*A Note on the Concept of Entropy*¹

I. E. SEGAL

Communicated by A. ZYGMUND

Introduction. From a physical point of view, there is no reason why entropy should not be defined for any state of any physical system. The present paper gives a mathematical formulation that is roughly of this order of generality, which it validates by deriving certain properties. It is shown that the entropy of a mixture of states is bounded from below by the corresponding average of the entropies of the constituent states, that the entropy of a state is increased by restricting it to a subsystem, and that pure and equilibrium states tend to be characterized by their entropies. No difficult technical considerations are involved, although some of the problems that are led up to seem to involve quite non-trivial questions of technique.

As is well known, entropy appears originally in thermodynamics, plays a part in the related subject of statistical mechanics, extends appropriately to quantum mechanics, and more recently has emerged as a fundamental notion for information theory, in a form somewhat related to earlier and simpler ideas concerning statistical estimation (see notably [7] and [8]). What is common to all these areas is the existence of a physical system that may be defined through its observables (or "random variables" in probability theory) and a concern with the states of this system as defined through the expectation values of the observables. A viable theory dealing with such matters has been set up on the basis of a few quite conservative postulates in [3], but the virtually exclusive theoretical prototype for the physical systems that have arisen or are even contemplated is included in the special case in which the observables may be imbedded in an associative algebra with a related adjunction operation, which has the effect that, technically, the observables may be represented by the self-adjoint elements of an algebra of operators in Hilbert space. Actually, the analytical form of the systems that arise in quantum mechanics and probability theory is such that they have certain general properties beyond that of being represented by an algebra of operators that is closed under uniform convergence and under the adjunction operation, which is all that develops from the general phenomenological and elementary algebraic considerations

¹ Research supported in part by the Office of Scientific Research.

described. Quite often these systems are, or can be technically related to, a ring of operators in the sense of MURRAY and VON NEUMANN, and the states of fundamental physical importance satisfy not merely the simple phenomenological requirements involved above, but such requirements as that the expectation values of observables depending on certain physical parameters should be continuous functions of the parameters (*cf. e.g.* Section 3 of [4]). Suffice it to say that, except for the case of quantum systems of an infinite number of degrees of freedom, all major theories satisfy the simple assumptions that the observables form (or can be extended in a physically meaningful way to form) a ring of operators, and that the physically strictly realizable states are strongly continuous on the unit sphere in this ring. In addition there is given in all cases a species of normalization—*e.g.* the equal weighting assigned to the letters of an alphabet in information theory—which can be formulated in compact terms by assuming given a reference integral or “trace” for suitable observables. It is within this phenomenological framework that entropy is primarily dealt with in the following.

Infinite quantum systems, however, are properly represented by systems of observables that are not such rings, but merely approximable by such in the sense of uniform convergence (*cf. loc. cit.*). It appears that the theory can be extended to such limit algebras, although the case of a perfectly general algebra of operators that is closed under uniform convergence and adjunction may well be beyond the scope of a useful analytical treatment.

Mathematical development. Let $\mathfrak{A}$ be a ring of operators in a complex Hilbert space $\mathfrak{H}$; a *regular* state of $\mathfrak{A}$ is defined as one that is continuous on the unit sphere in $\mathfrak{A}$ in the strong (operator) topology. We assume given a distinguished gage m on $\mathfrak{A}$ in the sense of [5]; this is a completely additive non-negative-valued function on the projections in $\mathfrak{A}$ that is invariant under transformation by the unitary operators in $\mathfrak{A}$, and is such that every non-zero projection in $\mathfrak{A}$ bounds a projection on which m is finite; it is no essential loss of generality to assume, as will be done here, that m vanishes only on the zero projection. (In the case of a conventional measure space, $\mathfrak{A}$ is essentially the set of all bounded measurable functions, while $m(P)$ is the measure of the set whose characteristic function is P .) The following definition is essentially in conformity with the usage in information theory (*cf.* [7]) and in quantum statistics (*cf.* [8]).

