to be a generalized probability space, it is sufficient that $j\mathbf{W} \subseteq \mathbf{V}'$ be $\sigma(\mathbf{V}', \mathbf{V})$-closed. But this is equivalent to $j\mathbf{W} = \mathbf{V}'$, since $j\mathbf{W}$ is $\sigma(\mathbf{V}', \mathbf{V})$-dense in $\mathbf{V}'$. Examples show the existence of generalized probability spaces with proper subspaces $j\mathbf{W} \subset \mathbf{V}'$.

(b) The condition of the definition means that all increasing norm-bounded sequences $\{l_n\}_{n\in\mathbf{N}}$ in $\mathbf{W}$ have a least upper bound in $\mathbf{W}$ satisfying

$$\left\langle v, \sup_{n\in\mathbf{N}} l_n \right\rangle = \sup_{n\in\mathbf{N}} \langle v, l_n \rangle$$

for $v \geq 0$. This least upper bound is given by

$$\sup_{n\in\mathbf{N}} l_n = \sigma(\mathbf{W}, \mathbf{V})\text{-}\lim_{n\to\infty} l_n$$

Furthermore, by

$$j\left(\sigma(\mathbf{W}, \mathbf{V})\text{-}\lim_{n\to\infty} l_n\right) = \sigma(\mathbf{V}', \mathbf{V})\text{-}\lim_{n\to\infty} jl_n = \sup_{n\in\mathbf{N}} jl_n$$

we obtain

$$j\left(\sup_{n\in\mathbf{N}} l_n\right) = \sup_{n\in\mathbf{N}} jl_n$$

i.e., $\sup_{n\in\mathbf{N}} l_n$ corresponds to the least upper bound of $\{jl_n\}_{n\in\mathbf{N}}$ taken in $\mathbf{V}'$ (and not only in $j\mathbf{W}$!).

(c) Positive linear functionals on a complete base-norm space with closed cone are automatically bounded, and analogously positive linear functionals Λ on an arbitrary order-unit norm space $(\mathbf{W}, e)$, the latter ones fulfilling $\|\Lambda\| = \Lambda(e)$ (see, e.g., Stulpe, 1986). Now let $(\mathbf{W}, e)$ be a complete order-unit norm space with closed cone such that $\sup_{n\in\mathbf{N}} l_n$ exists for increasing norm-bounded sequences. A positive linear functional $\Lambda: \mathbf{W} \to \mathbf{R}$ is called *normal* if

$$\Lambda\left(\sup_{n\in\mathbf{N}} l_n\right) = \sup_{n\in\mathbf{N}} \Lambda(l_n)$$

holds. This definition and the preceding statements can be transferred to positive linear maps between order-unit and base-norm spaces.

(d) If $\langle \mathbf{V}, \mathbf{W} \rangle$ is a generalized probability space, then the embedding map j as well as the functionals iv with $v \in \mathbf{V}$ and $v \geq 0$ are normal.

That $\langle \mathbf{V}, \mathbf{W} \rangle$ is a generalized probability space means physically that for any sequence of effects increasing in sensitivity there exists one effect that is more sensitive and is physically approximated by the given sequence.

We next consider some familiar examples of statistical dualities and generalized probability spaces.

Example 2.3.

(a) For any base-norm Banach space $\mathbf{V}$ and any order-unit norm Banach space $\mathbf{W}$, $\langle \mathbf{V}, \mathbf{V}' \rangle$ and $\langle \mathbf{W}', \mathbf{W} \rangle$ are statistical dualities by their canonical bilinear functionals. Moreover, $\langle \mathbf{V}, \mathbf{V}' \rangle$ is a generalized probability space, whereas in general for $\langle \mathbf{W}', \mathbf{W} \rangle$ this does not happen.

(b) A σ-finite measure space (Ω, Σ, μ) yields an example of type $\langle \mathbf{V}, \mathbf{V}' \rangle$, namely, $\mathbf{V} := \mathbf{L}^1_{\mathbf{R}}(\Omega, \Sigma, \mu)$, $\mathbf{W} := \mathbf{L}^\infty_{\mathbf{R}}(\Omega, \Sigma, \mu) \cong \mathbf{V}'$, $\langle \rho, f \rangle := \int \rho f \, d\mu$.

(c) Another special case of $\langle \mathbf{V}, \mathbf{V}' \rangle$ arises from a complex separable Hilbert space $\mathbf{H}$. The dual pairing of the space $\mathbf{V} := \mathbf{C}^1_s(\mathbf{H})$ of self-adjoint trace-class operators and the space $\mathbf{W} := \mathbf{B}_s(\mathbf{H})$ of all bounded self-adjoint operators in $\mathbf{H}$ by $\langle V, A \rangle := \operatorname{tr} VA$, $\mathbf{W} \cong \mathbf{V}'$, describes usual quantum mechanics.

(d) The last example can be generalized to $\mathbf{V} := \mathbf{M}_{*s}$ and $\mathbf{W} := \mathbf{M}_s \cong \mathbf{V}'$ where $\mathbf{M}_s$ consists of the self-adjoint elements of a von Neumann algebra $\mathbf{M}$ on an arbitrary complex Hilbert space and $\mathbf{M}_*$ denotes the predual of $\mathbf{M}$.

(e) Let M be a compact Hausdorff space, $\Xi_0(M)$ the σ-algebra of its Baire sets, and $\mathbf{M}_{\mathbf{R}}(M, \Xi_0(M))$ the space of bounded, signed Baire measures. An example of type $\langle \mathbf{W}', \mathbf{W} \rangle$ is given by $\mathbf{V} := \mathbf{M}_{\mathbf{R}}(M, \Xi_0(M))$, $\mathbf{W} := \mathbf{C}_{\mathbf{R}}(M)$, and $\langle \nu, f \rangle := \int f \, d\nu$, since $\mathbf{V} \cong \mathbf{W}'$.

(f) We obtain a more general example for $\langle \mathbf{W}', \mathbf{W} \rangle$ by the Segal algebra $\mathbf{A}_s$ of the self-adjoint elements of an arbitrary unital C^*-algebra $\mathbf{A}$, namely, $\mathbf{V} := \mathbf{A}^*_s$, $\mathbf{W} := \mathbf{A}_s$, and $\langle \omega, a \rangle := \omega(a)$, where $\mathbf{V} \cong \mathbf{W}'$.

(g) If $\mathbf{V} := \mathbf{M}_{\mathbf{R}}(\Omega, \Sigma)$ is the space of bounded, signed measures on an arbitrary measurable space (Ω, Σ) and $\mathbf{W} := \mathbf{F}_{\mathbf{R}}(\Omega, \Sigma)$ the space of all real-valued, bounded measurable functions, then a statistical duality $\langle \mathbf{V}, \mathbf{W} \rangle$ is defined by $\langle \nu, f \rangle := \int f \, d\nu$. The duality $\langle \mathbf{V}, \mathbf{W} \rangle$ is a generalized probability space where in general $j\mathbf{W}$ is a proper subspace of $\mathbf{V}'$ (and $i\mathbf{V}$ a proper subspace of $\mathbf{W}'$).

3. OPERATIONAL CONCEPTS

In this section, we will introduce operations, observables, and instruments in our formulation (cf. Davies and Lewis, 1970; Edwards, 1970, 1971; Davies, 1976; Ludwig, 1983, 1985; Kraus, 1983).

A measurement is *preparative* if the systems interacting with the apparatus are not absorbed. The classes of statistically equivalent preparative 0–1 measurements are called *operations.* Formally, operations are defined as follows.

