The Many-Worlds Interpretation of Quantum Mechanics

THE THEORY OF THE UNIVERSAL WAVE FUNCTION

Hugh Everett, III

I. INTRODUCTION

We begin, as a way of entering our subject, by characterizing a particular interpretation of quantum theory which, although not representative of the more careful formulations of some writers, is the most common form encountered in textbooks and university lectures on the subject.

A physical system is described completely by a state function ψ, which is an element of a Hilbert space, and which furthermore gives information only concerning the probabilities of the results of various observations which can be made on the system. The state function ψ is thought of as objectively characterizing the physical system, i.e., at all times an isolated system is thought of as possessing a state function, independently of our state of knowledge of it. On the other hand, ψ changes in a causal manner so long as the system remains isolated, obeying a differential equation. Thus there are two fundamentally different ways in which the state function can change:[1]

Process 1: The discontinuous change brought about by the observation of a quantity with eigenstates $\phi_1, \phi_2, \ldots$, in which the state ψ will be changed to the state ϕ_j with probability $|(\psi, \phi_j)|^2$.

Process 2: The continuous, deterministic change of state of the (isolated) system with time according to a wave equation $\frac{\partial \psi}{\partial t} = U\psi$, where U is a linear operator.

---

1 We use here the terminology of von Neumann [17].

The question of the consistency of the scheme arises if one contemplates regarding the observer and his object-system as a single (composite) physical system. Indeed, the situation becomes quite paradoxical if we allow for the existence of more than one observer. Let us consider the case of one observer A, who is performing measurements upon a system S, the totality (A + S) in turn forming the object-system for another observer, B.

If we are to deny the possibility of B's use of a quantum mechanical description (wave function obeying wave equation) for A + S, then we must be supplied with some alternative description for systems which contain observers (or measuring apparatus). Furthermore, we would have to have a criterion for telling precisely what type of systems would have the preferred positions of "measuring apparatus" or "observer" and be subject to the alternate description. Such a criterion is probably not capable of rigorous formulation.

On the other hand, if we do allow B to give a quantum description to A + S, by assigning a state function ψ^{A+S}, then, so long as B does not interact with A + S, its state changes causally according to Process 2, *even though* A *may be performing measurements upon* S. From B's point of view, nothing resembling Process 1 can occur (there are no discontinuities), and the question of the validity of A's use of Process 1 is raised. That is, *apparently* either A is incorrect in assuming Process 1, with its probabilistic implications, to apply to his measurements, or else B's state function, with its purely causal character, is an inadequate description of what is happening to A + S.

To better illustrate the paradoxes which can arise from strict adherence to this interpretation we consider the following amusing, but *extremely hypothetical* drama.

> Isolated somewhere out in space is a room containing an observer, A, who is about to perform a measurement upon a system S. After performing his measurement he will record the result in his notebook. We assume that he knows the state function of S (perhaps as a result

of previous measurement), and that it is not an eigenstate of the measurement he is about to perform. A, being an orthodox quantum theorist, then believes that the outcome of his measurement is undetermined and that the process is correctly described by Process 1.

In the meantime, however, there is another observer, B, outside the room, who is in possession of the state function of the entire room, including S, the measuring apparatus, and A, just prior to the measurement. B is only interested in what will be found in the notebook one week hence, so he computes the state function of the room for one week in the future according to Process 2. One week passes, and we find B still in possession of the state function of the room, which this equally orthodox quantum theorist believes to be a complete description of the room and its contents. If B's state function calculation tells beforehand exactly what is going to be in the notebook, then A is incorrect in his belief about the indeterminacy of the outcome of his measurement. We therefore assume that B's state function contains non-zero amplitudes over several of the notebook entries.

At this point, B opens the door to the room and looks at the notebook (performs his observation). Having observed the notebook entry, he turns to A and informs him in a patronizing manner that since his (B's) wave function just prior to his entry into the room, which he knows to have been a complete description of the room and its contents, had non-zero amplitude over other than the present result of the measurement, the result must have been decided only when B entered the room, so that A, his notebook entry, and his memory about what occurred one week ago had no independent objective existence until the intervention by B. In short, B implies that A owes his present objective existence to B's generous nature which compelled him to intervene on his behalf. However, to B's consternation, A does not react with anything like the respect and gratitude he should exhibit towards B, and at the end of a somewhat heated reply, in which A conveys in a colorful manner his opinion of B and his beliefs, he

rudely punctures B's ego by observing that if B's view is correct, then he has no reason to feel complacent, since the whole present situation may have no objective existence, but may depend upon the future actions of yet another observer.

It is now clear that the interpretation of quantum mechanics with which we began is untenable if we are to consider a universe containing more than one observer. We must therefore seek a suitable modification of this scheme, or an entirely different system of interpretation. Several alternatives which avoid the paradox are:

Alternative 1: To postulate the existence of only one observer in the universe. This is the solipsist position, in which each of us must hold the view that he alone is the only valid observer, with the rest of the universe and its inhabitants obeying at all times Process 2 except when under his observation.

This view is quite consistent, but one must feel uneasy when, for example, writing textbooks on quantum mechanics, describing Process 1, for the consumption of other persons to whom it does not apply.

Alternative 2: To limit the applicability of quantum mechanics by asserting that the quantum mechanical description fails when applied to observers, or to measuring apparatus, or more generally to systems approaching macroscopic size.

If we try to limit the applicability so as to exclude measuring apparatus, or in general systems of macroscopic size, we are faced with the difficulty of sharply defining the region of validity. For what n might a group of n particles be construed as forming a measuring device so that the quantum description fails? And to draw the line at human or animal observers, i.e., to assume that all mechanical aparata obey the usual laws, but that they are somehow not valid for living observers, does violence to the so-called

principle of psycho-physical parallelism,[2] and constitutes a view to be avoided, if possible. To do justice to this principle we must insist that we be able to conceive of mechanical devices (such as servomechanisms), obeying natural laws, which we would be willing to call observers.

Alternative 3: To admit the validity of the state function description, but to deny the possibility that B could ever be in possession of the state function of A + S. Thus one might argue that a determination of the state of A would constitute such a drastic intervention that A would cease to function as an observer.

The first objection to this view is that no matter what the state of A + S is, there is in principle a complete set of commuting operators for which it is an eigenstate, so that, at least, the determination of *these* quantities will not affect the state nor in any way disrupt the operation of A. There are no fundamental restrictions in the usual theory about the knowability of *any* state functions, and the introduction of any such restrictions to avoid the paradox must therefore require extra postulates.

The second objection is that it is not particularly relevant whether or not B actually *knows* the precise state function of A + S. If he merely *believes* that the system is described by a state function, which he does not presume to know, then the difficulty still exists. He must then believe that this state function changed deterministically, and hence that there was nothing probabilistic in A's determination.

---

[2] In the words of von Neumann ([17], p. 418): "...it is a fundamental requirement of the scientific viewpoint – the so-called principle of the psycho-physical parallelism – that it must be possible so to describe the extra-physical process of the subjective perception as if it were in reality in the physical world – *i.e.*, to assign to its parts equivalent physical processes in the objective environment, in ordinary space."

Alternative 4: To abandon the position that the state function is a *complete* description of a system. The state function is to be regarded not as a description of a single system, but of an ensemble of systems, so that the probabilistic assertions arise naturally from the incompleteness of the description.

It is assumed that the correct complete description, which would presumably involve further (hidden) parameters beyond the state function alone, would lead to a deterministic theory, from which the probabilistic aspects arise as a result of our ignorance of these extra parameters in the same manner as in classical statistical mechanics.

Alternative 5: To assume the universal validity of the quantum description, by the complete abandonment of Process 1. The general validity of pure wave mechanics, *without any statistical assertions*, is assumed for *all* physical systems, including observers and measuring apparata. Observation processes are to be described completely by the state function of the composite system which includes the observer and his object-system, and which at all times obeys the wave equation (Process 2).

This brief list of alternatives is not meant to be exhaustive, but has been presented in the spirit of a preliminary orientation. We have, in fact, omitted one of the foremost interpretations of quantum theory, namely the position of Niels Bohr. The discussion will be resumed in the final chapter, when we shall be in a position to give a more adequate appraisal of the various alternate interpretations. For the present, however, we shall concern ourselves only with the development of Alternative 5.

It is evident that Alternative 5 is a theory of many advantages. It has the virtue of logical simplicity and it is complete in the sense that it is applicable to the entire universe. All processes are considered equally (there are no "measurement processes" which play any preferred role), and the principle of psycho-physical parallelism is fully maintained. Since

the universal validity of the state function description is asserted, one can regard the state functions themselves as the fundamental entities, and one can even consider the state function of the whole universe. In this sense this theory can be called the theory of the "universal wave function," since all of physics is presumed to follow from this function alone. There remains, however, the question whether or not such a theory can be put into correspondence with our experience.

The present thesis is devoted to showing that this concept of a universal wave mechanics, together with the necessary correlation machinery for its interpretation, forms a logically self consistent description of a universe in which several observers are at work.

We shall be able to introduce into the theory systems which represent observers. Such systems can be conceived as automatically functioning machines (servomechanisms) possessing recording devices (memory) and which are capable of responding to their environment. The behavior of these observers shall always be treated within the framework of wave mechanics. Furthermore, we shall deduce the probabilistic assertions of Process 1 as *subjective* appearances to such observers, thus placing the theory in correspondence with experience. We are then led to the novel situation in which the formal theory is objectively continuous and causal, while subjectively discontinuous and probabilistic. While this point of view thus shall ultimately justify our use of the statistical assertions of the orthodox view, it enables us to do so in a logically consistent manner, allowing for the existence of other observers. At the same time it gives a deeper insight into the meaning of quantized systems, and the role played by quantum mechanical correlations.

In order to bring about this correspondence with experience for the pure wave mechanical theory, we shall exploit the correlation between subsystems of a composite system which is described by a state function. A subsystem of such a composite system does not, in general, possess an independent state function. That is, in general a composite system cannot be represented by a single pair of subsystem states, but can be repre-

sented only by a *superposition* of such pairs of subsystem states. For example, the Schrodinger wave function for a pair of particles, $\psi(x_1, x_2)$, cannot always be written in the form $\psi = \phi(x_1)\eta(x_2)$, but only in the form $\psi = \sum_{i,j} a_{ij}\phi^i(x_1)\eta^j(x_2)$. In the latter case, there is no single state for Particle 1 alone or Particle 2 alone, but only the superposition of such cases.

In fact, to any arbitrary choice of state for one subsystem there will correspond a *relative state* for the other subsystem, which will generally be dependent upon the choice of state for the first subsystem, so that the state of one subsystem is not independent, but correlated to the state of the remaining subsystem. Such correlations between systems arise from interaction of the systems, and from our point of view all measurement and observation processes are to be regarded simply as interactions between observer and object-system which produce strong correlations.

Let one regard an observer as a subsystem of the composite system: observer + object-system. It is then an inescapable consequence that after the interaction has taken place there will not, generally, exist a single observer state. There will, however, be a superposition of the composite system states, each element of which contains a definite observer state and a definite relative object-system state. Furthermore, as we shall see, *each* of these relative object-system states will be, approximately, the eigenstates of the observation corresponding to the value obtained by the observer which is described by the same element of the superposition. Thus, each element of the resulting superposition describes an observer who perceived a definite and generally different result, and to whom it appears that the object-system state has been transformed into the corresponding eigenstate. In this sense the usual assertions of Process 1 appear to hold on a subjective level to each observer described by an element of the superposition. We shall also see that correlation plays an important role in preserving consistency when several observers are present and allowed to interact with one another (to "consult" one another) as well as with other object-systems.

In order to develop a language for interpreting our pure wave mechanics for composite systems we shall find it useful to develop quantitative definitions for such notions as the "sharpness" or "definiteness" of an operator A for a state ψ, and the "degree of correlation" between the subsystems of a composite system or between a pair of operators in the subsystems, so that we can use these concepts in an unambiguous manner. The mathematical development of these notions will be carried out in the next chapter (II) using some concepts borrowed from Information Theory.[3] We shall develop there the general definitions of information and correlation, as well as some of their more important properties. Throughout Chapter II we shall use the language of probability theory to facilitate the exposition, and because it enables us to introduce in a unified manner a number of concepts that will be of later use. We shall nevertheless subsequently apply the mathematical definitions directly to state functions, by replacing probabilities by square amplitudes, *without*, however, *making any reference to probability models.*

Having set the stage, so to speak, with Chapter II, we turn to quantum mechanics in Chapter III. There we first investigate the quantum formalism of composite systems, particularly the concept of relative state functions, and the meaning of the representation of subsystems by non-interfering mixtures of states characterized by density matrices. The notions of information and correlation are then applied to quantum mechanics. The final section of this chapter discusses the measurement process, which is regarded simply as a correlation-inducing interaction between subsystems of a single isolated system. A simple example of such a measurement is given and discussed, and some general consequences of the superposition principle are considered.

---

3 The theory originated by Claude E. Shannon [19].

This will be followed by an abstract treatment of the problem of Observation (Chapter IV). In this chapter we make use only of the superposition principle, and general rules by which composite system states are formed of subsystem states, in order that our results shall have the greatest generality and be applicable to any form of quantum theory for which these principles hold. (Elsewhere, when giving examples, we restrict ourselves to the non-relativistic Schrödinger Theory for simplicity.) The validity of Process 1 as a subjective phenomenon is deduced, as well as the consistency of allowing several observers to interact with one another.

Chapter V supplements the abstract treatment of Chapter IV by discussing a number of diverse topics from the point of view of the theory of pure wave mechanics, including the existence and meaning of macroscopic objects in the light of their atomic constitution, amplification processes in measurement, questions of reversibility and irreversibility, and approximate measurement.

The final chapter summarizes the situation, and continues the discussion of alternate interpretations of quantum mechanics.

II. PROBABILITY, INFORMATION, AND CORRELATION

The present chapter is devoted to the mathematical development of the concepts of information and correlation. As mentioned in the introduction we shall use the language of probability theory throughout this chapter to facilitate the exposition, although we shall apply the mathematical definitions and formulas in later chapters without reference to probability models. We shall develop our definitions and theorems in full generality, for probability distributions over arbitrary sets, rather than merely for distributions over real numbers, with which we are mainly interested at present. We take this course because it is as easy as the restricted development, and because it gives a better insight into the subject.

The first three sections develop definitions and properties of information and correlation for probability distributions over *finite* sets only. In section four the definition of correlation is extended to distributions over arbitrary sets, and the general invariance of the correlation is proved. Section five then generalizes the definition of information to distributions over arbitrary sets. Finally, as illustrative examples, sections seven and eight give brief applications to stochastic processes and classical mechanics, respectively.

§1. *Finite joint distributions*

We assume that we have a collection of finite sets, $\mathcal{X},\mathcal{Y},\ldots,\mathcal{Z}$, whose elements are denoted by $x_i \in \mathcal{X}$, $y_j \in \mathcal{Y},\ldots$, $z_k \in \mathcal{Z}$, etc., and that we have a *joint probability distribution*, $P = P(x_i,y_j,\ldots,z_k)$, defined on the cartesian product of the sets, which represents the probability of the combined event $x_i,y_j,\ldots,$ and z_k. We then denote by $X,Y,\ldots,Z$ the random variables whose values are the elements of the sets $\mathcal{X},\mathcal{Y},\ldots,\mathcal{Z}$, with probabilities given by P.