Definition 1. The *entropy* of a regular state E of an operator ring $\mathfrak{A}$, with respect to a gage m on $\mathfrak{A}$, is said to exist in case $D \log D$ has a well defined integral with respect to m (as an extended real number), where D is the derivative dE/dm (*i.e.* the unique operator D such that $E(X) = m(XD)$ for all $X \in \mathfrak{A}$), and is then defined as $-m(D \log D)$ and denoted by $e(E)$ (or $e(D, m)$ if the dependence on m is particularly relevant). (That m extends by linearity, *etc.*, to suitable elements of $\mathfrak{A}$, in such a manner that E has the stated form, is shown in [5], esp. Th. 14, which shows also that D is non-negative and self-adjoint.)

In particular, defining E as *bounded* in case D is a bounded operator, such a state has finite entropy.

Theorem 1.² *The entropy is a concave function on the bounded states.*

Let K be the collection of all bounded states on the ring $\mathcal{A}$, relative to the gage m . It must be shown that if E and E' are in K , and if $0 < \alpha < 1$ and $\alpha + \alpha' = 1$, then

$$\alpha e(E) + \alpha' e(E') \leq e(\alpha E + \alpha' E')$$

(it is obvious that $\alpha E + \alpha' E'$ is in K). Now the derivatives dE/dm and dE'/dm , say D and D' , are in L^p , where the space L^p relevant here is that defined in [2]. Then by the analogue of the Minkowski inequality (*loc. cit.*)

$$[m((\alpha D + \alpha' D')^p)]^{1/p} \leq \alpha [m(D^p)]^{1/p} + \alpha' [m(D'^p)]^{1/p}.$$

Now each term in this inequality is a differentiable function of p , and differentiation commutes with m because, for any fixed self-adjoint operator T and general function f , $E[f(T)]$ may be represented as a conventional integral of f as a function of a real variable, with respect to a certain regular measure on the reals (Cor. 12.6 of [5]), so that the familiar conditions for the validity of differentiation under an integral sign are applicable. Now noting that both sides of the preceding inequality take the value unity when $p = 1$, it follows that the respective derivatives, evaluated at the point $p = 1$, satisfy a similar inequality. Performing in a straightforward manner this differentiation and substitution, and noting that

$$e(E) = -\frac{d}{dp} E(D^p) \Big|_{p=1}$$

(a representation of the entropy similar to that used for probability distributions on the line in [1]), the stated inequality follows.

Theorem 2. *The entropy of the restriction of a bounded state to a subring is not less than the entropy of the original state.*

In order for the entropy of the restriction of a state E to a subring $\mathcal{B}$ of $\mathcal{A}$ to be defined, relative to a gage m on $\mathcal{A}$, it is necessary that the restriction of m to $\mathcal{B}$ define a gage on $\mathcal{B}$, which amounts to the requirement that every projection in $\mathcal{B}$ bound a projection in $\mathcal{A}$ on which m is finite. With this assumption, there is for any integrable X in $\mathcal{A}$ a conditional expectation X' relative to $\mathcal{B}$ uniquely determined by the condition that $m(XY) = m(X'Y)$ for all $Y \in \mathcal{B}$. It follows readily that $|X'|_1 \leq |X|_1$ and $|X'|_2 \leq |X|_2$, and so by the extension to gage spaces due to KUNZE [2] of a well known convexity theorem of CALDERON and ZYGMUND it follows that $|X'|_p \leq |X|_p$. In particular, if D' is the conditional expectation of D relative to $\mathcal{B}$, then $m(D'^p) \leq m(D^p)$. Differentiation of both

² See Comment (1) at the end of this note.

sides of this inequality with respect to p and substitution of the value $p = 1$ (as is justified since both sides have the value unity for $p = 1$) gives, as in the proof of Theorem 2, the inequality

$$m(D' \log D') \leq m(D \log D),$$

which is equivalent to the stated relation between the respective entropies.