Definition 3.1. Let a statistical duality $\langle \mathbf{V}, \mathbf{W}\rangle$ be given. A positive linear map $T: \mathbf{V} \to \mathbf{V}$ is called an *operation* if:

(i) $Tv \in \bigcup_{0\le\alpha\le 1} \alpha K$ holds for all $v \in K$
(ii) The adjoint map T' with respect to $\langle \mathbf{V}, \mathbf{W}\rangle$ exists.

Remark 3.1.

(a) A positive linear map T satisfies condition (i) if and only if $\|Tv\| \le 1$ for $v \in K$, resp. if $\|T\| \le 1$. For any linear map T, (ii) is equivalent to $T^*(j_0\mathbf{W}) \subseteq j_0\mathbf{W}$, where j_0 denotes the canonical embedding of $\mathbf{W}$ into the algebraic dual space $\mathbf{V}^*$, and T^* the adjoint of T with respect to the duality $\langle \mathbf{V}, \mathbf{V}^*\rangle$. Furthermore, T' exists if and only if T is $\sigma(\mathbf{V}, \mathbf{W})$-continuous.

(b) T' has the properties $T' \ge 0$, $\|T\| = \|T'\| = \|T'e\|$, $T'L \subseteq L$, and $T'' = T$. Moreover, a $\sigma(\mathbf{V}, \mathbf{W})$-continuous linear map T is an operation if and only if $T'L \subseteq L$.

(c) If $\langle \mathbf{V}, \mathbf{W}\rangle$ is a generalized probability space and T an operation, then T' is normal. This follows from its $\sigma(\mathbf{W}, \mathbf{V})$-continuity or from

$$\left\langle v, T'\left(\sup_{n\in\mathbb{N}} l_n\right)\right\rangle = \left\langle Tv, \sup_{n\in\mathbb{N}} l_n\right\rangle = \sup_{n\in\mathbb{N}} \langle Tv, l_n\rangle = \left\langle v, \sup_{n\in\mathbb{N}} T'l_n\right\rangle$$

where $v \ge 0$.

Given an ensemble $v \in K$, then we interpret as follows:

(i) $\|Tv\| = \langle Tv, e\rangle = \langle v, T'e\rangle$ is the probability for the outcome 1 of the measurement described by T.
(ii) If $Tv \ne 0$, $Tv/\|Tv\|$ is the ensemble obtained by selection of systems according to the outcome 1.

As a consequence of (i), $l := T'e$ is the effect "measured by T."

Let us consider the *successive* measurement of $l = T'e$ and another effect $\tilde{l} \in L$ in an ensemble v. This means that $\tilde{l}$ is measured in the ensemble obtained by nonselective measurement of l. The joint probability for occurrence of the outcomes 1 of l and $\tilde{l}$ is given by

$$\langle v, l\rangle \left\langle \frac{Tv}{\|Tv\|}, \tilde{l}\right\rangle = \langle Tv, \tilde{l}\rangle = \langle v, T'\tilde{l}\rangle$$

i.e., $T'\tilde{l} \in L$ is the effect for occurrence of the outcomes 1 of l and $\tilde{l}$ by successive measurement.

The interpretation of operations in terms of preparative measurements is not the only one, in particular for two special classes of operations. Using a terminology similar to that of Ludwig (1983, 1985), we call an operation $T: \mathbf{V} \to \mathbf{V}$ with $TK \subseteq K$ a *mixture endomorphism.* For such a map $\|T\| = 1$ holds. The property $TK \subseteq K$ of a $\sigma(\mathbf{V}, \mathbf{W})$-continuous linear map T is equivalent to $T' \geq 0$ and $T'e = e$. A bijective mixture endomorphism is called a *mixture automorphism* if $TK = K$ holds and $(T^{-1})'$ exists. A $\sigma(\mathbf{V}, \mathbf{W})$-continuous linear map is a mixture automorphism if and only if T' is bijective, $T'L = L$ is fulfilled, and $((T')^{-1})'$ exists. Moreover, T and T' are automorphisms with respect to norm and order in $\mathbf{V}$ and $\mathbf{W}$.

Mixture endomorphisms are used to describe the dynamics of physical systems, and mixture automorphisms and their adjoints also have an important meaning as symmetry transformations.

Definition 3.2. Let $\langle \mathbf{V}, \mathbf{W} \rangle$ be a statistical duality and $I := \{(s, t) \in (\mathbf{R}_0^+)^2 \mid s \leq t\}$. A family $\{T_{st}\}_{(s,t) \in I}$ of mixture endomorphisms satisfying (i) $T_{tt} = id_{\mathbf{V}}$ for all $t \in \mathbf{R}_0^+$ and (ii) $T_{rt} = T_{st} T_{rs}$ for all $r, s, t \in \mathbf{R}_0^+$ with $r \leq s \leq t$ is called a *dynamical family.* One-parameter semigroups $\{T_t\}_{t \in \mathbf{R}_0^+}$ of mixture endomorphisms and one-parameter groups $\{T_t\}_{t \in \mathbf{R}}$ of mixture automorphisms are called *dynamical semigroups* and *dynamical groups,* respectively.

If the systems under consideration are developing in time according to a dynamical family $\{T_{st}\}_{(s,t) \in I}$, then the probability for the outcome 1 of an effect $l \in L$ at time $t \geq s$ in an ensemble $v \in K$ prepared at the time $s \geq 0$ is given by

$$\langle T_{st} v, l \rangle = \langle v, T'_{st} l \rangle$$

A dynamical semigroup corresponds to a dynamical family homogeneous in time, and a dynamical semigroup of mixture automorphisms can be extended to a dynamical group.

Now we will introduce *observables,* which, physically, are defined as statistically equivalent measuring apparatus with arbitrarily many outcomes. The mathematical definition of observables is the following one.

Definition 3.3. Let a statistical duality $\langle \mathbf{V}, \mathbf{W} \rangle$ and a measurable space (M, Ξ) be given. An *observable* on (M, Ξ) is defined as an effect-valued measure on Ξ, i.e., an observable is a map $F: \Xi \to L$, $B \to F(B)$, with the following properties:

(i) $F(\varnothing) = 0$, $F(M) = e$
(ii) $F(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} F(B_i) := \sigma(\mathbf{W}, \mathbf{V})\text{-}\lim_{n \to \infty} \sum_{i=1}^{n} F(B_i)$ for every sequence of mutually disjoint sets $B_i \in \Xi$.

We call F a *decision observable* if $F(\Xi) \subseteq \partial_e L$ (where $\partial_e L$ denotes the set of extreme points of L).

Remark 3.2. According to (ii), we postulate that $\sigma(\mathbf{W}, \mathbf{V})$-$\lim_{n\to\infty} \sum_{i=1}^{n} F(B_i)$ exists and is equal to $F(\bigcup_{i=1}^{\infty} B_i)$. If $\langle \mathbf{V}, \mathbf{W} \rangle$ is a generalized probability space, then $\sigma(\mathbf{W}, \mathbf{V})$-convergence of $\{\sum_{i=1}^{n} F(B_i)\}_{n\in\mathbf{N}}$ follows from the finite additivity of F.

For an ensemble $v \in K$, the observable F on (M, Ξ) defines a probability measure P_v^F on Ξ by

$$P_v^F(B) := \langle v, F(B) \rangle$$

We interpret M as the "value space" of F and $P_v^F(B)$ as the probability for occurrence of values of F in the set $B \in \Xi$. We call P_v^F the *probability distribution of F in the ensemble v*. In the special case $(M, \Xi) := (\mathbf{R}, \Xi(\mathbf{R}))$, where $\Xi(\mathbf{R})$ are the Borel sets of $\mathbf{R}$, the *expectation value of the observable F in v* is defined by

$$\langle F \rangle_v := \int \xi P_v^F(d\xi)$$

provided that the integral exists. If it does even exist for every $v \in K$ and if $\langle \mathbf{V}, \mathbf{W} \rangle$ is a generalized probability space, then the *F-integral* $\int \xi F(d\xi)$ exists as that uniquely determined element of $\mathbf{W}$ that satisfies $\langle v, \int \xi F(d\xi) \rangle = \int \xi \langle v, F(d\xi) \rangle$ for all $v \in \mathbf{V}$ [see Stulpe (1986) and the end of Section 5], and the expectation values of F are also given by

$$\langle F \rangle_v = \left\langle v, \int \xi F(d\xi) \right\rangle$$

where $v \in K$.