For any subset $Y,\ldots,Z$, of a set of random variables $W,\ldots,X, Y,\ldots,Z$, with joint probability distribution $P(w_i,\ldots,x_j,y_k,\ldots,z_\ell)$, the *marginal distribution*, $P(y_k,\ldots,z_\ell)$, is defined to be:

$$(1.1) \qquad P(y_k,\ldots,z_\ell) = \sum_{i,\ldots,j} P(w_i,\ldots,x_j,y_k,\ldots,z_\ell) \ ,$$

which represents the probability of the joint occurrence of $y_k,\ldots,z_\ell$, with no restrictions upon the remaining variables.

For any subset $Y,\ldots,Z$ of a set of random variables the *conditional distribution*, conditioned upon the values $W = w_i,\ldots,X = x_j$ for any remaining subset $W,\ldots,X$, and denoted by $P^{w_i,\ldots,x_j}(y_k,\ldots,z_\ell)$, is defined to be:[1]

$$(1.2) \qquad P^{w_i,\ldots,x_j}(y_k,\ldots,z_\ell) = \frac{P(w_i,\ldots,x_j,y_k,\ldots,z_\ell)}{P(w_i,\ldots,x_j)} \ ,$$

which represents the probability of the joint event $Y = y_k,\ldots,Z = z_\ell$, conditioned by the fact that $W,\ldots,X$ are known to have taken the values $w_i,\ldots,x_j$, respectively.

For any numerical valued function $F(y_k,\ldots,z_\ell)$, defined on the elements of the cartesian product of $\mathcal{Y},\ldots,\mathcal{Z}$, the *expectation*, denoted by Exp [F], is defined to be:

$$(1.3) \qquad \text{Exp [F]} = \sum_{k,\ldots,\ell} P(y_k,\ldots,z_\ell)\, F(y_k,\ldots,z_\ell) \ .$$

We note that if $P(y_k,\ldots,z_\ell)$ is a marginal distribution of some larger distribution $P(w_i,\ldots,x_j,y_k,\ldots,z_\ell)$ then

$$(1.4) \qquad \text{Exp [F]} = \sum_{k,\ldots,\ell} \left(\sum_{\ell,\ldots,j} P(w_i,\ldots,x_j,y_k,\ldots,z_\ell) \right) F(y_k,\ldots,z_\ell)$$

$$= \sum_{i,\ldots,j,k,\ldots,\ell} P(w_i,\ldots,x_j,y_k,\ldots,z_\ell)\, F(y_k,\ldots,z_\ell) \ ,$$

---

[1] We regard it as undefined if $P(w_i,\ldots,x_j) = 0$. In this case $P(w_i,\ldots,x_j, y_k,\ldots,z_\ell)$ is necessarily zero also.

so that if we wish to compute Exp [F] with respect to some joint distribution it suffices to use *any* marginal distribution of the original distribution which contains at least those variables which occur in F.

We shall also occasionally be interested in *conditional expectations*, which we define as:

$$\text{(1.5)} \qquad \text{Exp}^{w_i,\ldots,x_j}[F] = \sum_{k,\ldots,\ell} P^{w_i,\ldots,x_j}(y_k,\ldots,z_\ell)\, F(y_k,\ldots,z_\ell) \ ,$$

and we note the following easily verified rules for expectations:

$$\text{(1.6)} \qquad \text{Exp}[\text{Exp}[F]] = \text{Exp}[F] \ ,$$

$$\text{(1.7)} \qquad \text{Exp}^{u_i,\ldots,v_j}[\text{Exp}^{u_i,\ldots,v_j,\, w_k,\ldots,x_\ell}[F]] = \text{Exp}^{u_i,\ldots,v_j}[F] \ ,$$

$$\text{(1.8)} \qquad \text{Exp}[F+G] = \text{Exp}[F] + \text{Exp}[G] \ .$$

We should like finally to comment upon the notion of *independence*. Two random variables X and Y with joint distribution $P(x_i, y_j)$ will be said to be independent if and only if $P(x_i, y_j)$ is equal to $P(x_i)P(y_j)$ for all i, j. Similarly, the groups of random variables (U...V), (W...X),..., (Y...Z) will be called *mutually independent groups* if and only if $P(u_i,\ldots,v_j, w_k,\ldots,x_\ell,\ldots,y_m,\ldots,z_n)$ is always equal to $P(u_i,\ldots,v_j)$ $P(w_k,\ldots,x_\ell)\ldots P(y_m,\ldots,z_n)$.

Independence means that the random variables take on values which are not influenced by the values of other variables with respect to which they are independent. That is, the conditional distribution of one of two independent variables, Y, conditioned upon the value x_i for the other, is independent of x_i, so that knowledge about one variable tells nothing of the other.

§2. *Information for finite distributions*

Suppose that we have a single random variable X, with distribution $P(x_i)$. We then define[2] a number, I_X, called the *information* of X, to be:

---

[2] This definition corresponds to the negative of the *entropy* of a probability distribution as defined by Shannon [19].

$$(2.1) \qquad I_X = \sum_i P(x_i) \ln P(x_i) = \text{Exp}\,[\ln P(x_i)] \ ,$$

which is a function of the probabilities alone and not of any possible numerical values of the x_i's themselves.[3]

The information is essentially a measure of the sharpness of a probability distribution, that is, an inverse measure of its "spread." In this respect information plays a role similar to that of variance. However, it has a number of properties which make it a superior measure of the "sharpness" than the variance, not the least of which is the fact that it can be defined for distributions over arbitrary sets, while variance is defined only for distributions over real numbers.

Any change in the distribution $P(x_i)$ which "levels out" the probabilities decreases the information. It has the value zero for "perfectly sharp" distributions, in which the probability is one for one of the x_i and zero for all others, and ranges downward to $-\ln n$ for distributions over n elements which are equal over all of the x_i. The fact that the information is nonpositive is no liability, since we are seldom interested in the absolute information of a distribution, but only in differences.

We can generalize (2.1) to obtain the formula for the information of a *group* of random variables $X, Y, \ldots, Z$, with joint distribution $P(x_i, y_j, \ldots, z_k)$, which we denote by $I_{XY\ldots Z}$:

$$(2.2) \qquad I_{XY\ldots Z} = \sum_{i,j,\ldots,k} P(x_i, y_j, \ldots, z_k) \ln P(x_i, y_j, \ldots, z_k)$$

$$= \text{Exp}\,[\ln P(x_i, y_j, \ldots, z_k)] \ ,$$

---

[3] A good discussion of information is to be found in Shannon [19], or Woodward [21]. Note, however, that in the theory of communication one defines the information of a *state* x_i, which has *a priori* probability P_i, to be $-\ln P_i$. We prefer, however, to regard information as a property of the distribution itself.

which follows immediately from our previous definition, since the group of random variables $X, Y, \ldots, Z$ may be regarded as a single random variable W which takes its values in the cartesian product $\mathfrak{X} \times \mathfrak{Y} \times \cdots \times \mathfrak{Z}$.

Finally, we define a *conditional information*, $I_{XY\ldots Z}^{v_m,\ldots,w_n}$, to be:

$$(2.3)\quad I_{XY\ldots Z}^{v_m,\ldots,w_n} = \sum_{i,j,\ldots,k} P^{v_m,\ldots,w_n}(x_i, y_j, \ldots, z_k) \ln P^{v_m,\ldots,w_n}(x_i, y_j, \ldots, z_k)$$
$$= \mathrm{Exp}^{v_m,\ldots,w_n} [\ln P^{v_m,\ldots,w_n}(x_i, y_j, \ldots, z_k)] ,$$

a quantity which measures our information about $X, Y, \ldots, Z$ given that we know that $V \ldots W$ have taken the particular values $v_m, \ldots, w_n$.

For independent random variables $X, Y, \ldots, Z$, the following relationship is easily proved:

$$(2.4)\qquad I_{XY\ldots Z} = I_X + I_Y + \ldots + I_Z \quad (X, Y, \ldots, Z \text{ independent}) ,$$

so that the information of $XY \ldots Z$ is the sum of the individual quantities of information, which is in accord with our intuitive feeling that if we are given information about unrelated events, our total knowledge is the sum of the separate amounts of information. We shall generalize this definition later, in §5.

§3. *Correlation for finite distributions*

Suppose that we have a pair of random variables, X and Y, with joint distribution $P(x_i, y_j)$. If we say that X and Y are *correlated*, what we intuitively mean is that *one learns something about one variable when he is told the value of the other*. Let us focus our attention upon the variable X. If we are not informed of the value of Y, then our information concerning X, I_X, is calculated from the marginal distribution $P(x_i)$. However, if we are now told that Y has the value y_j, then our information about X changes to the information of the conditional distribution $P^{y_j}(x_i)$, $I_X^{y_j}$. According to what we have said, we wish the degree correlation to measure how much we learn about X by being informed of

Y's value. However, since the change of information, $I_X^{y_j} - I_X$, may depend upon the particular value, y_j, of Y which we are told, the natural thing to do to arrive at a single number to measure the strength of correlation is to consider the *expected* change in information about X, given that we are to be told the value of Y. This quantity we call the *correlation information*, or for brevity, the *correlation*, of X and Y, and denote it by $\{X, Y\}$. Thus:

$$\{X, Y\} = \text{Exp}\left[I_X^{y_j} - I_X\right] = \text{Exp}\left[I_X^{y_j}\right] - I_X \,. \tag{3.1}$$

Expanding the quantity $\text{Exp}\left[I_X^{y_j}\right]$ using (2.3) and the rules for expectations (1.6)–(1.8) we find:

$$\begin{aligned} \text{Exp}\left[I_X^{y_j}\right] &= \text{Exp}\left[\text{Exp}^{y_j}\left[\ln P^{y_j}(x_i)\right]\right] \\ &= \text{Exp}\left[\ln \frac{P(x_i, y_j)}{P(y_j)}\right] = \text{Exp}\left[\ln P(x_i, y_j)\right] - \text{Exp}\left[\ln P(y_j)\right] \\ &= I_{XY} - I_Y \,, \end{aligned} \tag{3.2}$$

and combining with (3.1) we have:

$$\{X, Y\} = I_{XY} - I_X - I_Y \,. \tag{3.3}$$

Thus the correlation is symmetric between X and Y, and hence also equal to the expected change of information about Y given that we will be told the value of X. Furthermore, according to (3.3) the correlation corresponds precisely to the amount of "missing information" if we possess only the marginal distributions, i.e., the loss of information if we choose to regard the variables as independent.

THEOREM 1. $\{X, Y\} = 0$ *if and only if* X *and* Y *are independent, and is otherwise strictly positive.* (Proof in Appendix I.)

In this respect the correlation so defined is superior to the usual correlation coefficients of statistics, such as covariance, etc., which can be zero even when the variables are not independent, and which can assume both positive and negative values. An inverse correlation is, after all, quite as useful as a direct correlation. Furthermore, it has the great advantage of depending upon the probabilities alone, and not upon any numerical values of x_i and y_j, so that it is defined for distributions over sets whose elements are of an arbitrary nature, and not only for distributions over numerical properties. For example, we might have a joint probability distribution for the political party and religious affiliation of individuals. Correlation and information are defined for such distributions, although they possess nothing like covariance or variance.

We can generalize (3.3) to define a *group correlation* for the groups of random variables (U...V), (W...X),..., (Y...Z), denoted by {U...V, W...X, ..., Y...Z} (where the groups are separated by commas), to be:

$$\{U...V, W...X, ..., Y...Z\} = I_{U...V\,W...X...Y...Z} - I_{U...V} - I_{W...X} - \cdots - I_{Y...Z}\,, \tag{3.4}$$

again measuring the information deficiency for the group marginals. Theorem 1 is also satisfied by the group correlation, so that it is zero if and only if the groups are mutually independent. We can, of course, also define conditional correlations in the obvious manner, denoting these quantities by appending the conditional values as superscripts, as before.

We conclude this section by listing some useful formulas and inequalities which are easily proved:

$$\{U, V, ..., W\} = \mathrm{Exp}\left[\ln \frac{P(u_i, v_j, ..., w_k)}{P(u_i)P(v_j)...P(w_k)}\right], \tag{3.5}$$

$$\{U,V,...,W\}^{x_i...y_j} = \mathrm{Exp}^{x_i...y_j}\left[\ln \frac{P^{x_i...y_j}(u_k, v_1, ..., w_m)}{P^{x_i...y_j}(u_k)P^{x_i...y_j}(v_1)...P^{x_i...y_j}(w_m)}\right] \tag{3.6}$$

(conditional correlation) ,

$$\{\ldots,U,V,\ldots\} = \{\ldots,UV,\ldots\} + \{U,V\} \ , \tag{3.7}$$
$$\{\ldots,U,V,\ldots,W,\ldots\} = \{\ldots,UV\ldots W,\ldots\} + \{U,V,\ldots,W\} \text{ (comma removal)}$$

$$\{\ldots,U,VW,\ldots\} - \{\ldots,UV,W,\ldots\} = \{U,V\} - \{V,W\} \text{ (commutator)} \ , \tag{3.8}$$

$$\{X\} = 0 \quad \text{(definition of bracket with no commas)} \ , \tag{3.9}$$

$$\{\ldots,XXV,\ldots\} = \{\ldots,XV,\ldots\} \tag{3.10}$$
(removal of repeated variable within a group) ,

$$\{\ldots,UV,VW,\ldots\} = \{\ldots,UV,W,\ldots\} - \{V,W\} - I_V \tag{3.11}$$
(removal of repeated variable in separate groups) ,

$$\{X,X\} = -I_X \quad \text{(self correlation)} \ , \tag{3.12}$$

$$\{U,VW,X\}^{\ldots w_j \ldots} = \{U,V,X\}^{\ldots w_j \ldots} \ , \tag{3.13}$$
$$\{U,W,X\}^{\ldots w_j \ldots} = \{U,X\}^{\ldots w_j \ldots}$$
(removal of conditioned variables) ,

$$\{XY,Z\} \geqq \{X,Z\} \ , \tag{3.14}$$

$$\{XY,Z\} \geqq \{X,Z\} + \{Y,Z\} - \{X,Y\} \ , \tag{3.15}$$

$$\{X,Y,Z\} \geqq \{X,Y\} + \{X,Z\} \ . \tag{3.16}$$

Note that in the above formulas any random variable W may be replaced by any group $XY\ldots Z$ and the relation holds true, since the set $XY\ldots Z$ may be regarded as the single random variable W, which takes its values in the cartesian product $\mathfrak{X}\times\mathfrak{Y}\times\cdots\times\mathfrak{Z}$.

§4. *Generalization and further properties of correlation*

Until now we have been concerned only with finite probability distributions, for which we have defined information and correlation. We shall now generalize the definition of correlation so as to be applicable to joint probability distributions over arbitrary sets of unrestricted cardinality.