Entropy of pure and equilibrium states. A pure state of $\mathfrak{A}$ is one that is not a mixture of two other states, while an equilibrium state may be defined (relative to a given gage m) as one that is invariant under (the induced action of) all m -preserving automorphisms of $\mathfrak{A}$. If this terminology is to be in essential accord with the usage of the same terms in physics, then the pure and equilibrium states should more or less coincide with the states of minimal and maximal entropy, respectively. On the other hand, states of these various types do not always exist (while pure states exist always, there may be none of the “regular” type needed for strict physical realizability), so no precise general statement is possible. The next theorem collects several results in this direction for special cases of interest, some of which are essentially contained in [7] and [8].

Theorem 3. *States of minimal entropy exist, and are characterized as the pure regular states, in the following cases:*

*the ring of all bounded linear operators on a Hilbert space;
a discrete measure space with all points equally weighted.*

An equilibrium state exists, and is characterized as a state of maximal entropy, in the following cases:

*factors of type II_1 or I_n ;
a finite measure space.*

For succinctness of statement, a methodical specification of the ring $\mathfrak{A}$ and the gage m has not been included in the enunciation of the theorem. These are simply the obvious ones, but, to be quite explicit, they are as follows. In the case of the ring of all bounded operators there is a unique gage (within a positive proportionality factor), whose value on a projection is proportional to the dimension of its range. In the case of a discrete measure space, $\mathfrak{A}$ consists of all bounded functions on the space, and the gage of a projection in $\mathfrak{A}$ is the number of elements in the set whose characteristic function gives the projection. In the case of a factor of finite type, $\mathfrak{A}$ is the factor in question, and m is the unique gage on $\mathfrak{A}$ (within a positive proportionality factor). For a finite measure space, $\mathfrak{A}$ is the ring of all bounded measurable functions, modulo null functions, while $m(P)$ is the measure of the set whose characteristic function gives P .

The theorem follows readily from the known forms for the regular states, together with, in the cases other than that of all bounded operators, a well known inequality based on the convexity of $t \log t$ (see *e.g. Inequalities*, by HARDY, LITTLEWOOD & POLYA, Cambridge, 1952) and also Cor. 12.6 of [5]. It is not diffi-

cult to verify that the equilibrium states mentioned are unique, provided, in the measure space case, that the space is homogeneous (cf. D. MAHARAM, *On homogeneous measure algebras*, Proc. Nat. Acad. Sci. **28**(1942), 313–326).

The probabilistic case. This is effectively coextensive with the case when $\mathfrak{A}$ is abelian. The approximation of arbitrary regular states by bounded ones can be carried out along completely standard lines, and Th. 1 is stated more exactly as follows for this case.

Corollary 1. *If p and p' are probability measures on the Borel field $\mathfrak{A}$ of subsets of the space R , with the entropy relative to the localizable measure r on R of such a measure defined as*

$$e(p) = - \int_R \log \frac{dp}{dr} dp,$$

and if $0 < \alpha < 1$ and $\alpha + \alpha' = 1$, then

$$e(\alpha p + \alpha' p') \geq \alpha e(p) + \alpha' e(p').$$

Here $\mathfrak{A}$ is the algebra of all bounded measurable functions (strictly speaking, of all multiplications by such as operators on $L_2(R, \mathfrak{A}, r)$), while the functional E corresponding to p is the corresponding expectation functional, $E(k) = \int_R k dp$. Localizability is an assumption required in case R is not a countable union of sets of finite measure to assure the existence of dp/dr —cf. [6]; it is also assumed, to avoid presently irrelevant pathology, that each set in $\mathfrak{A}$ of positive measure contains a subset of positive finite measure.

The entropies involved in the inequality of the corollary may be infinite or even not well defined, but in the latter case there is naturally no content.

In case R is a purely continuous space (no atoms), there are no pure regular states. If $r(R) = \infty$ there is also no equilibrium state, and the entropy may have any value from $-\infty$ to $+\infty$, inclusive, when well defined.