Instruments are associated with statistically equivalent preparative measurements of observables.

Definition 3.4. Let a statistical duality $\langle \mathbf{V}, \mathbf{W} \rangle$ and a measurable space (M, Ξ) be given. An *instrument J* on (M, Ξ) is an operation-valued measure on Ξ, i.e., J is a map $B \to J(B)$ assigning to every measurable set an operation such that the following conditions are satisfied:

(i) $J(\emptyset) = 0$
(ii) $J(M)v \in K$ for all $v \in K$
(iii) $J(\bigcup_{i=1}^{\infty} B_i)v = \sum_{i=1}^{\infty} J(B_i)v := \sigma(\mathbf{V}, \mathbf{W})$-$\lim_{n\to\infty} \sum_{i=1}^{n} J(B_i)v$ for every sequence of mutually disjoint sets $B_i \in \Xi$ and all $v \in \mathbf{V}$.

Remark 3.3.
(a) $J(M)$ is a mixture endomorphism.

(b) For $v \geq 0$, the sequence $\{\sum_{i=1}^{n} J(B_i)v\}_{n\in\mathbf{N}}$ according to (iii) is increasing and norm-bounded. This implies that the limit exists also in norm for any $v \in \mathbf{V}$.

Given an ensemble $v \in K$, then the following interpretation is evident:

(i) $\|J(B)v\| = \langle J(B)v, e\rangle = \langle v, J'(B)e\rangle$ is the probability that the observable "measured by J" takes values in the set $B \in \Xi$.

(ii) If $J(B)v \neq 0$, $J(B)v/\|J(B)v\|$ is the ensemble obtained by selection of systems according to outcomes in B.

By (i), the instrument J determines the observable $F := J'(\cdot)e$.

Note that, in general, $J(M)$ is not equal to the identity in $\mathbf{V}$. This is because a measuring apparatus representing J may influence the systems during the measurement. The additivity of J can be explained as follows. Let $B, B_1, B_2 \in \Xi$, $B = B_1 \cup B_2$, and $B_1 \cap B_2 = \varnothing$. Physically, the ensemble $J(B)v/\|J(B)v\|$ should be a mixture of $J(B_1)v/\|J(B_1)v\|$ and $J(B_2)v/\|J(B_2)v\|$ in ratio

$$P_v^F(B_1) : P_v^F(B_2) = \frac{P_v^F(B_1)}{P_v^F(B)} : \frac{P_v^F(B_2)}{P_v^F(B)}$$

That is,

$$\frac{J(B)v}{\|J(B)v\|} = \frac{\|J(B_1)v\|}{\|J(B)v\|}\frac{J(B_1)v}{\|J(B_1)v\|} + \frac{\|J(B_2)v\|}{\|J(B)v\|}\frac{J(B_2)v}{\|J(B_2)v\|}$$

and thus $J(B) = J(B_1) + J(B_2)$.

4. GENERALIZED CONDITIONAL EXPECTATIONS

In this section, we presuppose $\langle \mathbf{V}, \mathbf{W}\rangle$ to be a generalized probability space and (M, Ξ) a measurable space. Let an instrument J on (M, Ξ) determining the observable $F := J'(\cdot)e$, a fixed ensemble $v \in K$, and an effect $l \in L$ be given.

Consider the successive measurement of $F = J'(\cdot)e$ and l in the ensemble v. This means that l is measured in the ensemble $J(M)v$ obtained by nonselective measurement of F. The probability for outcomes of the observable F in the set $B \in \Xi$ is given by

$$P_v^F(B) := \langle v, F(B)\rangle = \langle J(B)v, e\rangle = \|J(B)v\| \tag{1}$$

and the joint probability for occurrence of values of F in B and for the outcome 1 of the effect l by

$$P_v^F(B)\left\langle \frac{J(B)v}{\|J(B)v\|}, l\right\rangle = \langle J(B)v, l\rangle = \langle v, J'(B)l\rangle$$

Now let l be an arbitrary element of **W**. We have that $B \to \langle J(B)v, l\rangle$ defines a bounded signed measure on Ξ which is absolutely continuous with respect to the probability measure P_v^F. Indeed, because of (1), it follows from $P_v^F(B)=0$ that $\langle J(B)v, l\rangle = 0$. Hence, we can apply the Radon–Nikodym theorem and obtain the following statement.

Proposition 4.1. For any $l \in \mathbf{W}$ there exists a Ξ-measurable function $E_v^J(l)$: $M \to \mathbf{R}$ such that

$$\langle J(B)v, l\rangle = \int_B E_v^J(l)\, dP_v^F \tag{2}$$

holds for all $B \in \Xi$. This function is determined uniquely P_v^F-a.e.

Definition 4.1. We call $E_v^J(l)$ a *version of the generalized conditional expectation of l given the instrument J and the ensemble v*; the equivalence class $[E_v^l(l)]$ is the *generalized conditional expectation of l given J and v.* For an effect $l \in L$, $E_v^J(l)$ is also called a *version of the generalized conditional probability of l.*

Let us remark that this definition is analogous to the definition of the corresponding notions in classical probability theory [see, e.g., Bauer (1972) and Section 1]. The type of generalized conditional expectation established here was introduced by Cycon and Hellwig (1977) implicity, for the quantum case explicitly by Hellwig and Stulpe (1983; Hellwig, 1981) and Ozawa (1985a), and for the general case explicitly by Stulpe (1986).

To give an interpretation of generalized conditional expectations, assume that there is a set $B \in \Xi$ and a version $E_v^J(l)$ such that $E_v^J(l)$ is constant on B and $P_v^F(B) \neq 0$ (e.g., $B := \{x\}$ for $x \in M$ with $\{x\} \in \Xi$ and $P_v^F(\{x\}) \neq 0$). Then equations (2) and (1) yield

$$E_v^J(l)(x) = \frac{\langle J(B)v, l\rangle}{P_v^F(B)} = \left\langle \frac{J(B)v}{\|J(B)v\|}, l \right\rangle \tag{3}$$

for all $x \in B$. This shows that for an effect $l \in L$ the value of any version $E_v^J(l)$ of the generalized conditional probability at each point $x \in B$ is just the elementary conditional probability which coincides with the probability for the outcome 1 of l in the ensemble obtained by the selection procedure according to outcomes of F in the set B. This remark motivates the denotations "generalized conditional probability" and "generalized conditional expectation," respectively.

The properties of versions of conditional expectations are listed in the next theorem.

Theorem 4.1. The following statements are valid for the versions of the generalized conditional expectations:

(i) $E_v^J(\alpha l + \beta \tilde{l}) = \alpha E_v^J(l) + \beta E_v^J(\tilde{l})$ P_v^F-a.e. for $l, \tilde{l} \in \mathbf{W}$ and $\alpha, \beta \in \mathbf{R}$

(ii) $E_v^J(l) \leq E_v^J(\tilde{l})$ P_v^F-a.e., where $l, \tilde{l} \in \mathbf{W}$ and $l \leq \tilde{l}$
(iii) $E_v^J(0) = 0$ and $E_v^J(e) = 1$ P_v^F-a.e.
(iv) $|E_v^J(l)| \leq \|l\|$ P_v^J-a.e. for $l \in \mathbf{W}$
(v) If $\{l_n\}_{n \in \mathbf{N}}$ is an increasing, norm-bounded sequence in $\mathbf{W}$, then

$$E_v^J\left(\sup_{n \in \mathbf{N}} l_n\right) = \sup_{n \in \mathbf{N}} E_v^J(l_n)$$

holds P_v^F-a.e.