We first consider the effects of refinement of a finite distribution. For example, we may discover that the event x_i is actually the disjunction of several exclusive events $\tilde{x}_i^1,\ldots,\tilde{x}_i^n$, so that x_i occurs if any one of the $\tilde{x}_i^\mu$ occurs, i.e., the single event x_i results from failing to distinguish between the $\tilde{x}_i^\mu$. The probability distribution which distinguishes between the $\tilde{x}_i^\mu$ will be called a *refinement* of the distribution which does not. In general, we shall say that a distribution $P' = P'(\tilde{x}_i^\mu,\ldots,\tilde{y}_j^\nu)$ is a refinement of $P = P(x_i,\ldots,y_j)$ if

$$(4.1) \qquad P(x_i,\ldots,y_j) = \sum_{\mu\ldots\nu} P'(\tilde{x}_i^\mu,\ldots,\tilde{y}_j^\nu) \quad (\text{all } i,\ldots,j) .$$

We now state an important theorem concerning the behavior of correlation under a refinement of a joint probability distributions:

THEOREM 2. *P' is a refinement of $P \Rightarrow \{X,\ldots,Y\}' \geqq \{X,\ldots,Y\}$ so that correlations never decrease upon refinement of a distribution.* (Proof in Appendix I, §3.)

As an example, suppose that we have a continuous probability density $P(x, y)$. By division of the axes into a finite number of intervals, $\bar{x}_i$, $\bar{y}_j$, we arrive at a finite joint distribution P_{ij}, by integration of $P(x, y)$ over the rectangle whose sides are the intervals $\bar{x}_i$ and $\bar{y}_j$, and which represents the probability that $X \epsilon \bar{x}_i$ and $Y \epsilon \bar{y}_j$. If we now subdivide the intervals, the new distribution P' will be a refinement of P, and by Theorem 2 the correlation $\{X,Y\}$ computed from P' will never be less than that computed from P. Theorem 2 is seen to be simply the mathematical verification of the intuitive notion that closer analysis of a situation in which quantities X and Y are dependent can never lessen the knowledge about Y which can be obtained from X.

This theorem allows us to give a general definition of correlation which will apply to joint distributions over completely arbitrary sets, i.e.,

for any probability measure[4] on an arbitrary product space, in the following manner:

Assume that we have a collection of arbitrary sets $\mathcal{X}, \mathcal{Y}, \ldots, \mathcal{Z}$, and a probability measure, $M_P(\mathcal{X} \times \mathcal{Y} \times \cdots \times \mathcal{Z})$, on their cartesian product. Let $\mathcal{P}^\mu$ be any finite partition of $\mathcal{X}$ into subsets $\mathcal{X}_i^\mu$, $\mathcal{Y}$ into subsets $\mathcal{Y}_j^\mu, \ldots$, and $\mathcal{Z}$ into subsets $\mathcal{Z}_k^\mu$, such that the sets $\mathcal{X}_i^\mu \times \mathcal{Y}_j^\mu \times \cdots \times \mathcal{Z}_k^\mu$ of the cartesian product are measurable in the probability measure M_P. Another partition $\mathcal{P}^\nu$ is a *refinement* of $\mathcal{P}^\mu$, $\mathcal{P}^\nu \subseteqq \mathcal{P}^\mu$, if $\mathcal{P}^\nu$ results from $\mathcal{P}^\mu$ by further subdivision of the subsets $\mathcal{X}_i^\mu, \mathcal{Y}_j^\mu, \ldots, \mathcal{Z}_k^\mu$. Each partition $\mathcal{P}^\mu$ results in a finite probability distribution, for which the correlation, $\{X, Y, \ldots, Z\}^{\mathcal{P}^\mu}$, is always defined through (3.3). Furthermore a refinement of a partition leads to a refinement of the probability distribution, so that by Theorem 2:

$$\mathcal{P}^\nu \subseteqq \mathcal{P}^\mu \Rightarrow \{X, Y, \ldots, Z\}^{\mathcal{P}^\nu} \geqq \{X, Y, \ldots, Z\}^{\mathcal{P}^\mu} \tag{4.8}$$

Now the set of all partitions is partially ordered under the refinement relation. Moreover, because for any pair of partitions $\mathcal{P}, \mathcal{P}'$ there is always a third partition $\mathcal{P}''$ which is a refinement of both (common lower bound), the set of all partitions forms a *directed set.*[5] For a function, f, on a directed set, δ, one defines a directed set limit, lim f,:

DEFINITION. lim f exists and is equal to $a \Leftrightarrow$ for every $\varepsilon > 0$ there exists an $\alpha \in \delta$ such that $|f(\beta) - a| < \varepsilon$ for every $\beta \in \delta$ for which $\beta \leqq \alpha$.

It is easily seen from the directed set property of common lower bounds that if this limit exists it is necessarily unique.

---

[4] A measure is a non-negative, countably additive set function, defined on some subsets of a given set. It is a probability measure if the measure of the entire set is unity. See Halmos [12].

[5] See Kelley [15], p. 65.

By (4.8) the correlation $\{X,Y,\ldots,Z\}^{\mathcal{P}}$ is a *monotone* function on the directed set of all partitions. Consequently the directed set limit, which we shall take as the basic definition of the correlation $\{X,Y,\ldots,Z\}$, *always exists*. (It may be infinite, but it is in every case well defined.) Thus:

DEFINITION. $\{X,Y,\ldots,Z\} = \lim \{X,Y,\ldots,Z\}^{\mathcal{P}}$,

and we have succeeded in our endeavor to give a completely general definition of correlation, applicable to all types of distributions.

It is an immediate consequence of (4.8) that this directed set limit is the supremum of $\{X,Y,\ldots,Z\}^{\mathcal{P}}$, so that:

$$\{X,Y,\ldots,Z\} = \sup_{\mathcal{P}} \{X,Y,\ldots,Z\}^{\mathcal{P}} \quad , \tag{4.9}$$

which we could equally well have taken as the definition.

Due to the fact that the correlation is defined as a limit for discrete distributions, Theorem 1 and all of the relations (3.7) to (3.15), which contain only correlation brackets, remain true for arbitrary distributions. Only (3.11) and (3.12), which contain information terms, cannot be extended.

We can now prove an important theorem about correlation which concerns its invariant nature. Let $\mathfrak{X},\mathfrak{Y},\ldots,\mathfrak{Z}$ be arbitrary sets with probability measure M_P on their cartesian product. Let f be any one-one mapping of $\mathfrak{X}$ onto a set $\mathfrak{U}$, g a one-one map of $\mathfrak{Y}$ onto $\mathfrak{V}$,..., and h a map of $\mathfrak{Z}$ onto $\mathfrak{W}$. Then a joint probability distribution over $\mathfrak{X}\times\mathfrak{Y}\times\ldots\times\mathfrak{Z}$ leads also to one over $\mathfrak{U}\times\mathfrak{V}\times\ldots\times\mathfrak{W}$ where the probability M'_P induced on the product $\mathfrak{U}\times\mathfrak{V}\times\ldots\times\mathfrak{W}$ is simply the measure which assigns to each subset of $\mathfrak{U}\times\mathfrak{V}\times\ldots\times\mathfrak{W}$ the measure which is the measure of its image set in $\mathfrak{X}\times\mathfrak{Y}\times\ldots\times\mathfrak{Z}$ for the original measure M_P. (We have simply transformed to a new set of random variables: U = f(X), V = g(Y), ..., W = h(Z).) Consider any partition $\mathcal{P}$ of $\mathfrak{X},\mathfrak{Y},\ldots,\mathfrak{Z}$ into the subsets $\{\mathfrak{X}_i\},\{\mathfrak{Y}_j\},\ldots,\{\mathfrak{Z}_k\}$ with probability distribution $P_{ij\ldots k} = M_P(\mathfrak{X}_i\times\mathfrak{Y}_j\times,\ldots,\times\mathfrak{Z}_k)$. Then there is a corresponding partition $\mathcal{P}'$ of $\mathfrak{U},\mathfrak{V},\ldots,\mathfrak{W}$ into the image

sets of the sets of $\mathcal{P}, \{\mathcal{U}_i\}, \{\mathcal{V}_j\}, \ldots, \{\mathcal{W}_k\}$, where $\mathcal{U}_i = f(\mathcal{X}_i)$, $\mathcal{V}_j = g(\mathcal{Y}_j), \ldots,$ $\mathcal{W}_k = h(\mathcal{Z}_k)$. But the probability distribution for $\mathcal{P}'$ is the same as that for $\mathcal{P}$, since $P'_{ij\ldots k} = M'_P(\mathcal{U}_i \times \mathcal{V}_j \times \cdots \times \mathcal{W}_k) = M_P(\mathcal{X}_i \times \mathcal{Y}_j \times \cdots \times \mathcal{Z}_k) = P_{ij\ldots k}$, so that:

$$\{X,Y,\ldots,Z\}^{\mathcal{P}} = \{U,V,\ldots,W\}^{\mathcal{P}'} \tag{4.10}$$

Due to the correspondence between the $\mathcal{P}$'s and $\mathcal{P}'$'s we have that:

$$\sup_{\mathcal{P}} \{X,Y,\ldots,Z\}^{\mathcal{P}} = \sup_{\mathcal{P}'} \{U,V,\ldots,W\}^{\mathcal{P}'}, \tag{4.11}$$

and by virtue of (4.9) we have proved the following theorem:

THEOREM 3. $\{X,Y,\ldots,Z\} = \{U,V,\ldots,W\}$, *where* $\mathcal{U}, \mathcal{V}, \ldots, \mathcal{W}$ *are any one-one images of* $\mathcal{X}, \mathcal{Y}, \ldots, \mathcal{Z}$, *respectively. In other notation:* $\{X,Y,\ldots,Z\} = \{f(X), g(Y),\ldots, h(Z)\}$ *for all one-one functions* $f, g, \ldots, h$.

This means that changing variables to functionally related variables preserves the correlation. Again this is plausible on intuitive grounds, since a knowledge of $f(x)$ is just as good as knowledge of x, provided that f is one-one.

A special consequence of Theorem 3 is that for any continuous probability density $P(x, y)$ over real numbers the correlation between $f(x)$ and $g(y)$ is the same as between x and y, where f and g are any real valued one-one functions. As an example consider a probability distribution for the position of two particles, so that the random variables are the position coordinates. Theorem 3 then assures us that the position correlation is *independent of the coordinate system*, even if different coordinate systems are used for each particle! Also for a joint distribution for a pair of events in space-time the correlation is invariant to arbitrary space-time coordinate transformations, again even allowing different transformations for the coordinates of each event.

These examples illustrate clearly the *intrinsic nature* of the correlation of various groups for joint probability distributions, which is implied by its invariance against arbitrary (one-one) transformations of the random variables. These correlation quantities are thus fundamental properties of probability distributions. A correlation is an *absolute* rather than *relative* quantity, in the sense that the correlation between (numerical valued) random variables is completely independent of the scale of measurement chosen for the variables.

§5. *Information for general distributions*

Although we now have a definition of correlation applicable to all probability distributions, we have not yet extended the definition of information past finite distributions. In order to make this extension we first generalize the definition that we gave for discrete distributions to a definition of *relative* information for a random variable, relative to a given underlying measure, called the *information measure*, on the values of the random variable.

If we assign a measure to the set of values of a random variable, X, which is simply the assignment of a positive number a_i to each value x_i in the finite case, we define the information of a probability distribution $P(x_i)$ *relative* to this *information measure* to be:

$$I_X = \sum_i P(x_i) \ln \frac{P(x_i)}{a_i} = \operatorname{Exp}\left[\ln \frac{P(x_i)}{a_i}\right] \tag{5.1}$$

If we have a joint distribution of random variables X,Y,...,Z, with information measures $\{a_i\},\{b_j\},\ldots,\{c_k\}$ on their values, then we define the total information relative to these measures to be:

$$I_{XY\ldots Z} = \sum_{ij\ldots k} P(x_i,y_j,\ldots,z_k) \ln \frac{P(x_i,y_j,\ldots,z_k)}{a_i b_j \ldots c_k} = \operatorname{Exp}\left[\ln \frac{P(x_i,y_j,\ldots,z_k)}{a_i b_j \ldots c_k}\right], \tag{5.2}$$

so that the information measure on the cartesian product set is *always* taken to be the product measure of the individual information measures.

We shall now alter our previous position slightly and consider information as always being defined relative to some information measure, so that our previous definition of information is to be regarded as the information relative to the measure for which all the a_i's, b_j's,... and c_k's are taken to be unity, which we shall henceforth call the *uniform measure.*

Let us now compute the correlation $\{X,Y,\ldots,Z\}'$ by (3.4) using the relative information:

$$(5.3)\quad \{X,Y,\ldots,Z\}' = I'_{XY\ldots Z} - I'_X - I'_Y - \ldots - I'_Z$$

$$= \operatorname{Exp}\left[\ln \frac{P(x_i,y_j,\ldots,z_k)}{a_i b_j \ldots c_k}\right] - \operatorname{Exp}\left[\ln \frac{P(x_i)}{a_i}\right] - \ldots -$$

$$\operatorname{Exp}\left[\ln \frac{P(s_k)}{c_k}\right]$$

$$= \operatorname{Exp}\left[\ln \frac{P(x_i,y_j,\ldots,z_k)}{P(x_i)P(y_j)\ldots P(z_k)}\right] = \{X,Y,\ldots,Z\}\ ,$$

so that the correlation for discrete distributions, as defined by (3.4), is independent of the choice of information measure, and the correlation remains an absolute, not relative quantity. It can, however, be computed from the information relative to any information measure through (3.4).

If we consider refinements, of our distributions, as before, and realize that such a refinement is also a refinement of the information measure, then we can prove a relation analogous to Theorem 2:

THEOREM 4. *The information of a distribution relative to a given information measure never decreases under refinement.* (Proof in Appendix I.)

Therefore, just as for correlation, we can define the information of a probability measure M_P on the cartesian product of arbitrary sets

$\mathfrak{X}, \mathfrak{Y}, \ldots, \mathfrak{Z}$, relative to the information measures $\mu_X, \mu_Y, \ldots, \mu_Z$, on the individual sets, by considering finite partitions $\mathcal{P}$ into subsets $\{\mathfrak{X}_i\}$, $\{\mathfrak{Y}_j\}, \ldots, \{\mathfrak{Z}_k\}$, for which we take as the definition of the information:

$$(5.4) \qquad I^{\mathcal{P}}_{XY\ldots Z} = \sum_{ij\ldots k} M_P(\mathfrak{X}_i, \mathfrak{Y}_j, \ldots, \mathfrak{Z}_k) \ln \frac{M_P(\mathfrak{X}_i, \mathfrak{Y}_j, \ldots, \mathfrak{Z}_k)}{\mu_X(\mathfrak{X}_i)\mu_Y(\mathfrak{Y}_j)\ldots\mu_Z(\mathfrak{Z}_k)}$$

Then $I^{\mathcal{P}}_{XY\ldots Z}$ is, as was $\{X,Y,\ldots,Z\}^{\mathcal{P}}$, a monotone function upon the directed set of partitions (by Theorem 4), and as before we take the directed set limit for our definition:

$$(5.5) \qquad I_{XY\ldots Z} = \lim I^{\mathcal{P}}_{XY\ldots Z} = \sup_{\mathcal{P}} I^{\mathcal{P}}_{XY\ldots Z}$$

which is then the information relative to the information measures $\mu_X, \mu_Y, \ldots, \mu_Z$.