The quantum-mechanical case. In the case of a particle system with a finite number of degrees of freedom, it is generally assumed, and seems to create no foundational difficulties as long as the states are restricted to be regular, that the bounded observables of the system are mathematically representable by the bounded self-adjoint operator on a Hilbert space. Non-normalizable states are important from an analytical-theoretical point of view, but do not represent physically realizable particle states, having only slightly ambiguous interpretations as “beams” of particles. A “packet” of such states can represent a *bona fide* physical state, but this will be normalizable, and so determine a regular state. *E.g.* a plane wave e^{ikx} is such a non-normalizable state for a free particle, which is more convenient analytically than the normalizable packet $\int e^{ikx} f(k) dk$ ($f \in L_2$), but which from a foundational point of view is a less satisfactory notion and can lead to inconsistencies. It is in this sense that the restriction to the

consideration of regular states is appropriate in the quantum mechanics of finite systems.

On the other hand, a finite quantum-mechanical system is inevitably insecure from a foundational viewpoint since the phenomena being treated generally have no complete theoretical explanation except in terms of an infinite system. A Bose-Einstein quantum field is shown in [4] to be appropriately representable by a limit in the sense of uniform convergence of factors each of which is of type I_∞ . A rather conservative physical requirement leads to a class of states which turn out to be precisely those whose restrictions to the approximating type I_∞ factors are regular in the foregoing sense. This makes it possible to define the entropy of such a state as the limit of the entropies as defined above for the restrictions, when the limit exists.

Comments on further problems. (1) Theorem 1 is stated without proof in [8] for the case of the ring of all bounded operators, in the stronger form

$$\operatorname{tr} F(\alpha A + \alpha' A') \leq \alpha \operatorname{tr} F(A) + \alpha' \operatorname{tr} F(A')$$

for an essentially arbitrary convex function F . No simple direct proof is known to us, but the result follows readily from an inequality of WEYL concerning the relation between the proper values of $A + B$ on the one hand and those of A and B on the other, for self-adjoint operators A and B (Theorem 1 of *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen*, Math. Ann. **71**(1912), 441–479). There is every reason to expect that the corresponding inequality holds in an arbitrary ring of operators, in terms of the relative trace, and such a generalized inequality would imply the present Theorem 1 in a correspondingly strengthened form.

(2) Another direction of strengthening for Theorem 1 is towards removal of the restriction to bounded states. Presumably the states of finite entropy form a convex set on which the entropy is *strictly* concave; but in the case of a non-abelian ring some further technical elaboration seems needed for dealing readily with such questions.

(3) Similarly, Theorem 2 should hold for any state of well-defined entropy, and the inequality may well be strict in the sense that the equality holds only when the restricted state has a *unique* regular extension—at least in the case of the rings of major interest.

(4) A general theory of the entropy of one state relative to another state, rather than relative to a gage, does not seem possible—in particular, Theorems 1 and 2 would appear to fail—but a treatment of the possible existence of such an entropy function as defined by its properties (carried out for the case of a finite discrete measure space in [7], Appendix 2) would be more conclusive.

(5) In the case of a Bose-Einstein field, it would be of interest to determine whether every regular state has a well defined entropy, and whether the pure states are those of vanishing entropy.

Acknowledgement. The author is indebted to JOHN WARD for a stimulating discussion of entropy some time ago.

REFERENCES

- [1] I. I. HIRSCHMAN, JR., *A note on entropy*. Amer. Jour. Math. **79** (1957), 152–156.
- [2] R. A. KUNZE, L_p *Fourier transforms on locally compact unimodular groups*. Trans. Amer. Math. Soc. **89** (1958), 519–540.
- [3] I. E. SEGAL, *Postulates for general quantum mechanics*. Annals of Math. **48** (1947), 930–948.
- [4] ———, *Foundations of the theory of dynamical systems of infinitely many degrees of freedom*. I. Mat.-fys. Medd., K. Danske Vidensk. Selsk. **31**, nr. 12 (1959), 1–39.
- [5] ———, *A non-commutative extension of abstract integration*. Annals of Math. **57** (1953), 401–457.
- [6] ———, *Equivalences of measure spaces*. Amer. Jour. Math. **73** (1951), 275–313.
- [7] C. E. SHANNON, *A mathematical theory of communication*. Bell Syst. Tech. Jour. **27** (1948), 379–423 and 623–656.
- [8] J. VON NEUMANN, *Mathematische Grundlagen der Quantenmechanik*. Berlin, 1932.

University of Chicago
Chicago, Illinois