Proof. From

$$\int_B E_v^J(\alpha l + \beta \tilde{l})\, dP_v^F = \langle J(B)v, \alpha l + \beta \tilde{l} \rangle = \int_B (\alpha E_v^J(l) + \beta E_v^J(\tilde{l}))\, dP_v^F$$

for all $B \in \Xi$ it follows that $E_v^J(\alpha l + \beta \tilde{l}) = \alpha E_v^J(l) + \beta E_v^J(\tilde{l})$ holds P_v^F-a.e. Because of $v \in K$ we have $\langle J(B)v, l \rangle \leq \langle J(B)v, \tilde{l} \rangle$ for $l \leq \tilde{l}$, which implies $E_v^J(l) \leq E_v^J(\tilde{l})$ P_v^F-a.e. Moreover, $\langle J(B)v, e \rangle = P_v^F(B) = \int_B \chi_M\, dP_v^F$ gives $E_v^J(e) = \chi_M$ P_v^F-a.e. (χ_A denotes the characteristic function of any subset $A \subseteq M$).

Now, from $-\|l\| e \leq l \leq \|l\| e$ we obtain $-\|l\| \chi_M \leq E_v^J(l) \leq \|l\| \chi_M$ P_v^F-a.e., resp. $|E_v^J(l)(x)| \leq \|l\|$ for P_v^F-almost all $x \in M$.

Let $\{l_n\}_{n \in \mathbf{N}}$ be an increasing, norm-bounded sequence in $\mathbf{W}$ with $l := \sup_{n \in \mathbf{N}} l_n$. The versions $E_v^J(l_n)$ satisfy

$$-\|l_1\| \leq E_v^J(l_1) \leq E_v^J(l_n) \leq E_v^J(l_{n+1}) \leq E_v^J(l) \leq \|l\| \quad P_v^F\text{-a.e}$$

which implies that $\sup_{n \in \mathbf{N}} E_v^J(l_n)$ exists P_v^F-a.e. and is P_v^F-a.e. equal to a bounded, measurable function. The definition of $E_v^J(l)$ and $E_v^J(l_n)$ and the monotone convergence theorem (or the dominated convergence theorem) now yield

$$\int_B E_v^J\left(\sup_{n \in \mathbf{N}} l_n\right) dP_v^F = \left\langle J(B)v, \sup_{n \in \mathbf{N}} l_n \right\rangle = \sup_{n \in \mathbf{N}} \langle J(B)v, l_n \rangle$$

$$= \sup_{n \in \mathbf{N}} \int_B E_v^J(l_n)\, dP_v^F = \int_B \sup_{n \in \mathbf{N}} E_v^J(l_n)\, dP_v^F$$

Since $B \in \Xi$ is arbitrary, it follows the assertion (v). ■

Corollary 4.1. The map $\varepsilon_v^J\colon \mathbf{W} \to \mathbf{L}_{\mathbf{R}}^\infty(M, \Xi, P_v^F)$ defined by $\varepsilon_v^J(l) := [E_v^J(l)]$ is linear, positive, and normal and satisfies $\|\varepsilon_v^J\| = 1$ and $\varepsilon_v^J(e) = [\chi_M]$.

Definition 4.2. We call ε_v^J the *generalized conditional expectation given J and v.*

5. GENERALIZED CONDITIONAL DISTRIBUTIONS

Again, we suppose $\langle \mathbf{V}, \mathbf{W} \rangle$ is a generalized probability space. Let us consider the successive measurement of two observables $F := J'(\cdot)e$ and G in an ensemble $v \in K$, the first one associated with an instrument J on some measurable space (M, Ξ) and the second one defined on some measurable space (M', Ξ'). In this situation, the joint probabilities

$$\langle J(B)v, G(B') \rangle = \int_B E_v^J(G(B'))\, dP_v^F$$

$(B \in \Xi, B' \in \Xi')$ are of interest. Some properties of the versions of the corresponding generalized conditional expectations can easily be derived from Theorem 4.1 and are quoted in the following proposition.

Proposition 5.1. For versions $E_v^J(G(B'))$ of the generalized conditional probabilities of $G(B')$, $B' \in \Xi$, the following equalities and inequalities hold P_v^F-a.e.:

(i) $0 \leq E_v^J(G(B')) \leq 1$
(ii) $E_v^J(G(\varnothing)) = 0$, $E_v^J(G(M')) = 1$
(iii) $E_v^J(G(B_1')) \leq E_v^J(G(B_2'))$ for $B_1' \subseteq B_2'$
(iv) $E_v^J(G(\bigcup_{i=1}^{\infty} B_i')) = \sum_{i=1}^{\infty} E_v^J(G(B_i'))$ for $B_i' \cap B_j' = \varnothing \ (i \neq j)$

These statements do not imply that $B' \to E_v^J(G(B'))(x)$ is a probability measure on Ξ' for P_v^F-almost every $x \in M$, because the equation (iv), for instance, only holds up to a set of measure zero depending on the sequence of disjoint sets B_i' and, in general, there are more than countably many such sequences. However, as we shall show by Theorem 5.1, under a technical assumption it is always possible to choose the versions $E_v^J(G(B'))$ in such a manner that they define a probability measure for each $x \in M$. In this case the map $P\colon M \times \Xi' \to [0, 1]$ given by

$$P(x, B') := E_v^J(G(B'))(x)$$

is a *Markov kernel*, i.e., P is a Ξ-measurable function in the first argument and a probability measure on Ξ' in the second one.

Definition 5.1. A Markov kernel $P := P_v^{J,G}$ on $M \times \Xi'$ is called a *generalized conditional distribution of the observable G given the instrument J and the ensemble v* if for every $B' \in \Xi'$, $P_v^{J,G}(\cdot, B')$ is a version of the conditional expectation of $G(B')$, i.e.,

$$P(\cdot, B') := P_v^{J,G}(\cdot, B') = E_v^J(G(B'))$$

P_v^F-a.e.

The next proposition and the following theorem give information about existence and uniqueness of generalized conditional distributions; they are closely related to corresponding classical statements (cf. Bauer, 1972).

Proposition 5.2. Let P_1 and P_2 be two generalized conditional distributions of G given (J, v). Then, for every $B' \in \Xi'$, there is a set $N_{B'} \in \Xi$ with $P_v^F(N_{B'}) = 0$ such that

$$P_1(x, B') = P_2(x, B') \tag{4}$$

is fulfilled for all $x \in M \setminus N_{B'}$. If the σ-algebra Ξ' is countably generated, then there exists a set N of measure zero such that equation (4) holds for all $x \in M \setminus N$ and all $B' \in \Xi'$.