Now, for functions f, g on a directed set the existence of $\lim f$ and $\lim g$ is a sufficient condition for the existence of $\lim (f+g)$, which is then $\lim f + \lim g$, provided that this is not indeterminate. Therefore:

THEOREM 5. $\{X,\ldots,Y\} = \lim \{X,\ldots,Y\}^{\mathcal{P}} = \lim \left[I^{\mathcal{P}}_{X\ldots Y} - I^{\mathcal{P}}_{X} - \ldots - I^{\mathcal{P}}_{Y}\right] = I_{X\ldots Y} - I_X - \ldots - I_Y$, *where the information is taken relative to any information measure for which the expression is not indeterminate. It is sufficient for the validity of the above expression that the basic measures* $\mu_X, \ldots, \mu_Y$ *be such that none of the marginal informations* $I_X \ldots I_Y$ *shall be positively infinite.*

The latter statement holds since, because of the general relation $I_{X\ldots Y} \geqq I_X + \ldots + I_Y$, the determinateness of the expression is guaranteed so long as all of the $I_X, \ldots, I_Y$ are $< +\infty$.

Henceforth, unless otherwise noted, we shall understand that information is to be computed with respect to the uniform measure for discrete distributions, and Lebesgue measure for continuous distributions over real

numbers. In case of a mixed distribution, with a continuous density $P(x,y,\ldots,z)$ plus discrete "lumps" $P'(x_i,y_j,\ldots,z_k)$, we shall understand the information measure to be the uniform measure over the discrete range, and Lebesgue measure over the continuous range. These conventions then lead us to the expressions:

$$(5.6)\qquad I_{XY\ldots Z} = \begin{cases} \displaystyle\sum_{ij\ldots k} P(x_i,y_j,\ldots,z_k) \ln P(x_i,y_j,\ldots,z_k) \Big\} \text{(discrete)} \\ \displaystyle\int P(x,y,\ldots,z) \ln P(x,y,\ldots,z)\, dxdy\ldots dz \Big\} \text{(cont.)} \\ \left.\begin{array}{l} \displaystyle\sum_{i\ldots k} P'(x_i,\ldots,z_k) \ln P(x_i,\ldots,z_k) \\ \displaystyle + \int P(x,\ldots,z) \ln P(x,\ldots,z)\, dx\ldots dz \end{array}\right\} \text{(mixed)} \\ \qquad \text{(unless otherwise noted)} \end{cases}$$

The mixed case occurs often in quantum mechanics, for quantities which have both a discrete and continuous spectrum.

§6. *Example: Information decay in stochastic processes*

As an example illustrating the usefulness of the concept of relative information we shall consider briefly stochastic processes.[6] Suppose that we have a stationary Markov[7] process with a finite number of states S_i, and that the process occurs at discrete (integral) times $1,2,\ldots,n,\ldots$, at which times the transition probability from the state S_i to the state S_j is T_{ij}. The probabilities T_{ij} then form what is called a *stochastic*

6 See Feller [10], or Doob [6].

7 A Markov process is a stochastic process whose future development depends only upon its present state, and not on its past history.

matrix, i.e., the elements are between 0 and 1, and $\sum_i T_{ij} = 1$ for all i. If at any time k the probability distribution over the states is $\{P_i^k\}$ then at the next time the probabilities will be $P_j^{k+1} = \sum P_i^k T_{ij}$.

In the special case where the matrix is *doubly-stochastic*, which means that $\sum_i T_{ij}$, as well as $\sum_j T_{ij}$, equals unity, and which amounts to a principle of detailed balancing holding, it is known that the entropy of a probability distribution over the states, defined as $H = -\sum_i P_i \ln P_i$, is a monotone increasing function of the time. This entropy is, however, simply the negative of the information relative to the uniform measure.

One can extend this result to more general stochastic processes only if one uses the more general definition of relative information. For an arbitrary stationary process the choice of an information measure which is stationary, i.e., for which

$$a_j = \sum_i a_i T_{ij} \quad \text{(all j)} \tag{6.1}$$

leads to the desired result. In this case the *relative* information,

$$I = \sum_i P_i \ln \frac{P_i}{a_i} , \tag{6.2}$$

is a monotone decreasing function of time and constitutes a suitable basis for the definition of the entropy $H = -I$. Note that this definition leads to the previous result for doubly-stochastic processes, since the uniform measure, $a_i = 1$ (all i), is obviously stationary in this case.

One can furthermore drop the requirement that the stochastic process be stationary, and even allow that there are completely different sets of states, $\{S_i^n\}$, at each time n, so that the process is now given by a sequence of matrices T_{ij}^n representing the transition probability at time n from state S_i^n to state S_j^{n+1}. In this case probability distributions change according to:

$$P_j^{n+1} = \sum_i P_i^n T_{ij}^n \quad . \tag{6.3}$$

If we then choose *any* time-dependent information measure which satisfies the relations:

$$a_j^{n+1} = \sum a_i^n T_{ij}^n \quad \text{(all j, n)} \quad , \tag{6.4}$$

then the information of a probability distribution is again monotone decreasing with time. (Proof in Appendix I.)

All of these results are easily extended to the continuous case, and we see that the concept of relative information allows us to define entropy for quite general stochastic processes.

§7. *Example: Conservation of information in classical mechanics*

As a second illustrative example we consider briefly the classical mechanics of a group of particles. The system at any instant is represented by a point, $(x^1,y^1,z^1,p_x^1,p_y^1,p_z^1,\ldots,x^n,y^n,z^n,p_x^n,p_y^n,p_z^n)$, in the phase space of all position and momentum coordinates. The natural motion of the system then carries each point into another, defining a continuous transformation of the phase space into itself. According to Liouville's theorem the measure of a set of points of the phase space is invariant under this transformation.[8] This invariance of measure implies that if we begin with a probability distribution over the phase space, rather than a single point, the total information

$$I_{total} = I_{X^1Y^1Z^1P_x^1P_y^1P_z^1 \ldots X^nY^nZ^nP_x^nP_y^nP_z^n} \quad , \tag{7.1}$$

which is the information of the *joint* distribution for all positions and momenta, remains *constant in time.*

---

8 See Khinchin [16], p. 15.

In order to see that the total information is conserved, consider any partition $\mathcal{P}$ of the phase space at one time, t_0, with its information relative to the phase space measure, $I^{\mathcal{P}}(t_0)$. At a later time t_1 a partition $\mathcal{P}'$, into the image sets of $\mathcal{P}$ under the mapping of the space into itself, is induced, for which the probabilities for the sets of $\mathcal{P}'$ are the same as those of the corresponding sets of $\mathcal{P}$, and furthermore for which the measures are the same, by Liouville's theorem. Thus corresponding to each partition $\mathcal{P}$ at time t_0 with information $I^{\mathcal{P}}(t_0)$, there is a partition $\mathcal{P}'$ at time t_1 with information $I^{\mathcal{P}'}(t_1)$, which is the same:

$$I^{\mathcal{P}'}(t_1) = I^{\mathcal{P}}(t_0) \ . \tag{7.2}$$

Due to the correspondence of the $\mathcal{P}$'s and $\mathcal{P}'$s the supremums of each over all partitions must be equal, and by (5.5) we have proved that

$$I_{\text{total}}(t_1) = I_{\text{total}}(t_0) \ , \tag{7.3}$$

and the total information is conserved.

Now it is known that the individual (marginal) position and momentum distributions tend to decay, except for rare fluctuations, into the uniform and Maxwellian distributions respectively, for which the classical entropy is a maximum. This entropy is, however, except for the factor of Boltzman's constant, simply the negative of the marginal information

$$I_{\text{marginal}} = I_{X_1} + I_{Y_1} + I_{Z_1} + \ldots + I_{P_x^n} + I_{P_y^n} + I_{P_z^n} \ , \tag{7.4}$$

which thus tends towards a minimum. But this decay of marginal information is exactly compensated by an increase of the total correlation information

$$\{\text{total}\} = I_{\text{total}} - I_{\text{marginal}} \ , \tag{7.5}$$

since the total information remains constant. Therefore, if one were to define the *total entropy* to be the negative of the total information, one could replace the usual second law of thermodynamics by a law of

conservation of total entropy, where the increase in the standard (marginal) entropy is exactly compensated by a (negative) *correlation entropy*. The usual second law then results simply from our renunciation of all correlation knowledge (*stosszahlansatz*), and not from any intrinsic behavior of classical systems. The situation for classical mechanics is thus in sharp contrast to that of stochastic processes, which are intrinsically irreversible.

III. QUANTUM MECHANICS

Having mathematically formulated the ideas of information and correlation for probability distributions, we turn to the field of quantum mechanics. In this chapter we assume that the states of physical systems are represented by points in a Hilbert space, and that the time dependence of the state of an isolated system is governed by a linear wave equation.

It is well known that state functions lead to distributions over eigenvalues of Hermitian operators (square amplitudes of the expansion coefficients of the state in terms of the basis consisting of eigenfunctions of the operator) which have the mathematical properties of probability distributions (non-negative and normalized). The standard interpretation of quantum mechanics regards these distributions as actually giving the probabilities that the various eigenvalues of the operator will be observed, when a measurement represented by the operator is performed.

A feature of great importance to our interpretation is the fact that a state function of a *composite* system leads to *joint* distributions over subsystem quantities, rather than independent subsystem distributions, i.e., the quantities in different subsystems may be correlated with one another. The first section of this chapter is accordingly devoted to the development of the formalism of composite systems, and the connection of composite system states and their derived joint distributions with the various possible subsystem conditional and marginal distributions. We shall see that there exist *relative state functions* which correctly give the conditional distributions for all subsystem operators, while marginal distributions can *not* generally be represented by state functions, but only by *density matrices*.

In Section 2 the concepts of information and correlation, developed in the preceding chapter, are applied to quantum mechanics, by defining

information and correlation for operators on systems with prescribed states. It is also shown that for composite systems there exists a quantity which can be thought of as the fundamental correlation between subsystems, and a closely related *canonical representation* of the composite system state. In addition, a stronger form of the uncertainty principle, phrased in information language, is indicated.

The third section takes up the question of measurement in quantum mechanics, viewed as a correlation producing interaction between physical systems. A simple example of such a measurement is given and discussed. Finally some general consequences of the superposition principle are considered.

It is convenient at this point to introduce some notational conventions. We shall be concerned with points ψ in a Hilbert space $\mathcal{H}$, with scalar product (ψ_1, ψ_2). A *state* is a point ψ for which $(\psi, \psi) = 1$. For any linear operator A we define a functional, $<A>\psi$, called the *expectation of* A *for* ψ, to be:

$$<A>\psi = (\psi, A\psi) .$$

A class of operators of particular interest is the class of *projection operators.* The operator $[\phi]$, called the projection on ϕ, is defined through:

$$[\phi]\psi = (\phi, \psi)\phi .$$

For a complete orthonormal set $\{\phi_i\}$ and a state ψ we define a *square-amplitude distribution,* P_i, called the distribution of ψ over $\{\phi_i\}$ through:

$$P_i = |(\phi_i, \psi)|^2 = <[\phi_i]>\psi .$$

In the probabilistic interpretation this distribution represents the probability distribution over the results of a measurement with eigenstates ϕ_i, performed upon a system in the state ψ. (Hereafter when referring to the probabilistic interpretation we shall say briefly "the probability that the system will be found in ϕ_i", rather than the more cumbersome phrase "the probability that the measurement of a quantity B, with eigenfunc-

tions $\{\phi_i\}$, shall yield the eigenvalue corresponding to ϕ_i," which is meant.)

For two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, we form the *direct product* Hilbert space $\mathcal{H}_3 = \mathcal{H}_1 \otimes \mathcal{H}_2$ (tensor product) which is taken to be the space of all possible[1] sums of formal products of points of $\mathcal{H}_1$ and $\mathcal{H}_2$, i.e., the elements of $\mathcal{H}_3$ are those of the form $\sum_i a_i \xi_i \eta_i$ where $\xi_i \in \mathcal{H}_1$ and $\eta_i \in \mathcal{H}_2$. The scalar product in $\mathcal{H}_3$ is taken to be $\left(\sum_i a_i \xi_i \eta_i, \sum_j b_j \xi_j \eta_j\right) = \sum_{ij} a_i^* b_j (\xi_i, \xi_j)(\eta_i, \eta_j)$. It is then easily seen that if $\{\xi_i\}$ and $\{\eta_i\}$ form complete orthonormal sets in $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, then the set of all formal products $\{\xi_i \eta_j\}$ is a complete orthonormal set in $\mathcal{H}_3$. For any pair of operators A, B, in $\mathcal{H}_1$ and $\mathcal{H}_2$ there corresponds an operator $C = A \otimes B$, the direct product of A and B, in $\mathcal{H}_3$, which can be defined by its effect on the elements $\xi_i \eta_j$ of $\mathcal{H}_3$:

$$C \xi_i \eta_j = A \otimes B \xi_i \eta_j = (A \xi_i)(B \eta_j) .$$

§1. *Composite systems*

It is well known that if the states of a pair of systems S_1 and S_2, are represented by points in Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, then the states of the *composite system* $S = S_1 + S_2$ (the two systems S_1 and S_2 regarded as a single system S) are represented correctly by points of the direct product $\mathcal{H}_1 \otimes \mathcal{H}_2$. This fact has far reaching consequences which we wish to investigate in some detail. Thus if $\{\xi_i\}$ is a complete orthonormal set for $\mathcal{H}_1$, and $\{\eta_j\}$ for $\mathcal{H}_2$, the general state of $S = S_1 + S_2$ has the form:

$$\psi^S = \sum_{ij} a_{ij} \xi_i \eta_j \quad \left(\sum_{ij} a_{ij}^* a_{ij} = 1\right) . \tag{1.1}$$

---

[1] More rigorously, one considers only *finite* sums, then completes the resulting space to arrive at $\mathcal{H}_1 \otimes \mathcal{H}_2$.