Proof. The first assertion is immediately implied by the definition of conditional distributions. To prove the second one, let Γ' be a countable generator of Ξ' which can be assumed to contain M' and to be stable against finite intersections. For each set $B' \in \Gamma'$ there exists a set $N_{B'}$ of measure zero so that (4) holds for $x \notin N_{B'}$. The countable union $N := \bigcup_{B' \in \Gamma'} N_{B'}$ is also a set of measure zero, and (4) is fulfilled for any $x \notin N$ and any $B' \in \Gamma'$. That is, for $x \notin N$ the measures $P_1(x, \cdot)$ and $P_2(x, \cdot)$ coincide on Γ', and hence, by a well-known uniqueness theorem, they are equal. ■

Under technical assumptions concerning the measurable spaces (M, Ξ) and (M', Ξ'), it was proved by Davies and Lewis (1970, Davies 1976) by means of functional-analytic methods that a probability measure μ on the product $\Xi \otimes \Xi'$ is defined by $\mu(B \times B') := \langle J(B)v, G(B')\rangle$. Using this fact, the following theorem can be derived from the corresponding classical statement. However, we prove it by mimicking the classical proof, for four reasons. First, the classical statement is a special case of our general one, as we shall see in Section 6. Second, our proof requires only probabilistic methods. Third, we have to make a technical presupposition only for (M', Ξ'). Finally, we wish to present a self-contained formulation: a polish space is a topological space whose topology can be derived from a complete and separable metric.

Theorem 5.1. Let M' be a polish space and $\Xi' := \Xi(M')$ the σ-algebra of its Borel sets. Then there exists a generalized conditional distribution P of G given J and v.

Proof. If Γ' is a class of subsets of M' and $\tilde{\Gamma}' := \Gamma' \cup \{C_{M'}B' | B' \in \Gamma'\} \cup \{M', \varnothing\}$ ($C_{M'}B'$ denotes the complement of B'), then the algebra Δ' generated by Γ' consists of all finite unions of finite intersections of sets of $\tilde{\Gamma}'$. Now let Γ' be a countable generator of $\Xi(M')$; then Δ' is a countable generator of $\Xi(M')$, which is an algebra. Denote the elements of Δ' by D_i', where $i \in \mathbb{N}$ (replace $i \in \mathbb{N}$ by $i = 1, 2, \ldots, r$ if the class Δ' is finite). $B' \to \langle J(M)v, G(B')\rangle$ is a probability measure on $\Xi(M')$. From the inner regularity of this finite measure on the Borel sets of the polish space M' it follows that for each $D_i' \in \Delta'$ there exists a sequence $\{K_{ij}'\}_{j \in \mathbb{N}}$ of compact subsets of

D'_i such that $K'_{ij} \subseteq K'_{i,j+1}$ and

$$\langle J(M)v, G(D'_i)\rangle = \sup_{j\in\mathbf{N}} \langle J(M)v, G(K'_{ij})\rangle \tag{5}$$

hold. Again, the algebra $\tilde{\Delta}'$ generated by Δ' and all sets K'_{ij} is countable.

Let $E^J_v(G(B'))$ be a version of the conditional expectation of $G(B')$, $B' \in \Xi(M')$, and define

$$\tilde{P}(x, B') := E^J_v(G(B'))(x) \tag{6}$$

where $x \in M$ and $B' \in \tilde{\Delta}'$. According to Proposition 5.1, we can choose versions $E^J_v(G(B'))$ that satisfy

$$0 \le \tilde{P}(x, B') \le 1, \qquad \tilde{P}(x, \varnothing) = 0, \qquad \tilde{P}(x, M') = 1 \tag{7}$$

for all $x \in M$. Moreover, Proposition 5.1 implies that for any finite sequence $B'_1, B'_2, \ldots, B'_n$ of mutually disjoint sets of $\tilde{\Delta}'$, the relation

$$\tilde{P}\left(x, \bigcup_{i=1}^{n} B'_i\right) = \sum_{i=1}^{n} \tilde{P}(x, B'_i) \tag{8}$$

is fulfilled for P^F_v-almost every $x \in M$. Since $\tilde{\Delta}'$ is countable, there are only finitely many such sequences, and there exists a set N_0 of measure zero such that

$$B' \to \tilde{P}(x, B')$$

is finitely additive on $\tilde{\Delta}'$ for every $x \notin N_0$.

From the properties of the observable G as a measure and the monotony of the sequence $\{K'_{ij}\}_{j\in\mathbf{N}}$ it follows that $G(\bigcup_{j\in\mathbf{N}} K'_{ij}) = \sup_{j\in\mathbf{N}} G(K'_{ij})$ holds, which implies

$$E^J_v\left(G\left(\bigcup_{j\in\mathbf{N}} K'_{ij}\right)\right) = \sup_{j\in\mathbf{N}} \tilde{P}(\cdot, K'_{ij}) \tag{9}$$

P^F_v-a.e. according to (v) of Theorem 4.1 and equation (6). Furthermore, we have

$$\left\langle J(M)v, G\left(\bigcup_{j\in\mathbf{N}} K'_{ij}\right)\right\rangle = \sup_{j\in\mathbf{N}} \langle J(M)v, G(K'_{ij})\rangle$$

and therefore for the set $H'_i := D'_i \setminus \bigcup_{j\in\mathbf{N}} K'_{ij}$

$$0 = \langle J(M)v, G(H'_i)\rangle = \int E^J_v(G(H'_i))\, dP^F_v$$

where we have taken account of equation (5). Because of $E_v^J(G(H_i')) \geq 0$ P_v^J-a.e., this equation yields $E_v^J(G(H_i')) = 0$ P_v^F-a.e. Hence,

$$E_v^J(G(D_i')) = E_v^J\left(G\left(\bigcup_{j\in\mathbf{N}} K_{ij}'\right) + G(H_i')\right) = E_v^J\left(G\left(\bigcup_{j\in\mathbf{N}} K_{ij}'\right)\right)$$

holds P_v^J-a.e. From this, (9), and (6) we finally obtain

$$\tilde{P}(x, D_i') = \sup_{j\in\mathbf{N}} \tilde{P}(x, K_{ij}') \tag{10}$$

for all $x \notin N_i$, where the N_i are sets of measure zero.

The set $N := N_0 \cup \bigcup_{i\in\mathbf{N}} N_i$ is a set with $P_v^F(N) = 0$ such that the map $\tilde{P}: M \times \tilde{\Delta}' \to \mathbf{R}$ defined by equation (6) satisfies (7), (8) and (10) for all $x \in C_M N$, where B', $B_i' \in \tilde{\Delta}'$ and $D_i' \in \Delta'$. For every $x \in C_M N$, $\tilde{P}(x, \cdot)$ is a finitely additive measure on $\tilde{\Delta}'$ whose restriction to Δ' is continuous from above at $\varnothing$, as we will show now.

To that end, let $x \in C_M N$ and $\{B_n'\}_{n\in\mathbf{N}}$ be a sequence of sets of Δ' with $B_n' \supseteq B_{n+1}'$ and $\bigcap_{n\in\mathbf{N}} B_n' = \varnothing$. According to (10), for every $\varepsilon > 0$ and every $n \in \mathbf{N}$ there exists a compact set K_n' (namely a suitable K_{ij}') such that $K_n' \subseteq B_n'$ and

$$\tilde{P}(x, B_n') - \tilde{P}(x, K_n') = \tilde{P}(x, B_n' \backslash K_n') < \varepsilon/2^n$$

hold. From $\bigcap_{n\in\mathbf{N}} B_n' = \varnothing$ it follows that $\bigcap_{n\in\mathbf{N}} K_n' = \varnothing$, and the compactness of the K_n' implies the existence of a number $n_0 \in \mathbf{N}$ such that $\bigcap_{n=1}^{n_0} K_n' = \varnothing$. We conclude that for any $x' \in B_{n_0}'$ there is a number $j \in \mathbf{N}$, $1 \leq j \leq n_0$, such that $x' \in B_j' \backslash K_j'$. Thus,

$$B_{n_0}' \subseteq \bigcup_{i=1}^{n_0} B_i' \backslash K_i'$$

holds. Because $B' \to \tilde{P}(x, B')$ is a finitely additive measure on $\tilde{\Delta}'$, we obtain

$$\tilde{P}(x, B_{n_0}') \leq \sum_{i=1}^{n_0} \tilde{P}(x, B_i' \backslash K_i') < \sum_{i=1}^{n_0} \frac{\varepsilon}{2^i} < \varepsilon$$

In consequence, $\inf_{n\in\mathbf{N}} \tilde{P}(x, B_n') = 0$, and $B' \to \tilde{P}(x, B')$ is a finitely additive measure on the algebra Δ' which is continuous from above at $\varnothing$. Equivalently, for every $x \in C_M N$, $\tilde{P}(x, \cdot)$ is σ-additive on Δ'.

By a well-known theorem, for every $x \in C_M N$, $\tilde{P}(x, \cdot): \Delta' \to \mathbf{R}$ can uniquely be extended to a σ-additive measure $\hat{P}(x, \cdot)$ on the σ-algebra $\Xi(M')$, since the algebra Δ' generates $\Xi(M')$. The measure $\hat{P}(x, \cdot)$ is a probability measure according to (7). If the set N of measure zero is empty, define $P := \hat{P}$; otherwise, let μ be any probability measure on $\Xi(M')$ and

$$P(x, B') := \begin{cases} \mu(B') & \text{for} \quad x \in N \\ \hat{P}(x, B') & \text{for} \quad x \notin N \end{cases} \tag{11}$$

For each $x \in M$, $P(x, \cdot)$ is a probability measure on $\Xi(M')$. Equation (6), $N \in \Xi$, and the definition (11) of P imply that $P(\cdot, B')$ is a Ξ-measurable function for each $B' \in \Delta'$. We observe that the class Θ' of all sets $B' \in \Xi(M')$ for which $x \to P(x, B')$ is Ξ-measurable is monotone. Because Θ' contains Δ', it also contains the monotone class $\Theta(\Delta')$ generated by Δ', which coincides with the σ-algebra $\Sigma(\Delta')$ generated by the algebra Δ'. Summarizing, we obtain $\Xi(M') = \Sigma(\Delta') = \Theta(\Delta') \subseteq \Theta' \subseteq \Xi(M')$ and conclude $\Theta' = \Xi(M')$. Hence, $P: M \times \Xi(M') \to \mathbf{R}$ is a Markov kernel.

By construction of P, $P(\cdot, B')$ is a version of the conditional expectation of $G(B')$ for every $B' \in \Delta'$, i.e.,

$$\langle J(B)v, G(B')\rangle = \int_B P(\cdot, B')\, dP_v^F \tag{12}$$

holds for all $B' \in \Delta'$ and all $B \in \Xi$. If B is fixed, then

$$B' \to \langle J(B)v, G(B')\rangle, \qquad B' \to \int \chi_B\, P(\cdot, B')\, dP_v^F$$

are finite measures on $\Xi(M')$ coinciding on Δ', which implies their equality. Hence, equation (12) is valid for all $B \in \Xi$ and all $B' \in \Xi(M')$, i.e., $P(\cdot, B')$ is a version of the generalized conditional expectation of $G(B')$ and P is a conditional distribution of G. ■

In conclusion, the weak assumption that M' is a polish space and $\Xi' = \Xi(M')$ ensures the existence of generalized conditional distributions as well as their uniqueness in the sense that two conditional distributions differ at most on a set $N \times \Xi(M')$, where N is a set of measure zero. Nevertheless, there are important examples of observables and instruments for which the generalized conditional expectations and distributions can even be calculated and where (M', Ξ') need not be a polish space (Hellwig and Stulpe, 1983; Stulpe, 1986, 1987; this paper, Section 6).

Finally, the generalized conditional distribution of the observable G on (M', Ξ') is useful for calculating the conditional expectations for those $l \in \mathbf{W}$ that can be represented as a certain kind of weak integral of functions with respect to G. We call a Ξ'-measurable (not necessarily bounded) function $f: M' \to \mathbf{R}$ *integrable with respect to* G if for any $v \in \mathbf{V}$ the integral $\int f\, d\langle v, G(\cdot)\rangle =: \lambda(v)$ exists (equivalently, if for any $v \in K$, $\int f\, dP_v^G$ exists). Then, as a positive linear functional on $\mathbf{V}$, λ is bounded, and due to the properties of the measurable function f, it is given by an element $l \in \mathbf{W}$ because the observable G is effect-valued and $\langle \mathbf{V}, \mathbf{W}\rangle$ is a generalized probability space. Hence, we arrive at $\langle v, l\rangle = \int f\, d\langle v, G(\cdot)\rangle$ for all $v \in \mathbf{V}$ where the uniquely determined element $l =: \int f\, dG$ is called the *G-integral of* f (cf. Stulpe, 1986).

Proposition 5.3. If P is a generalized conditional distribution of G given (J, v) and f: $M' \to \mathbf{R}$ a G-integrable function, then

$$E_v^J\left(\int f\,dG\right)(x) = \int f(x')\,P(x, dx')$$

holds for P_v^F-almost all $x \in M$.

Proof: For characteristic functions $\chi_{B'}$, $B' \in \Xi'$, the assertion is true by definition; it is also valid for positive measurable functions with finitely many values, i.e., for positive simple functions. If f is any positive G-integrable function, then there exists a sequence $\{f_n\}_{n\in\mathbf{N}}$ of positive simple functions satisfying $0 \le f_n \le f_{n+1}$ such that $f = \sup_{n\in\mathbf{N}} f_n$, which implies $0 \le \int f_n\,dG \le \int f_{n+1}\,dG$ and $\int f\,dG = \sup_{n\in\mathbf{N}} \int f_n\,dG$. In consequence, by (v) of Theorem 4.1, we obtain

$$E_v^J\left(\int f\,dG\right) = E_v^J\left(\sup_{n\in\mathbf{N}} \int f_n\,dG\right) = \sup_{n\in\mathbf{N}} E_v^J\left(\int f_n\,dG\right)$$
$$= \sup_{n\in\mathbf{N}} \int f_n(x')\,P(\cdot\,, dx') = \int f(x')\,P(\cdot\,, dx')$$

to hold P_v^F-a.e., i.e., given a version $E_v^J(\int f\,dG)$, then for P_v^F-almost all $x \in M$, $\int f(x')\,P(x, dx')$ exists and is equal to $E_v^J(\int f\,dG)(x)$. Now our assertion for an arbitrary G-integrable function f follows from the decomposition $f = f^+ - f^-$, where f^+ is the positive part of f and f^- the negative part. ∎

6. *A POSTERIORI* ENSEMBLES: SOME EXAMPLES

In this last section, we first consider *a posteriori* ensembles in generalized probability theory, which were defined by Ozawa (1985a,b) in the von Neumann algebra formulation of quantum mechanics, and then we will discuss some main examples.

Definition 6.1. Let $\langle \mathbf{V}, \mathbf{W}\rangle$ be a generalized probability space, J an instrument on (M, Ξ), $F := J'(\cdot)e$, and $v \in K$ a fixed ensemble. A map Φ_v^J: $M \to K$, $x \to \Phi_v^J(x) =: \Phi_x$, is called a *family* $\{\Phi_x\}_{x\in M}$ *of a posteriori ensembles given J and v* if:

(i) Φ_v^J is $\mathbf{W}$-weakly Ξ-measurable, i.e., $\langle \Phi_v^J(\cdot), l\rangle$ is measurable for every $l \in \mathbf{W}$
(ii) For every $l \in \mathbf{W}$ and any version $E_v^J(l)$ of the generalized conditional expectation of l given J and v,

$$E_v^J(l) = \langle \Phi_v^J(\cdot), l\rangle \tag{13}$$

holds P_v^F-a.e.