In this case we shall call $P_{ij} = a^*_{ij}a_{ij}$ the *joint square-amplitude distribution* of ψ^S over $\{\xi_i\}$ and $\{\eta_j\}$. In the standard probabilistic interpretation $a^*_{ij}a_{ij}$ represents the joint probability that S_1 will be found in the state ξ_i *and* S_2 will be found in the state η_i. Following the probabilistic model we now derive some distributions from the state ψ^S. Let A be a Hermitian operator in S_1 with eigenfunctions ϕ_i and eigenvalues λ_i, and B an operator in S_2 with eigenfunctions θ_j and eigenvalues μ_j. Then the joint distribution of ψ^S over $\{\phi_i\}$ and $\{\phi_j\}$, P_{ij}, is:

$$P_{ij} = P(\phi_i \text{ and } \theta_j) = |(\phi_i\theta_j, \psi^S)|^2 . \tag{1.2}$$

The *marginal* distributions, of ψ^S over $\{\phi_i\}$ and of ψ^S over $\{\phi_j\}$, are:

$$P_i = P(\phi_i) = \sum_j P_{ij} = \sum_j |(\phi_i\theta_j, \psi^S)|^2 , \tag{1.3}$$

$$P_j = P(\theta_j) = \sum_i P_{ij} = \sum_i |(\phi_i\theta_j, \psi^S)|^2 ,$$

and the *conditional distributions* P_i^j and P_j^i are:

$$P_i^j = P(\phi_i \text{ conditioned on } \phi_j) = \frac{P_{ij}}{P_j} , \tag{1.4}$$

$$P_j^i = P(\phi_j \text{ conditioned on } \phi_i) = \frac{P_{ij}}{P_i} .$$

We now define the *conditional expectation* of an operator A on S_1, conditioned on θ_j in S_2, denoted by $\mathrm{Exp}^{\theta_j}[A]$, to be:

$$\begin{aligned} \mathrm{Exp}^{\theta_j}[A] &= \sum_i \lambda_i P_i^j = (1/P_j)\sum_i P_{ij}\lambda_i \\ &= (1/P_j)\sum_i \lambda_i |(\phi_i\theta_j, \psi^S)|^2 \\ &= (1/P_j)\sum_i |(\phi_i\theta_j, \psi^S)|^2 (\phi_i, A\phi_i) , \end{aligned} \tag{1.5}$$

and we define the *marginal expectation* of A on S_1 to be:

$$(1.6) \qquad \mathrm{Exp}\,[A] = \sum_i P_i \lambda_i = \sum_{ij} \lambda_i P_{ij} = \sum_{ij} |(\phi_i \theta_j, \psi^S)|^2 (\phi_i, A\phi_i)$$

We shall now introduce projection operators to get more convenient forms of the conditional and marginal expectations, which will also exhibit more clearly the degree of dependence of these quantities upon the chosen basis $\{\phi_i \theta_j\}$. Let the operators $[\phi_i]$ and $[\phi_j]$ be the projections on ϕ_i in S_1 and ϕ_j in S_2 respectively, and let I^1 and I^2 be the identity operators in S_1 and S_2. Then, making use of the identity $\psi^S = \sum_{ij} (\phi_i \theta_j, \psi^S) \phi_i \theta_j$ for any complete orthonormal set $\{\phi_i \theta_j\}$, we have:

$$(1.7) \quad <[\phi_i][\theta_j]>_{\psi^S} = (\psi^S, [\phi_i][\theta_j]\psi^S) =$$

$$\left(\sum_{k\ell} (\phi_k \theta_\ell, \psi^S) \phi_k \theta_\ell, [\phi_i][\theta_j] \sum_{mn} (\phi_m \theta_n, \psi^S) \phi_m \theta_n\right)$$

$$= \sum_{k\ell mn} (\phi_k \theta_\ell, \psi^S)^* (\phi_m \theta_n, \psi^S) \delta_{km} \delta_{\ell n} \delta_{im} \delta_{jn}$$

$$= (\phi_i \theta_j, \psi^S)^* (\phi_i \theta_j, \psi^S) = P_{ij} \ ,$$

so that the joint distribution is given simply by $<[\phi_i][\phi_j]>_{\psi^S}$.

For the marginal distribution we have:

$$(1.8) \quad P_i = \sum_j P_{ij} = \sum_j <[\phi_i][\theta_j]>_{\psi^S} = <[\phi_i]\left(\sum_j [\theta_j]\right)>_{\psi^S} = <[\phi_i] I^2>_{\psi^S} \ ,$$

and we see that the marginal distribution over the ϕ_i is *independent* of the set $\{\theta_j\}$ chosen in S_2. This result has the consequence in the ordinary interpretation that the expected outcome of measurement in one subsystem of a composite system is not influenced by the choice of quantity to be measured in the other subsystem. This expectation is, in fact, the expectation for the case in which no measurement at all (identity operator) is performed in the other subsystem. Thus no measurement in S_2 can

affect the expected outcome of a measurement in S_1, *so long as the result of any* S_2 *measurement remains unknown.* The case is quite different, however, if this result *is* known, and we must turn to the conditional distributions and expectations in such a case.

We now introduce the concept of a *relative state-function,* which will play a central role in our interpretation of pure wave mechanics. Consider a composite system $S = S_1 + S_2$ in the state ψ^S. To every state η of S_2 we associate a state of S_1, ψ^{η}_{rel}, called the relative state in S_1 for η in S_2, through:

(1.9) DEFINITION. $$\psi^{\eta}_{rel} = N \sum_i (\phi_i \eta, \psi^S) \phi_i ,$$

where $\{\phi_i\}$ is *any* complete orthonormal set in S_1 and N is a normalization constant.[2]

The first property of ψ^{η}_{rel} is its uniqueness,[3] i.e., its dependence upon the choice of the basis $\{\phi_i\}$ is only apparent. To prove this, choose another basis $\{\xi_k\}$, with $\phi_i = \sum_k b_{ik} \xi_k$. Then $\sum_i b^*_{ij} b_{ik} = \delta_{jk}$, and:

$$\sum_i (\phi_i \eta, \psi^S) \phi_i = \sum_i \left(\sum_j b_{ij} \xi_j \eta, \psi^S \right) \left(\sum_k b_{ik} \xi_k \right)$$

$$= \sum_{jk} \left(\sum_i b^*_{ij} b_{ik} \right) (\xi_j \eta, \psi^S) \xi_k = \sum_{jk} \delta_{jk} (\xi_j \eta, \psi^S) \xi_k$$

$$= \sum_k (\xi_k \eta, \psi^S) \xi_k \ .$$

The second property of the relative state, which justifies its name, is that $\psi^{\theta_j}_{rel}$ correctly gives the *conditional expectations* of all operators in S_1, conditioned by the state θ_j in S_2. As before let A be an operator in S_1 with eigenstates ϕ_i and eigenvalues λ_i. Then:

---

2 In case $\sum_i (\phi_i \eta, \psi^S) \phi_i = 0$ (unnormalizable) then choose any function for the relative function. This ambiguity has no consequences of any importance to us. See in this connection the remarks on p. 40.

3 Except if $\sum_i (\phi_i \eta, \psi^S) \phi_i = 0$. There is still, of course, no dependence upon the basis.

$$(1.10)\qquad <A>\psi^{\theta_j}_{rel} = \left(\psi^{\theta_j}_{rel}, A\psi^{\theta_j}_{rel}\right)$$

$$= \left(N \sum_i (\phi_i \theta_j, \psi^S)\phi_i, AN \sum_{im} (\phi_m \theta_j, \psi^S)\phi_m\right)$$

$$= N^2 \sum_{im} (\phi_i \theta_j, \psi^S)^* (\phi_m \theta_j, \psi^S)\lambda_m \delta_{im}$$

$$= N^2 \sum_i \lambda_i P_{ij} \ .$$

At this point the normalizer N^2 can be conveniently evaluated by using (1.10) to compute: $<I^1>\psi^{\theta_j}_{rel} = N^2 \sum_i 1\ P_{ij} = N^2 P_j = 1$, so that

$$(1.11)\qquad N^2 = 1/P_j \ .$$

Substitution of (1.11) in (1.10) yields:

$$(1.12)\qquad <A>\psi^{\theta_j}_{rel} = (1/P_j) \sum_i \lambda_i P_{ij} = \sum_i \lambda_i P^j_i = Exp^{\theta_j}[A] \ ,$$

and we see that the conditional expectations of operators are given by the relative states. (This includes, of course, the conditional distributions themselves, since they may be obtained as expectations of projection operators.)

An important representation of a composite system state ψ^S, in terms of an orthonormal set $\{\theta_j\}$ in one subsystem S_2 and the set of relative states $\left\{\psi^{\theta_j}_{rel}\right\}$ in S_1 is:

$$(1.13)\qquad \psi^S = \sum_{ij} (\phi_i \theta_j, \psi^S)\phi_i \theta_j = \sum_j \left(\sum_i (\phi_i \theta_j, \psi^S)\phi_i\right)\theta_j$$

$$= \sum_j \frac{1}{N_j}\left[N_j \sum_i (\phi_i \theta_j, \psi^S)\phi_i\right]\theta_j$$

$$= \sum_j \frac{1}{N_j}\psi^{\theta_j}_{rel}\,\theta_j \ , \quad \text{where } 1/N_j^2 = P_j = <I^1[\theta_j]>\psi^S$$

Thus, for *any* orthonormal set in one subsystem, the state of the composite system is a single superposition of elements consisting of a state of the given set and its relative state in the other subsystem. (The relative states, however, are not necessarily orthogonal.) We notice further that a particular element, $\psi_{\text{rel}}^{\theta_j}\theta_j$, is quite independent of the choice of basis $\{\theta_k\}$, $k \neq j$, for the orthogonal space of θ_j, since $\psi_{\text{rel}}^{\theta_j}$ depends *only* on θ_j and not on the other θ_k for $k \neq j$. We remark at this point that the ambiguity in the relative state which arises when $\sum_i (\phi_i\theta_j, \psi^S)\phi_i = 0$ (see p. 38) is unimportant for this representation, since although *any* state $\psi_{\text{rel}}^{\theta_j}$ can be regarded as the relative state in this case, the term $\psi_{\text{rel}}^{\theta_j}\theta_j$ will occur in (1.13) with coefficient zero.

Now that we have found subsystem states which correctly give conditional expectations, we might inquire whether there exist subsystem states which give marginal expectations. The answer is, unfortunately, no. Let us compute the marginal expectation of A in S_1 using the representation (1.13):

$$\text{(1.14)} \qquad \text{Exp}[A] = \langle A\, I^2 \rangle \psi^S = \left(\sum_j \frac{1}{N_j} \psi_{\text{rel}}^{\theta_j}\theta_j, A\, I^2 \sum_k \frac{1}{N_k} \psi_{\text{rel}}^{\theta_k}\theta_k \right)$$

$$= \sum_{jk} \frac{1}{N_j N_k} \left(\psi_{\text{rel}}^{\theta_j}, A\psi_{\text{rel}}^{\theta_j} \right) \delta_{jk}$$

$$= \sum_j \frac{1}{N_j^2} \left(\psi_{\text{rel}}^{\theta_j}, A\psi_{\text{rel}}^{\theta_j} \right) = \sum_j P_j \langle A \rangle \psi_{\text{rel}}^{\theta_j} .$$

Now suppose that there exists a state in S_1, ψ', which correctly gives the marginal expectation (1.14) for *all* operators A (i.e., such that $\text{Exp}[A] = \langle A \rangle \psi'$ for all A). One such operator is $[\psi']$, the projection on ψ', for which $\langle [\psi'] \rangle \psi' = 1$. But, from (1.14) we have that $\text{Exp}[\psi'] = \sum_j P_j \langle \psi' \rangle \psi_{\text{rel}}^{\theta_j}$, which is < 1 unless, for all j, $P_j = 0$ or $\psi_{\text{rel}}^{\theta_j} = \psi'$, a condition which is not generally true. Therefore *there exists in general no state for* S_1 *which correctly gives the marginal expectations for all operators in* S_1.

However, even though there is generally no single state describing marginal expectations, we see that there is always a *mixture* of states, namely the states $\psi_{\text{rel}}^{\theta_j}$ *weighted* with P_j, which does yield the correct expectations. The distinction between a mixture, M, of states ϕ_i, weighted by P_i, and a *pure state* ψ which is a superposition, $\psi = \sum a_i \phi_i$, is that there are *no interference phenomena* between the various states of a mixture. The expectation of an operator A for the mixture is $\text{Exp}^M[A] = \sum_i P_i \langle A \rangle_{\phi_i} = \sum_i P_i (\phi_i, A\phi_i)$, while the expectation for the pure state ψ is $\langle A \rangle_\psi = \left(\sum_i a_i \phi_i, A \sum_j a_j \phi_j \right) = \sum_{ij} a_i^* a_j (\phi_i, A\phi_j)$, which is *not* the same as that of the mixture with weights $P_i = a_i^* a_i$, due to the presence of the interference terms $(\phi_i, A\phi_j)$ for $j \neq i$.

It is convenient to represent such a mixture by a *density matrix*,[4] ρ. If the mixture consists of the states ψ_j weighted by P_j, and if we are working in a basis consisting of the complete orthonormal set $\{\phi_i\}$, where $\psi_j = \sum_i a_i^j \phi_i$, then we define the elements of the density matrix for the mixture to be:

$$\rho_{k\ell} = \sum_j P_j \, a_\ell^{j*} \, a_k^j \qquad (a_i^j = (\phi_i, \psi_j)) \; . \tag{1.15}$$

Then if A is any operator, with matrix representation $A_{i\ell} = (\phi_i, A\phi_\ell)$ in the chosen basis, its expectation for the mixture is:

$$\begin{aligned} \text{Exp}^M[A] &= \sum_j P_j (\psi_j, A\psi_j) = \sum_j P_j \left[\sum_{i\ell} a_i^{j*} a_\ell^j (\phi_i, A\phi_\ell) \right] \\ &= \sum_{i\ell} \left(\sum_j P_j \, a_i^{j*} a_\ell^j \right) (\phi_i, A\phi_\ell) = \sum_{i,\ell} \rho_{\ell i} A_{i\ell} \\ &= \text{Trace } (\rho A) \; . \end{aligned} \tag{1.16}$$

[4] Also called a *statistical operator* (von Neumann [17]).

Therefore any mixture is adequately represented by a density matrix.[5] Note also that $\rho^*_{k\ell} = \rho_{\ell k}$, so that ρ is Hermitian.

Let us now find the density matrices ρ^1 and ρ^2 for the subsystems S_1 and S_2 of a system $S = S_1 + S_2$ in the state ψ^S. Furthermore, let us choose the orthonormal bases $\{\xi_i\}$ and $\{\eta_j\}$ in S_1 and S_2 respectively, and let A be an operator in S_1, B an operator in S_2. Then:

$$(1.17)\quad \mathrm{Exp}[A] = \langle AI^2 \rangle_{\psi^S} = \left(\sum_{ij} (\xi_i \eta_j, \psi^S) \xi_i \eta_j, AI \sum_{\ell m} (\xi_\ell \eta_m, \psi^S) \xi_\ell \eta_m \right)$$

$$= \sum_{ij\ell m} (\xi_i \eta_j, \psi^S)^* (\xi_\ell \eta_m, \psi^S) (\xi_i, A\xi_\ell)(\eta_j, \eta_m)$$

$$= \sum_{i\ell} \left[\sum_j (\xi_i \eta_j, \psi^S)^* (\xi_\ell \eta_j, \psi^S) \right] (\xi_i, A\xi_\ell)$$

$$= \mathrm{Trace}\,(\rho^1 A)\ ,$$

where we have defined ρ^1 in the $\{\xi_i\}$ basis to be:

$$(1.18)\qquad \rho^1_{\ell i} = \sum_j (\xi_i \eta_j, \psi^S)^* (\xi_\ell \eta_j, \psi^S)\ .$$

In a similar fashion we find that ρ^2 is given, in the $\{\eta_j\}$ basis, by:

$$(1.19)\qquad \rho^2_{mn} = \sum_i (\xi_i \eta_n, \psi^S)^* (\xi_i \eta_m, \psi^S)\ .$$

It can be easily shown that here again the dependence of ρ^1 upon the choice of basis $\{\eta_j\}$ in S_2, and of ρ^2 upon $\{\xi_i\}$, is only apparent.

---

[5] A better, coordinate free representation of a mixture is in terms of the operator which the density matrix represents. For a mixture of states ψ_n (not necessarily orthogonal) with weights ρ_n, the density operator is $\rho = \sum_n \rho_n [\psi_n]$, where $[\psi_n]$ stands for the projection operator on ψ_n.

In summary, we have seen in this section that a state of a composite system leads to *joint* distributions over subsystem quantities which are generally not independent. Conditional distributions and expectations for subsystems are obtained from *relative states*, and subsystem marginal distributions and expectations are given by *density matrices.*

There does not, in general, exist anything like a single state for one subsystem of a composite system. That is, subsystems do *not* possess states independent of the states of the remainder of the system, so that the subsystem states are generally *correlated.* One can arbitrarily choose a state for one subsystem, and be led to the *relative state* for the other subsystem. Thus we are faced with a fundamental *relativity of states*, which is implied by the formalism of composite systems. It is meaningless to ask the absolute state of a subsystem – one can only ask the state relative to a given state of the remainder of the system.