In this context, we will call v the *a priori ensemble.*

Equation (13) is equivalent to

$$\langle J(B)v, l\rangle = \int_B \langle \Phi_v^J(x), l\rangle \, P_v^F(dx) \tag{14}$$

for all $B \in \Xi$ and all $l \in \mathbf{W}$. That is,

$$J(B)v = \int_B \Phi_v^J(x) \, P_v^F(dx) = \int \Phi_v^J \, dP_v^F \tag{15}$$

where the latter two integrals are understood in the $\mathbf{W}$-weak sense. If there exists a point $x \in M$ such that $\{x\} \in \Xi$ and $P_v^F(\{x\}) \neq 0$, then we obtain from (14)

$$\Phi_x = \Phi_v^J(x) = J(\{x\})v / \| J(\{x\})v \|$$

showing that Φ_x coincides with the ensemble obtained by the selection procedure according to the outcome x of F [compare equation (3)]. However, as (15) shows, in general Φ_v^J is only an ensemble-valued density of the positive-vector-valued measure $B \to J(B)v$ [Ludwig (1985, 1986) calls the measure $J(\cdot)v$ a *preparator of the ensemble* $J(M)v$].

If Φ_v^J is a family of *a posteriori* ensembles given (J, v) and G an observable on arbitrary (M', Ξ'), then

$$P(x, B') := P_v^{J,G}(x, B') := \langle \Phi_v^J(x), G(B')\rangle$$

$(x \in M, B' \in \Xi')$ defines a Markov kernel P, which, according to (13), is a generalized conditional distribution of G given (J, v). Thus, the existence of a family of *a posteriori* ensembles ensures the existence of generalized conditional distributions of any observable.

Our first example concerns classical probability theory, whose structure is at best reflected by the generalized probability space of Example 2.3(g). The ensembles are given as the probability measures on the measurable space (Ω, Σ), the effects as the measurable functions f fulfilling $0 \leq f \leq \chi_\Omega$, and the decision effects are the characteristic functions χ_A, $A \in \Sigma$. A random variable X on (Ω, Σ) into another measurable space (M, Ξ) is associated with its canonical instrument J^X defined by

$$J^X(B)\nu := \nu(X^{-1}(B) \cap \cdot) \tag{16}$$

$(B \in \Xi, \nu \in \mathbf{V})$, which determines the decision observable

$$B \to F^X(B) := (J^X)'(B)e = \chi_{X^{-1}(B)}$$

The probability distribution of F^X in an ensemble $\mu \in K$ is given by

$$P_\mu^{F^X}(B) := \langle \mu, F^X(B)\rangle = \int \chi_{X^{-1}(B)} \, d\mu = \mu(X^{-1}(B)) \tag{17}$$

Now let Y be a further random variable from (Ω, Σ) to (M', Ξ') and assume that a generalized conditional distribution P of $B' \to F^Y(B') := \chi_{Y^{-1}(B')}$ given (J^X, μ) exists. By Theorem 5.1, this is guaranteed if M' is a polish space and $\Xi' = \Xi(M')$. The Markov kernel P is defined by

$$\langle J^X(B)\mu, F^Y(B')\rangle = \int_B P(x, B')\, P_\mu^{F^X}(dx)$$

($B \in \Xi$, $B' \in \Xi'$), which, using (16) and (17), can be reformulated classically as

$$\mu(X^{-1}(B) \cap Y^{-1}(B')) = \int_B P(\cdot, B')\, d(\mu \circ X^{-1}) = \int_{X^{-1}(B)} P(\cdot, B') \circ X\, d\mu$$
$$= \int_{X^{-1}(B)} P(X(\omega), B')\, \mu(d\omega)$$

According to Proposition 5.3,

$$E_\mu^{J^X}\left(\int f dF^Y\right)(x) = \int f(x')\, P(x, dx') \tag{18}$$

holds for $P_\mu^{F^X}$-almost all $x \in M$ and every F^Y-integrable function $f\colon M' \to \mathbf{R}$. From

$$\left\langle \nu, \int f dF^Y \right\rangle = \int f d\langle \nu, F^Y(\cdot)\rangle = \int f d(\nu \circ Y^{-1}) = \int f \circ Y d\nu$$

being valid for all $\nu \in \mathbf{V}$ it follows that f is F^Y-integrable if and only if f is measurable and $f \circ Y$ bounded (equivalently, if f is an F^Y-essentially bounded, measurable function). Now, $\langle \nu, \int f dF^Y\rangle = \int f \circ Y d\nu = \langle \nu, f \circ Y\rangle$ implies

$$\int f dF^Y = f \circ Y$$

Hence, in the classical case equation (18) means the well-known formula

$$E^X(f \circ Y) := E_\mu^{J^X}(f \circ Y) = \int f(x')\, P(\cdot, dx') \tag{19}$$

holding $\mu \circ X^{-1}$-a.e. for all measurable functions f on M' with $f \circ Y \in \mathbf{W}$.

We conclude this example by proving that in classical probability theory there always exists a family of *a posteriori* ensembles given the canonical instrument J^X of any random variable $X\colon \Omega \to M$ and given any *a priori* ensemble $\mu \in K$, provided that Ω is a polish space and $\Sigma = \Xi(\Omega)$. To that end, consider a *conditional distribution P of $Y := id_\Omega$ given X*, which is also called an *expectation kernel.* Inserting $Y := id_\Omega$ into (19) yields

$$E^X(f) = E_\mu^{J^X}(f) = \int f(\omega)\, P(\cdot, d\omega)$$

$\mu \circ X^{-1}$-a.e. for every $f \in \mathbf{W}$. Now the probability measures $\Phi_x = \Phi_\mu^{J^X}(x) := P(x, \cdot)$ on Σ constitute a family $\{\Phi_x\}_{x \in M}$ of *a posteriori* ensembles because $x \to \langle \Phi_x, f\rangle = \int f(\omega)\, P(x, d\omega)$ is a measurable, bounded function on M and

$$E_\mu^{J^X}(f)(x) = \langle \Phi_x, f\rangle$$

holds for $P_\mu^{F^X}$-almost all $x \in M$ and all $f \in \mathbf{W}$. Conversely, it is easy to show that a family of *a posteriori* ensembles defines an expectation kernel. Thus, in classical probability theory families of *a posteriori* ensembles and expectation kernels are corresponding bijectively, and the condition (Ω, Σ) being a polish space is sufficient for the existence of both.

Next let us consider usual quantum mechanics as established by the generalized probability space Example 2.3(c). Here the ensembles are described by the statistical operators acting in the Hilbert space $\mathbf{H}$, the effects by the bounded self-adjoint operators A fulfilling $0 \le A \le 1$ (1 denoting the unit operator), and the decision effects are the orthogonal projections in $\mathbf{H}$. We will calculate the *quantum conditional expectations* given a *von Neumann-Lüders instrument* J and an ensemble $W \in K$. Let M be the set $\mathbf{N}$ of positive integers or a finite subset, $\Xi := \Pi(M)$ the power set of M, and let $\{E_n\}_{n \in \mathbf{N}}$ be a sequence of orthogonal projections corresponding to an orthogonal decomposition of $\mathbf{H}$. Then such an instrument may be defined on (M, Ξ) by

$$J(B)V := \sum_{i \in B} E_i V E_i \tag{20}$$

($B \in \Xi$, $V \in \mathbf{V}$), where the possibly infinite sum converges with respect to the trace norm. J determines the decision observable F given by

$$F(B) := J'(\cdot)e = \sum_{i \in B} E_i$$

[the sum converging in the strong operator topology in $\mathbf{B}(\mathbf{H})$]; the probability distribution of F in the ensemble W is given by

$$P_W^F(B) := \langle W, F(B)\rangle = \operatorname{tr} WF(B) = \sum_{i \in B} \operatorname{tr} WE_i \tag{21}$$

Insertion of (20) and (21) into equation (2) now yields

$$\operatorname{tr}(J(B)W)A = \sum_{i \in B} \operatorname{tr} E_i W E_i A = \sum_{i \in B} E_W^J(A)(i) \operatorname{tr} WE_i$$

for any bounded, self-adjoint operator $A \in \mathbf{W}$ and all $B \in \Xi$. Hence, we obtain

$$E^J_W(A)(i) = \frac{\operatorname{tr} E_i W E_i A}{\operatorname{tr} W E_i}$$

for every $i \in M$ with $\operatorname{tr} WE_i \neq 0$ (if $\operatorname{tr} WE_i = 0$, then $\{i\}$ is a set of measure zero). Moreover, for a family $\{\Phi_i\}_{i \in M}$ of *a posteriori* ensembles

$$\Phi_i = \Phi^J_W(i) = \frac{E_i W E_i}{\operatorname{tr} W E_i} = \frac{J(\{i\}) W}{\| J(\{i\}) W \|}$$

necessarily holds if $\operatorname{tr} WE_i \neq 0$, and the conditional distribution P of an observable G on an arbitrary measurable space (M', Ξ') is given by

$$P(i, B') = \operatorname{tr} \Phi_i G(B') = \frac{\operatorname{tr} E_i W E_i G(B')}{\operatorname{tr} W E_i}$$

($i \in M$, $B' \in \Xi'$, $\operatorname{tr} WE_i \neq 0$).

As final example we mention the so-called *nuclear instruments*, which can be defined in quantum mechanics as well as in the general theory. Let a generalized probability space $\langle \mathbf{V}, \mathbf{W} \rangle$ and an observable F on (M, Ξ) be given, and let $\Phi \colon M \to K$ be a map such that for every $B \in \Xi$ the map $\chi_B(\cdot)\Phi(\cdot)$ is $\mathbf{W}$-weakly integrable with respect to every measure $\langle v, F(\cdot)\rangle$, $v \in \mathbf{V}$ [this property of Φ follows from its strong measurability, and in the von Neumann algebra case as described in Example 2.3(d), solely the $\mathbf{W}$-weak measurability of Φ is sufficient]. Under these assumptions F and Φ define an instrument by the $\mathbf{W}$-weak integral

$$J(B)v := J^{F,\Phi}(B)v := \int_B \Phi(x)\langle v, F(dx)\rangle \tag{22}$$

($B \in \Xi$, $v \in \mathbf{V}$); conversely, J determines the observable $F = J'(\cdot)e$ uniquely, but in general not Φ. Instruments of the form $J = J^{F,\Phi}$ are called *nuclear.* By comparison of (22) with (15), resp. (14), we see that a family $\{\Phi_x\}_{x \in M}$ of *a posteriori* ensembles given J and $v \in K$ is constituted by

$$\Phi_x = \Phi^J_v(x) = \Phi^{J^{F,\Phi}}_v(x) := \Phi(x)$$

resp., $\Phi^J_v := \Phi$. In particular, $\{\Phi_x\}_{x \in M}$ does not depend on $v \in K$.

Applying abstract operator-theoretic methods, Ozawa (1985b) proved the surprising result that in conventional quantum mechanics there always exists a family of *a posteriori* ensembles given any arbitrary instrument and any ensemble. By an example he pointed out that, however, this is not true for the von Neumann algebra case. Thus, as expected, in generalized probability theory families of *a posteriori* ensembles do not exist in general, whereas generalized conditional distributions for observables defined on polish spaces do always exist, as we have shown by application of methods of classical probability theory.

ACKNOWLEDGMENTS

The author is obliged to Prof. S. P. Gudder for reading and commenting on the manuscript. He also thanks the Department of Mathematics and Computer Science of the University of Denver for use of its facilities. He is grateful for a Wigner scholarship granted by the Technical University of Berlin.

REFERENCES

Bauer, H. (1972). *Probability Theory and Elements of Measure Theory*, Holt, Rinehart and Winston, New York.

Cycon, H., and Hellwig, K.-E. (1977). Conditional expectations in generalized probability theory, *Journal of Mathematical Physics*, **18**, 1154–1161.

Davies, E. B. (1976). *Quantum Theory of Open Systems*, Academic Press, London.

Davies, E. B., and Lewis, J. T. (1970). An operational approach to quantum probability, *Communications in Mathematical Physics*, **17**, 239–260.

Edwards, C. M. (1970). The operational approach to algebraic quantum theory I, *Communications in Mathematical Physics*, **16**, 207–230.

Edwards, C. M. (1971). Classes of operations in quantum theory, *Communications in Mathematical Physics*, **20**, 26–56.

Gudder, S. P. (1979a). Axiomatic operational quantum mechanics, *Reports on Mathematical Physics*, **16**, 147–166.

Gudder, S. P. (1979b). *Stochastic Methods in Quantum Mechanics*, North-Holland, Amsterdam.

Gudder, S. P., and Marchand, J.-P. (1977). Conditional expectations on von Neumann algebras: A new approach, *Reports on Mathematical Physics*, **12**, 317–329.

Haag, R., and Kastler, D. (1964). An algebraic approach to quantum field theory, *Journal of Mathematical Physics*, **5**, 848–861.

Hellwig, K.-E. (1981). Conditional expectations and duals of instruments, in *Grundlagen der Exakten Naturwissenschaften*, Vol. **5**, H. Neumann, ed., pp. 113–124, Bibliographisches Institut, Mannheim.

Hellwig, K.-E., and Stulpe, W. (1983). A formulation of quantum stochastic processes and some of its properties, *Foundations of Physics*, **13**, 673–699.

Kraus, K. (1983). *States, Effects, and Operations*, Lecture Notes in Physics **190**, Springer-Verlag, Berlin.

Ludwig, G. (1983). *Foundations of Quantum Mechanics I*. Springer-Verlag, New York.

Ludwig, G. (1985). *An Axiomatic Basis for Quantum Mechanics*, Vol. I, *Derivation of Hilbert Space Structure*, Springer-Verlag, Berlin.

Nagel, R. J. (1974). Order unit and base norm spaces, in *Foundations of Quantum Mechanics and Ordered Linear Spaces*, A. Hartkaemper and H. Neumann, eds., pp. **29**, 23–29, Springer-Verlag, Berlin.

Nakamura, M., and Umgaki, H. (1962). On von Neumann's theory of measurement in quantum statistics, *Mathematica Japonica*, **7**, 151–157.

Ozawa, M. (1985a). Concepts of conditional expectations in quantum theory, *Journal of Mathematical Physics*, **26**, 1948–1955.

Ozawa, M. (1985b). Conditional probability and *a posteriori* states in quantum mechanics, *Publications RIMS Kyoto University*, **21**, 279–295.

Stulpe, W. (1986). Bedingte Erwartungen und stochastische Prozesse in der generalisierten Wahrscheinlichkeitstheorie—Beschreibung sukzessiver Messungen mit zufälligem Ausgang, Doctoral dissertation, Berlin.

Stulpe, W. (1987). Conditional expectations and stochastic processes in quantum probability, in *Information, Complexity, and Control in Quantum Physics*, A. Blaquiere, S. Diner, and G. Lochak, eds., pp. 223–234, CISM Courses and Lectures No. **294**, Springer-Verlag, Vienna.

Umegaki, H. (1954). Conditional expectations in an operator algebra I, *Tohoku Mathematics Journal*, **6**, 177–181.

Umegaki, H. (1964). Conditional expectations in an operator algebra IV, *Kodai Mathematics Seminar Report*, **14**, 59–85.

Werner, R. (1983). Physical uniformities on the state space of nonrelativistic quantum mechanics, *Foundations of Physics*, **13**, 859–881.