§2. *Information and correlation in quantum mechanics*

We wish to be able to discuss information and correlation for Hermitian operators $A, B, \ldots$, with respect to a state function ψ. These quantities are to be computed, through the formulas of the preceding chapter, from the square amplitudes of the coefficients of the expansion of ψ in terms of the eigenstates of the operators.

We have already seen (p. 34) that a state ψ and an orthonormal basis $\{\phi_i\}$ leads to a square amplitude distribution of ψ over the set $\{\phi_i\}$:

$$P_i = |(\phi_i, \psi)|^2 = <[\phi_i]>_\psi \ , \tag{2.1}$$

so that we can define the *information of the basis* $\{\phi_i\}$ *for the state* ψ, $I_{\{\phi_i\}}(\psi)$, to be simply the information of this distribution relative to the uniform measure:

$$I_{\{\phi_i\}}(\psi) = \sum_i P_i \ln P_i = \sum_i |(\phi_i, \psi)|^2 \ln |(\phi_i, \psi)|^2 \ . \tag{2.2}$$

We define the *information of an operator* A, for the state ψ, $I_A(\psi)$, to be the information in the square amplitude distribution over its *eigenvalues*, i.e., the information of the probability distribution over the results of a determination of A which is prescribed in the probabilistic interpretation. For a *non-degenerate* operator A this distribution is the same as the distribution (2.1) over the eigenstates. But because the information is dependent only on the distribution, and not on numerical values, the information of the distribution over eigenvalues of A is precisely the information of the eigenbasis of A, $\{\phi_i\}$. Therefore:

$$(2.3)\quad I_A(\psi) = I_{\{\phi_i\}}(\psi) = \sum_i <[\phi_i]>_\psi \ln <[\phi_i]>_\psi \quad \text{(A non-degenerate)}.$$

We see that for fixed ψ, the information of all non-degenerate operators having the same set of eigenstates is the same.

In the case of *degenerate* operators it will be convenient to take, as the definition of information, the information of the square amplitude distribution over the eigenvalues *relative* to the information measure which consists of the *multiplicity* of the eigenvalues, rather than the uniform measure. This definition preserves the choice of uniform measure over the *eigenstates*, in distinction to the eigenvalues. If ϕ_{ij} (j from 1 to m_i) are a complete orthonormal set of eigenstates for A′, with distinct eigenvalues λ_i (degenerate with respect to j), then the multiplicity of the i^{th} eigenvalue is m_i and the information $I_{A'}(\psi)$ is defined to be:

$$(2.4)\qquad I_{A'}(\psi) = \sum_i \left(\sum_j <[\phi_{ij}]>_\psi \right) \ln \frac{\sum_j <[\phi_{ij}]>_\psi}{m_i} .$$

The usefulness of this definition lies in the fact that any operator A″ which distinguishes further between any of the degenerate states of A′ leads to a refinement of the relative density, in the sense of Theorem 4, and consequently has equal or greater information. A non-degenerate operator thus represents the maximal refinement and possesses maximal information.

It is convenient to introduce a new notation for the projection operators which are *relevant* for a specified operator. As before let A have eigenfunctions ϕ_{ij} and distinct eigenvalues λ_i. Then define the projections A_i, the projections on the *eigenspaces* of different eigenvalues of A, to be:

$$A_i = \sum_{j=1}^{m_i} [\phi_{ij}] \ . \tag{2.5}$$

To each such projection there is associated a number m_i, the multiplicity of the degeneracy, which is the dimension of the i^{th} eigenspace. In this notation the distribution over the eigenvalues of A for the state ψ, P_i, becomes simply:

$$P_i = P(\lambda_i) = <A_i>\psi \ , \tag{2.6}$$

and the information, given by (2.4), becomes:

$$I_A = \sum_i <A_i>\psi \ln \frac{<A_i>\psi}{m_i} \ . \tag{2.7}$$

Similarly, for a pair of operators, A in S_1 and B in S_2, for the composite system $S = S_1 + S_2$ with state ψ^S, the *joint* distribution over eigenvalues is:

$$P_{ij} = P(\lambda_i, \mu_j) = <A_iB_j>\psi^S \ , \tag{2.8}$$

and the marginal distributions are:

$$P_i = \sum_j P_{ij} = <A_i\left(\sum_j B_j\right)>\psi^S = <A_iI^2>\psi^S \ , \tag{2.9}$$

$$P_j = \sum_i P_{ij} = <\left(\sum_i A_i\right)B_j>\psi^S = <I^1B_j>\psi^S \ .$$

The *joint* information, I_{AB}, is given by:

$$I_{AB} = \sum_{ij} P_{ij} \ln \frac{P_{ij}}{m_in_j} = \sum_{ij} <A_iB_j>\psi^S \ln \frac{<A_iB_j>\psi^S}{m_in_j} \ , \tag{2.10}$$

where m_i and n_j are the multiplicities of the eigenvalues λ_i and μ_j. The marginal information quantities are given by:

$$(2.11) \qquad I_A = \sum_i <A_i I^2>_\psi^S \ln \frac{<A_i I^2>_\psi^S}{m_i} ,$$

$$I_B = \sum_j <I^1 B_j>_\psi^S \ln \frac{<I^1 B_j>_\psi^S}{n_j} ,$$

and finally the correlation, $\{A,B\}_\psi^S$ is given by:

$$(2.12) \quad \{A,B\}_\psi^S = \sum_{ij} P_{ij} \ln \frac{P_{ij}}{P_i P_j} = \sum_{ij} <A_i B_j>_\psi^S \ln \frac{<A_i B_j>_\psi^S}{<A_i I>_\psi^S <I B_j>_\psi^S} ,$$

where we note that the expression does not involve the multiplicities, as do the information expressions, a circumstance which simply reflects the independence of correlation on any information measure. These expressions of course generalize trivially to distributions over more than two variables (composite systems of more than two subsystems).

In addition to the correlation of pairs of subsystem operators, given by (2.12), there always exists a unique quantity $\{S_1, S_2\}$, the *canonical correlation*, which has some special properties and may be regarded as the fundamental correlation between the two subsystems S_1 and S_2 of the composite system S. As we remarked earlier a density matrix is Hermitian, so that there is a representation in which it is diagonal.[6] In

---

[6] The density matrix of a subsystem always has a pure discrete spectrum, if the composite system is in a state. To see this we note that the choice of any orthonormal basis in S_2 leads to a discrete (i.e., denumerable) set of relative states in S_1. The density matrix in S_1 then represents *this* discrete mixture, $\psi_{rel}^{\theta_j}$ weighted by P_j. This means that the expectation of the identity, $\mathrm{Exp}[I] = \Sigma_j P_j(\psi_{rel}^{\theta_j}, I\psi_{rel}^{\theta_j}) = \Sigma_j P_j = 1 = \mathrm{Trace}(\rho I) = \mathrm{Trace}(\rho)$. Therefore ρ has a finite trace and is a completely continuous operator, having necessarily a pure discrete spectrum. (See von Neumann [17], p. 89, footnote 115.)

particular, for the decomposition of S (with state ψ^S) into S_1 and S_2, we can choose a representation in which both ρ^{S_1} and ρ^{S_2} are diagonal. (This choice is always possible because ρ^{S_1} is independent of the basis in S_2 and vice-versa.) Such a representation will be called a *canonical representation.* This means that it is always possible to represent the state ψ^S by a *single* superposition:

$$\psi^S = \sum_i a_i \xi_i \eta_i \ , \tag{2.13}$$

where *both* the $\{\xi_i\}$ and the $\{\eta_i\}$ constitute orthonormal sets of states for S_1 and S_2 respectively.

To construct such a representation choose the basis $\{\eta_i\}$ for S_2 so that ρ^{S_2} is diagonal:

$$\rho^{S_2}_{ij} = \lambda_i \delta_{ij} \ , \tag{2.14}$$

and let the ξ_i be the *relative* states in S_1 for the η_i in S_2:

$$\xi_i = N_i \sum_j (\phi_j \eta_i, \psi^S) \phi_j \quad \text{(any basis } \{\phi_j\}) \ . \tag{2.15}$$

Then, according to (1.13), ψ^S is represented in the form (2.13) where the $\{\eta_i\}$ are orthonormal by choice, and the $\{\xi_i\}$ are normal since they are relative states. We therefore need only show that the states $\{\xi_i\}$ are orthogonal:

$$\begin{aligned}(\xi_j, \xi_k) &= \Big(N_j \sum_\ell (\phi_\ell \eta_j, \psi^S) \phi_\ell, \ N_k \sum_m (\phi_m \eta_k, \psi^S) \phi_m\Big) \\ &= \sum_{\ell m} N_j^* N_k (\phi_\ell \eta_j, \psi^S)^* (\phi_m \eta_k, \psi^S) \delta_{\ell m} \\ &= N_j^* N_k \sum_\ell (\phi_\ell \eta_j, \psi^S)^* (\phi_\ell \eta_k, \psi^S) \\ &= N_j^* N_k \rho^{S_2}_{kj} = N_j^* N_k \lambda_k \delta_{kj} = 0 \ , \quad \text{for } j \neq k \ ,\end{aligned} \tag{2.16}$$

since we supposed ρ^{S_2} to be diagonal in this representation. We have therefore constructed a canonical representation (2.13).

The density matrix ρ^{S_1} is also automatically diagonal, by the choice of representation consisting of the basis in S_2 which makes ρ^{S_2} diagonal and the corresponding relative states in S_1. Since $\{\xi_i\}$ are orthonormal we have:

$$(2.17)\qquad \rho^{S_1} = \sum_k (\xi_i \eta_k, \psi^S)^* (\xi_j \eta_k, \psi^S) =$$

$$\sum_k \left(\xi_i \eta_k, \sum_m a_m \xi_m \eta_m\right)^* \left(\xi_j \eta_k, \sum_\ell a_\ell \xi_\ell \eta_\ell\right)$$

$$= \sum_{k\ell m} a_m^* a_\ell \delta_{im} \delta_{km} \delta_{j\ell} \delta_{k\ell} = \sum_k a_i^* a_j \delta_{ki} \delta_{kj}$$

$$= a_i^* a_i \delta_{ij} = P_i \delta_{ij} ,$$

where $P_i = a_i^* a_i$ is the marginal distribution over the $\{\xi_i\}$. Similar computation shows that the elements of ρ^{S_2} are the *same*:

$$(2.18)\qquad \rho^{S_2}_{k\ell} = a_k^* a_k \delta_{k\ell} = P_k \delta_{k\ell} .$$

Thus in the canonical representation both density matrices are diagonal and have the same elements, P_k, which give the marginal square amplitude distribution over both of the sets $\{\xi_i\}$ and $\{\eta_i\}$ forming the basis of the representation.

Now, any pair of operators, $\tilde{A}$ in S_1 and $\tilde{B}$ in S_2, which have as non-degenerate eigenfunctions the sets $\{\xi_i\}$ and $\{\eta_j\}$ (i.e., operators which define the canonical representation), are "perfectly" correlated in the sense that there is a one-one correspondence between their eigenvalues. The joint square amplitude distribution for eigenvalues λ_i of $\tilde{A}$ and μ_j of $\tilde{B}$ is:

$$(2.19)\qquad P(\lambda_i \text{ and } \mu_j) = P(\xi_i \text{ and } \eta_j) = P_{ij} = a_i^* a_i \delta_{ij} = P_i \delta_{ij} .$$

Therefore, the correlation between these operators, $\{\tilde{A},\tilde{B}\}\psi^S$ is:

$$(2.20)\quad \{\tilde{A},\tilde{B}\}\psi^S = \sum_{ij} P(\lambda_i \text{ and } \mu_j) \ln \frac{P(\lambda_i \& \mu_j)}{P(\lambda_i)P(\mu_j)} = \sum_{ij} P_i \delta_{ij} \ln \frac{P_i \delta_{ij}}{P_i P_j}$$

$$= -\sum_i P_i \ln P_i \ .$$

We shall denote this quantity by $\{S_1, S_2\}\psi^S$ and call it the *canonical correlation* of the subsystems S_1 and S_2 for the system state ψ^S. It is the correlation between any pair of non-degenerate subsystem operators which define the canonical representation.

In the canonical representation, where the density matrices are diagonal ((2.17) and (2.18)), the canonical correlation is given by:

$$(2.21)\quad \{S_1,S_2\}\psi^S = -\sum_i P_i \ln P_i = -\text{Trace}(\rho^{S_1} \ln \rho^{S_1})$$

$$= -\text{Trace}(\rho^{S_2} \ln \rho^{S_2}) \ .$$

But the trace is invariant for unitary transformations, so that (2.21) holds independently of the representation, and we have therefore established the *uniqueness* of $\{S_1,S_2\}\psi^S$.

It is also interesting to note that the quantity $-\text{Trace}(\rho \ln \rho)$ is (apart from a factor of Boltzman's constant) just the *entropy* of a mixture of states characterized by the density matrix ρ.[7] Therefore the entropy of the mixture characteristic of a subsystem S_1 for the state $\psi^S = \psi^{S_1+S_2}$ is exactly matched by a correlation information $\{S_1,S_2\}$, which represents the correlation between any pair of operators $\tilde{A}$, $\tilde{B}$, which define the canonical representation. The situation is thus quite similar to that of classical mechanics.[8]

---

7 See von Neumann [17], p. 296.

8 Cf. Chapter II, §7.

Another special property of the canonical representation is that any operators $\widetilde{A}$, $\widetilde{B}$ defining a canonical representation have *maximum marginal information*, in the sense that for any other discrete spectrum operators, A on S_1, B on S_2, $I_A \leqq I_{\widetilde{A}}$ and $I_B \leqq I_{\widetilde{B}}$. If the canonical representation is (2.13), with $\{\xi_i\}$, $\{\eta_i\}$ non-degenerate eigenfunctions of $\widetilde{A}$, $\widetilde{B}$, respectively, and A, B any pair of non-degenerate operators with eigenfunctions $\{\phi_k\}$ and $\{\theta_\ell\}$, where $\xi_i = \sum_k c_{ik}\phi_k$, $\eta_i = \sum_\ell d_{i\ell}\theta_\ell$, then ψ^S in ϕ,θ representation is:

$$\psi^S = \sum_{ik\ell} a_i c_{ik} d_{i\ell} \phi_k \theta_\ell = \sum_{k\ell} \left(\sum_i a_i c_{ik} d_{i\ell} \right) \phi_k \theta_\ell \ , \tag{2.22}$$

and the joint square amplitude distribution for ϕ_k, θ_ℓ is:

$$P_{k\ell} = \left| \left(\sum_i a_i c_{ik} d_{i\ell} \right) \right|^2 = \sum_{im} a_i^* a_m c_{ik}^* c_{mk} d_{i\ell}^* d_{m\ell} \ , \tag{2.23}$$

while the marginals are:

$$\begin{aligned} P_k &= \sum_\ell P_{k\ell} = \sum_{im} a_i^* a_m c_{ik}^* c_{mk} \sum_\ell d_{i\ell}^* d_{m\ell} \\ &= \sum_{im} a_i^* a_m c_{ik}^* c_{mk} \delta_{im} = \sum_i a_i^* a_i c_{ik}^* c_{ik} \ , \end{aligned} \tag{2.24}$$

and similarly

$$P_\ell = \sum_k P_{k\ell} = \sum_i a_i^* a_i d_{i\ell}^* d_{i\ell} \ . \tag{2.25}$$

Then the marginal information I_A is:

$$\begin{aligned} I_A &= \sum_k P_k \ln P_k = \sum_k \left(\sum_i a_i^* a_i c_{ik}^* c_{ik} \right) \ln \left(\sum_i a_i^* a_i c_{ik}^* c_{ik} \right) \\ &= \sum_k \left(\sum_i a_i^* a_i T_{ik} \right) \ln \left(\sum_i a_i^* a_i T_{ik} \right) \ , \end{aligned} \tag{2.26}$$

where $T_{ik} = c_{ik}^* c_{ik}$ is doubly-stochastic ($\sum_i T_{ik} = \sum_k T_{ik} = 1$ follows from unitary nature of the c_{ik}). Therefore (by Corollary 2, §4, Appendix I):

$$(2.27) \qquad I_A = \sum_k \left(\sum_i a_i^* a_i T_{ik} \right) \ln \left(\sum_i a_i^* a_i T_{ik} \right)$$

$$\leqq \sum_i a_i^* a_i \ln a_i^* a_i = I_{\tilde{A}} ,$$

and we have proved that $\tilde{A}$ has maximal marginal information among the discrete spectrum operators. Identical proof holds for $\tilde{B}$.

While this result was proved only for non-degenerate operators, it is immediately extended to the degenerate case, since as a consequence of our definition of information for a degenerate operator, (2.4), its information is still less than that of an operator which removes the degeneracy. We have thus proved:

THEOREM. $I_A \leqq I_{\tilde{A}}$, *where* $\tilde{A}$ *is any non-degenerate operator defining the canonical representation, and* A *is any operator with discrete spectrum.*

We conclude the discussion of the canonical representation by conjecturing that in addition to the maximum marginal information properties of $\tilde{A}$, $\tilde{B}$, which define the representation, they are also *maximally correlated,* by which we mean that for any pair of operators C in S_1, D in S_2, $\{C,D\} \leqq \{\tilde{A},\tilde{B}\}$, i.e.,:

$$(2.28) \qquad \text{CONJECTURE.}^9 \quad \{C,D\}\psi^S \leqq \{\tilde{A},\tilde{B}\}\psi^S = \{S_1,S_2\}\psi^S$$

for *all* C on S_1, D on S_2.

As a final topic for this section we point out that the uncertainty principle can probably be phrased in a stronger form in terms of information. The usual form of this principle is stated in terms of *variances,* namely:

---

9 The relations $\{C,\tilde{B}\} \leqq \{\tilde{A},\tilde{B}\} = \{S_1,S_2\}$ and $\{\tilde{A},D\} \leqq \{S_1,S_2\}$ for all C on S_1, D on S_2, can be proved easily in a manner analogous to (2.27). These do not, however, necessarily imply the general relation (2.28).

(2.29) $$\sigma_x^2 \sigma_k^2 \geqq \frac{1}{4} \qquad \text{for all } \psi(x) ,$$

$$\text{where} \quad \sigma_x^2 = \langle x^2 \rangle_\psi - [\langle x \rangle_\psi]^2 \quad \text{and}$$

$$\sigma_k^2 = \left\langle \left(-i \frac{\partial}{\partial x}\right)^2 \right\rangle_\psi - \left[\left\langle -i \frac{\partial}{\partial x} \right\rangle_\psi\right]^2 = \left\langle \left(\frac{P}{\hbar}\right)^2 \right\rangle_\psi - \left[\left\langle \frac{P}{\hbar} \right\rangle_\psi\right]^2 .$$

The conjectured information form of this principle is:

(2.30) $$I_x + I_k \leqq \ln (1/\pi e) \qquad \text{for all } \psi(x).$$

Although this inequality has not yet been proved with complete rigor, it is made highly probable by the circumstance that *equality* holds for $\psi(x)$ of the form $\psi(x) = (1/2\pi)^{\frac{1}{4}}$ exponent $\left[\frac{x^2}{4\sigma_x^2}\right]$ the so called "minimum uncertainty packets" which give normal distributions for both position and momentum, and that furthermore the first variation of $(I_x + I_k)$ vanishes for such $\psi(x)$. (See Appendix I, §6.) Thus, although $\ln (1/\pi e)$ has not been proved an absolute maximum of $I_x + I_k$, it is at least a stationary value.

The principle (2.30) is *stronger* than (2.29), since it implies (2.29) but is not implied by it. To see that it implies (2.29) we use the well known fact (easily established by a variation calculation: that, for fixed variance σ^2, the distribution of minimum information is a normal distribution, which has information $I = \ln (1/\sigma\sqrt{2\pi e})$. This gives us the general inequality involving information and variance:

(2.31) $$I \geqq \ln (1/\sigma\sqrt{2\pi e}) \qquad \text{(for all distributions)} .$$

Substitution of (2.31) into (2.30) then yields:

(2.32) $$\ln (1/\sigma_x\sqrt{2\pi e}) + \ln (1/\sigma_k\sqrt{2\pi e}) \leqq I_x + I_k \leqq \ln (1/\pi e)$$

$$\Rightarrow (1/\sigma_x\sigma_k 2\pi e) \leqq (1/\pi e) \Rightarrow \sigma_x^2\sigma_k^2 \geqq \frac{1}{4} ,$$

so that our principle implies the standard principle (2.29).

To show that (2.29) does *not imply* (2.30) it suffices to give a counter-example. The distributions $P(x) = \frac{1}{2}\delta(x) + \frac{1}{2}\delta(x-10)$ and $P(k) = \frac{1}{2}\delta(k) + \frac{1}{2}\delta(k-10)$, which consist simply of spikes at 0 and 10, clearly satisfy (2.29), while they both have infinite information and thus do *not* satisfy (2.30). Therefore it is possible to have arbitrarily high information about *both* x and k (or p) and still satisfy (2.13). We have, then, another illustration that information concepts are more powerful and more natural than the older measures based upon variance.

§3. *Measurement*

We now consider the question of measurement in quantum mechanics, which we desire to treat as a natural process within the theory of pure wave mechanics. From our point of view there is no fundamental distinction between "measuring apparata" and other physical systems. For us, therefore, a measurement is simply a special case of interaction between physical systems – an interaction which has the property of *correlating* a quantity in one subsystem with a quantity in another.

Nearly every interaction between systems produces *some* correlation however. Suppose that at some instant a pair of systems are independent, so that the composite system state function is a product of subsystem states ($\psi^S = \psi^{S_1}\psi^{S_2}$). Then this condition obviously holds only instantaneously if the systems are interacting[10]– the independence is immediately destroyed and the systems become correlated. We could, then, take the position that the two interacting systems are continually "measuring" one another, if we wished. At each instant t we could put the composite system into canonical representation, and choose a pair of operators $\tilde{A}(t)$

[10] If U_t^S is the unitary operator generating the time dependence for the state function of the composite system $S = S_1 + S_2$, so that $\psi_t^S = U_t^S \psi_0^S$, then we shall say that S_1 and S_2 have not interacted during the time interval $[0,t]$ if and only if U_t^S is the direct product of two subsystem unitary operators, i.e., if $U_t^S = U_t^{S_1} \otimes U_t^{S_2}$.

in S_1 and $\tilde{B}(t)$ in S_2 which define this representation. We might then reasonably assert that the quantity $\tilde{A}$ in S_1 is measured by $\tilde{B}$ in S_2 (or vice-versa), since there is a one-one correspondence between their values.

Such a viewpoint, however, does not correspond closely with our intuitive idea of what constitutes "measurement," since the quantities $\tilde{A}$ and $\tilde{B}$ which turn out to be measured depend not only on the time, but also upon the initial state of the composite system. A more reasonable position is to associate the term "measurement" with a fixed interaction H between systems,[11] and to define the "measured quantities" not as those quantities $\tilde{A}(t)$, $\tilde{B}(t)$ which are instantaneously canonically correlated, but as the limit of the instantaneous canonical operators as the time goes to infinity, $\tilde{A}_\infty$, $\tilde{B}_\infty$ – provided that this limit exists and is independent of the initial state.[12] In such a case we are able to associate the "measured quantities," $\tilde{A}_\infty$, $\tilde{B}_\infty$, with the interaction H independently of the actual system states and the time. We can therefore say that H is an interaction which causes the quantity $\tilde{A}_\infty$ in S_1 to be measured by $\tilde{B}_\infty$ in S_2. For finite times of interaction the measurement is only approximate, approaching exactness as the time of interaction increases indefinitely.

There is still one more requirement that we must impose on an interaction before we shall call it a measurement. If H is to produce a measurement of A in S_1 by B in S_2, then we require that H shall

---

[11] Here H means the *total* Hamiltonian of S, not just an interaction part.

[12] Actually, rather than referring to canonical operators $\tilde{A}$, $\tilde{B}$, which are not unique, we should refer to the *bases* of the canonical representation, $\{\xi_i\}$ in S_1 and $\{\eta_j\}$ in S_2, since *any* operators $\tilde{A} = \Sigma_i \lambda_i [\xi_i]$, $\tilde{B} = \Sigma_j \mu_j [\eta_j]$, with the completely arbitrary eigenvalues λ_i, μ_j, are canonical. The limit then refers to the limit of the canonical bases, if it exists in some appropriate sense. However, we shall, for convenience, continue to represent the canonical bases by operators.

never decrease the information in the marginal distribution of A. If H is to produce a measurement of A by correlating it with B, we expect that a knowledge of B shall give us more information about A than we had before the measurement took place, since otherwise the measurement would be useless. Now, H might produce a correlation between A and B by simply destroying the marginal information of A, without improving the expected conditional information of A given B, so that a knowledge of B would give us no more information about A than we possessed originally. Therefore in order to be sure that we will gain information about A by knowing B, when B has become correlated with A, it is necessary that the marginal information about A has not decreased. The expected information gain in this case is assured to be not less than the correlation $\{A,B\}$.

The restriction that H shall not decrease the marginal information of A has the interesting consequence that the eigenstates of A will not be distrubed, i.e., initial states of the form $\psi_0^S = \phi\,\eta_0$, where ϕ is an *eigenfunction of A, must be transformed after any time interval into* states of the form $\psi_t^S = \phi\,\eta_t$, since otherwise the marginal information of A, which was initially perfect, would be decreased. This condition, in turn, is connected with the *repeatability* of measurements, as we shall subsequently see, and could alternately have been chosen as the condition for measurement.

We shall therefore accept the following definition. An interaction H is a measurement of A in S_1 by B in S_2 if H does not destroy the marginal information of A (equivalently: if H does not disturb the eigenstates of A in the above sense) and if furthermore the correlation $\{A,B\}$ increases toward its maximum[13] with time.

[13] The maximum of $\{A,B\}$ is $-I_A$ if A has only a discrete spectrum, and ∞ if it has a continuous spectrum.

We now illustrate the production of correlation with an example of a simplified measurement due to von Neumann.[14] Suppose that we have a system of only one coordinate, q, (such as position of a particle), and an apparatus of one coordinate r (for example the position of a meter needle). Further suppose that they are initially independent, so that the combined wave function is $\psi_0^{S+A} = \phi(q)\eta(r)$, where $\phi(q)$ is the initial system wave function, and $\eta(r)$ is the initial apparatus function. Finally suppose that the masses are sufficiently large or the time of interaction sufficiently small that the kinetic portion of the energy may be neglected, so that during the time of measurement the Hamiltonian shall consist only of an interaction, which we shall take to be:

$$H_I = -i\hbar q \frac{\partial}{\partial r} . \tag{3.1}$$

Then it is easily verified that the state $\psi_t^{S+A}(q,r)$:

$$\psi_t^{S+A}(q,r) = \phi(q)\eta(r-qt) . \tag{3.2}$$

is a solution of the Schrödinger equation

$$i\hbar \frac{\partial \psi_t^{S+A}}{\partial t} = H_I \psi_t^{S+A} \tag{3.3}$$

for the specified initial conditions at time $t = 0$.

Translating (3.2) into square amplitudes we get:

$$P_t(q,r) = P_1(q)P_2(r-qt) , \tag{3.4}$$

where $$P_1(q) = \phi^*(q)\phi(q) , \quad P_2(r) = \eta^*(r)\eta(r) ,$$

and $$P_t(q,r) = \psi_t^{S+A^*}(q,r)\psi_t^{S+A}(q,r) ,$$

---

[14] von Neumann [17], p. 442.

and we note that for a fixed time, t, the conditional square amplitude distribution for r has been translated by an amount depending upon the value of q, while the marginal distribution for q has been unaltered. We see thus that a correlation has been introduced between q and r by this interaction, which allows us to interpret it as a measurement. It is instructive to see quantitatively how fast this correlation takes place. We note that:

$$(3.5)\qquad I_{QR}(t) = \iint P_t(q,r)\,\ln P_t(q,r)\,dqdr$$

$$= \iint P_1(q)P_2(r-qt)\,\ln P_1(q)P_2(r-qt)\,dqdr$$

$$= \iint P_1(q)P_2(\omega)\,\ln P_1(q)P_2(\omega)\,dqd\omega$$

$$= I_{QR}(0)\ ,$$

so that the information of the joint distribution does not change. Furthermore, since the marginal distribution for q is unchanged:

$$(3.6)\qquad I_Q(t) = I_Q(0)\ ,$$

and the only quantity which can change is the marginal information, I_R, of r, whose distribution is:

$$(3.7)\qquad P_t(r) = \int P_t(r,q)\,dq = \int P_1(q)P_2(r-qt)\,dq\ .$$

Application of a special inequality (proved in §5, Appendix I) to (3.7) yields the relation:

$$(3.8)\qquad I_R(t) \leqq I_Q(0) - \ln t\ ,$$

so that, except for the additive constant $I_Q(0)$, the marginal information I_R tends to decrease at least as fast as $\ln t$ with time during the interaction. This implies the relation for the correlation:

$$(3.9) \qquad \{Q,R\}_t = I_{QR}(t) - I_Q(t) - I_R(t) \geqq I_{RQ}(t) - I_Q(t) - I_Q(0) + \ln t \ .$$

But at $t = 0$ the distributions for R and Q were independent, so that $I_{RQ}(0) = I_R(0) + I_Q(0)$. Substitution of this relation, (3.5), and (3.6) into (3.9) then yields the final result:

$$(3.10) \qquad \{Q,R\}_t \geqq I_R(0) - I_Q(0) + \ln t \ .$$

Therefore the correlation is built up at least as fast as ln t, except for an additive constant representing the difference of the information of the initial distributions $P_2(r)$ and $P_1(q)$. Since the correlation goes to infinity with increasing time, and the marginal system distribution is not changed, the interaction (3.1) satisfies our definition of a measurement of q by r.

Even though the apparatus does not indicate any definite system value (since there are no independent system or apparatus states), one can nevertheless look upon the total wave function (3.2) as a *superposition* of pairs of subsystem states, each element of which has a definite q value and a correspondingly displaced apparatus state.[15] Thus we can write (3.2) as:

$$(3.11) \qquad \psi_t^{S+A} = \int \phi(q')\,\delta(q-q')\,\eta(r-q't)\,dq' \ ,$$

which is a superposition of states $\psi_{q'} = \delta(q-q')\,\eta(r-q't)$. Each of these elements, $\psi_{q'}$, of the superposition describes a state in which the system has the definite value $q = q'$, and in which the apparatus has a state that is displaced from its original state by the amount $q't$. These elements $\psi_{q'}$ are then superposed with coefficients $\phi(q')$ to form the total state (3.11).

---

15 See discussion of relative states, p. 38.

Conversely, if we transform to the representation where the *apparatus* is definite, we write (3.2) as:

$$\psi_t^{S+A} = \int (1/N_{r'})\,\xi^{r'}(q)\,\delta(r-r')\,dr' , \tag{3.12}$$

where $$\xi^{r'}(q) = N_{r'}\,\phi(q)\,\eta(r'-qt)$$

and $$(1/N_{r'})^2 = \int \phi^*(q)\,\phi(q)\,\eta^*(r'-qt)\,\eta(r-qt)\,dq .$$

Then the $\xi^{r'}(q)$ are the relative system state functions for the apparatus states $\delta(r-r')$ of definite value $r = r'$.

We notice that these relative system states, $\xi^{r'}(q)$, are nearly eigenstates for the values $q = r'/t$, if the degree of correlation between q and r is sufficiently high, i.e., if t is sufficiently large, or $\eta(r)$ sufficiently sharp (near $\delta(r)$) then $\xi^{r'}(q)$ is nearly $\delta(q-r'/t)$.

This property, that the relative system states become approximate eigenstates of the measurement, is in fact common to all measurements. If we adopt as a measure of the nearness of a state ψ to being an eigenfunction of an operator A the information $I_A(\psi)$, which is reasonable because $I_A(\psi)$ measures the sharpness of the distribution of A for ψ, then it is a consequence of our definition of a measurement that the relative system states tend to become eigenstates as the interaction proceeds. Since $\mathrm{Exp}[I_Q^r] = I_Q + \{Q,R\}$, and I_Q remains constant while $\{Q,R\}$ tends toward its maximum (or infinity) during the interaction, we have that $\mathrm{Exp}[I_Q^r]$ tends to a maximum (or infinity). But I_Q^r is just the information in the relative system states, which we have adopted as a measure of the nearness to an eigenstate. Therefore, at least in expectation, the relative system states approach eigenstates.

We have seen that (3.12) is a superposition of states $\psi_{r'}$, *for each of which* the apparatus has recorded a definite value r', and the system is left in approximately the eigenstate of the measurement corresponding to $q = r'/t$. The discontinuous "jump" into an eigenstate is thus only a

relative proposition, dependent upon our decomposition of the total wave function into the superposition, and relative to a particularly chosen apparatus value. So far as the complete theory is concerned all elements of the superposition exist simultaneously, and the entire process is quite continuous.

We have here only a special case of the following general principle which will hold for any situation which is treated entirely wave mechanically:

PRINCIPLE. For any situation in which the existence of a property R_i for a subsystem S_1 of a composite system S will imply the later property Q_i for S, then it is also true that an initial state for S_1 of the form $\psi^{S_1} = \sum a_i \psi^{S_1}_{[R_i]}$ which is a *superposition* of states with the properties R_i, will result in a later state for S of the form $\psi^S = \sum_i a_i \psi^S_{[Q_i]}$, which is *also a superposition*, of states with the property Q_i. That is, for any arrangement of an interaction between two systems S_1 and S_2, which has the property that each initial state $\phi_i^{S_1} \psi^{S_2}$ will result in a final situation with total state $\psi_i^{S_1+S_2}$, an initial state of S_1 of the form $\sum_i a_i \phi_i^{S_1}$ will lead, after interaction, to the superposition $\sum_i a_i \psi_i^{S_1+S_2}$ for the whole system.

This follows immediately from the superposition principle for solutions of a linear wave equation. It therefore holds for *any* system of quantum mechanics for which the superposition principle holds, both particle and field theories, relativistic or not, and is applicable to all physical systems, regardless of size.

This principle has the far reaching implication that for any possible measurement, for which the initial system state is not an eigenstate, the resulting state of the composite system leads to *no* definite system state nor any definite apparatus state. The system will not be put into one or another of its eigenstates with the apparatus indicating the corresponding value, and nothing resembling Process 1 can take place.

To see that this is indeed the case, suppose that we have a measuring arrangement with the following properties. The initial apparatus state is ψ_0^A. If the system is initially in an eigenstate of the measurement, ϕ_i^S, then after a specified time of interaction the total state $\phi_i^S\psi_0^A$ will be transformed into a state $\phi_i^S\psi_i^A$, i.e., the system eigenstate shall not be disturbed, and the apparatus state is changed to ψ_i^A, which is different for each ϕ_i^S. (ψ_i^A may for example be a state describing the apparatus as indicating, by the position of a meter needle, the eigenvalue of ϕ_i^S.) However, if the initial system state is *not an eigenstate* but a superposition $\sum_i a_i\phi_i^S$, then the final composite system state is *also a superposition*, $\sum_i a_i\phi_i^S\psi_i^\lambda$. This follows from the superposition principle since all we need do is superpose our solutions for the eigenstates, $\phi_i^S\psi_0^A \rightarrow \phi_i^S\psi_i^A$, to arrive at the solution, $\sum_i a_i\phi_i^S\psi_0^A \rightarrow \sum_i a_i\phi_i^S\psi_i^A$, for the general case. Thus in general after a measurement has been performed there will be no definite system state nor any definite apparatus state, even though there is a correlation. It seems as though nothing can ever be settled by such a measurement. Furthermore this result is independent of the *size* of the apparatus, and remains true for apparatus of quite macroscopic dimensions.

Suppose, for example, that we coupled a spin measuring device to a cannonball, so that if the spin is up the cannonball will be shifted one foot to the left, while if the spin is down it will be shifted an equal distance to the right. If we now perform a measurement with this arrangement upon a particle whose spin is a superposition of up and down, then the resulting total state will also be a superposition of two states, one in which the cannonball is to the left, and one in which it is to the right. There is no definite position for our macroscopic cannonball!

This behavior seems to be quite at variance with our observations, since macroscopic objects always appear to us to have definite positions. Can we reconcile this prediction of the purely wave mechanical theory

with experience, or must we abandon it as untenable? In order to answer this question we must consider the problem of observation itself within the framework of the theory.

IV. OBSERVATION

We shall now give an abstract treatment of the problem of observation. In keeping with the spirit of our investigation of the consequences of pure wave mechanics we have no alternative but to introduce observers, considered as purely physical systems, into the theory.

We saw in the last chapter that in general a measurement (coupling of system and apparatus) had the outcome that neither the system nor the apparatus had any definite state after the interaction – a result seemingly at variance with our experience. However, we do not do justice to the theory of pure wave mechanics until we have investigated what the theory itself says about the *appearance* of phenomena to observers, rather than hastily concluding that the theory must be incorrect because the actual states of systems as given by the theory seem to contradict our observations.

We shall see that the introduction of observers can be accomplished in a reasonable manner, and that the theory then predicts that the *appearance* of phenomena, as the subjective experience of these observers, is precisely in accordance with the predictions of the usual probabilistic interpretation of quantum mechanics.

§1. *Formulation of the problem*

We are faced with the task of making deductions about the appearance of phenomena on a subjective level, to observers which are considered as purely physical systems and are treated within the theory. In order to accomplish this it is necessary to identify some objective properties of such an observer (states) with subjective knowledge (i.e., perceptions). Thus, in order to say that an observer O has observed the event α, it

is necessary that the state of O has become changed from its former state to a new state which is dependent upon α.

It will suffice for our purposes to consider our observers to possess memories (i.e., parts of a relatively permanent nature whose states are in correspondence with the past experience of the observer). In order to make deductions about the subjective experience of an observer it is sufficient to examine the contents of the memory.

As models for observers we can, if we wish, consider automatically functioning machines, possessing sensory apparata and coupled to recording devices capable of registering past sensory data and machine configurations. We can further suppose that the machine is so constructed that its present actions shall be determined not only by its present sensory data, but by the contents of its memory as well. Such a machine will then be capable of performing a sequence of observations (measurements), and furthermore of deciding upon its future experiments on the basis of past results. We note that if we consider that current sensory data, as well as machine configuration, is immediately recorded in the memory, then the actions of the machine at a given instant can be regarded as a function of the memory contents only, and all relevant experience of the machine is contained in the memory.

For such machines we are justified in using such phrases as "the machine has perceived A" or "the machine is aware of A" if the occurrence of A is represented in the memory, since the future behavior of the machine will be based upon the occurrence of A. In fact, all of the customary language of subjective experience is quite applicable to such machines, and forms the most natural and useful mode of expression when dealing with their behavior, as is well known to individuals who work with complex automata.

When dealing quantum mechanically with a system representing an observer we shall ascribe a state function, ψ^O, to it. When the State ψ^O describes an observer whose memory contains representations of the

events A,B,...,C we shall denote this fact by appending the memory sequence in brackets as a subscript, writing:

$$\psi^{O}_{[A,B,\ldots,C]} .$$

The symbols A,B,...,C, which we shall assume to be ordered time wise, shall therefore stand for memory configurations which are in correspondence with the past experience of the observer. These configurations can be thought of as punches in a paper tape, impressions on a magnetic reel, configurations of a relay switching circuit, or even configurations of brain cells. We only require that they be capable of the interpretation "The observer has experienced the succession of events A,B,...,C." (We shall sometimes write dots in a memory sequence, [...A,B,...,C], to indicate the possible presence of previous memories which are irrelevant to the case being considered.)

Our problem is, then, to treat the interaction of such observer-systems with other physical systems (observations), within the framework of wave mechanics, and to deduce the resulting memory configurations, which we can then interpret as the subjective experiences of the observers.

We begin by defining what shall constitute a "good" observation. A good observation of a quantity A, with eigenfunctions $\{\phi_i\}$ for a system S, by an observer whose initial state is $\psi^{O}_{[\ldots]}$, shall consist of an interaction which, in a specified period of time, transforms each (total) state

$$\psi^{S+O} = \phi_i \psi^{O}_{[\ldots]}$$

into a new state

$$\psi^{S+O'} = \phi_i \psi^{O}_{i[\ldots,a_i]},$$

where a_i characterizes the state ϕ_i. (It might stand for a recording of the eigenvalue, for example.) That is, our requirement is that the system state, *if it is an eigenstate*, shall be unchanged, and that the observer

state shall change so as to describe an observer that is "aware" of which eigenfunction it is, i.e., some property is recorded in the memory of the observer which characterizes ϕ_i, such as the eigenvalue. The requirement that the eigenstates for the system be unchanged is necessary if the observation is to be significant (repeatable), and the requirement that the observer state change in a manner which is different for each eigenfunction is necessary if we are to be able to call the interaction an observation at all.

§2. *Deductions*

From these requirements we shall first deduce the result of an observation upon a system which is *not* in an eigenstate of the observation. We know, by our previous remark upon what constitutes a good observation that the interaction transforms states $\phi_i\psi^O_{[\ldots]}$ into states $\phi_i\psi^O_{i[\ldots,a_i]}$. Consequently we can simply superpose these solutions of the wave equation to arrive at the final state for the case of an arbitrary initial system state. Thus if the initial system state is not an eigenstate, but a general state $\sum_i a_i\phi_i$, we get for the final total state:

$$\psi^{S+O} = \sum_i a_i\phi_i\psi^O_{i[\ldots,a_i]}. \tag{2.1}$$

This remains true also in the presence of further systems which do not interact for the time of measurement. Thus, if systems $S_1,S_2,\ldots,S_n$ are present as well as O, with original states $\psi^{S_1},\psi^{S_2},\ldots,\psi^{S_n}$, and the only interaction during the time of measurement is between S_1 and O, the result of the measurement will be the transformation of the initial total state:

$$\psi^{S_1+S_2+\ldots+S_n+O} = \psi^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^O_{[\ldots]}$$

into the final state:

$$\psi'^{S_1+S_2+\ldots+S_n+O} = \sum_i a_i\phi_i^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^O_{i[\ldots,a_i]} \tag{2.2}$$

where $a_i = \left(\phi_i^{S_1}, \psi^{S_1}\right)$ and $\phi_i^{S_1}$ are eigenfunctions of the observation.

Thus we arrive at the general rule for the transformation of total state functions which describe systems within which observation processes occur:

Rule 1. The observation of a quantity A, with eigenfunctions $\phi_i^{S_1}$, in a system S_1 by the observer O, transforms the total state according to:

$$\psi^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^{O}_{[\ldots]} \rightarrow \sum_i a_i\phi_i^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^{O}_{i[\ldots,a_i]},$$

where $a_i = \left(\phi_i^{S_1}, \psi^{S_1}\right)$.

If we next consider a *second* observation to be made, where our total state is now a superposition, we can apply *Rule* 1 separately to each element of the superposition, since each element separately obeys the wave equation and behaves independently of the remaining elements, and then superpose the results to obtain the final solution. We formulate this as:

Rule 2. *Rule* 1 may be applied separately to each element of a superposition of total system states, the results being superposed to obtain the final total state. Thus, a determination of B, with eigenfunctions $\eta_j^{S_2}$, on S_2 by the observer O transforms the total state

$$\sum_i a_i\phi_i^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^{O}_{i[\ldots,a_i]}$$

into the state

$$\sum_{i,j} a_i b_j \phi_i^{S_1}\eta_j^{S_2}\psi^{S_3}\ldots\psi^{S_n}\psi^{O}_{ij[\ldots,a_i,\beta_j]}$$

where $b_j = \left(\eta_j^{S_2}, \psi^{S_2}\right)$, which follows from the application of *Rule* 1 to each element $\phi_i^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^{O}_{i[\ldots,a_i]}$, and then superposing the results with the coefficients a_i.

These two rules, which follow directly from the superposition principle, give us a convenient method for determining final total states for any number of observation processes in any combinations. We must now seek the interpretation of such final total states.

Let us consider the simple case of a single observation of a quantity A, with eigenfunctions ϕ_i, in the system S with initial state ψ^S, by an observer O whose initial state is $\psi^O_{[\ldots]}$. The final result is, as we have seen, the superposition:

$$\psi' {}^{S+O} = \sum_i a_i \phi_i \psi^O_{i[\ldots,a_i]} \,. \tag{2.3}$$

We note that there is no longer any independent system state or observer state, although the two have become correlated in a one-one manner. However, in each *element* of the superposition (2.3), $\phi_i \psi^O_{i[\ldots,a_i]}$, the object-system state is a particular eigenstate of the observer, and *furthermore the observer-system state describes the observer as definitely perceiving that particular system state.*[1] It is this correlation which allows one to maintain the interpretation that a measurement has been performed.

We now carry the discussion a step further and allow the observer-system to repeat the observation. Then according to *Rule* 2 we arrive at the total state after the second observation:

---

[1] At this point we encounter a language difficulty. Whereas before the observation we had a single observer state afterwards there were a number of different states for the observer, all occurring in a superposition. Each of these separate states is a state for an observer, so that we can speak of the different observers described by the different states. On the other hand, the same physical system is involved, and from this viewpoint it is the *same* observer, which is in different states for different elements of the superposition (i.e., has had different experiences in the separate elements of the superposition). In this situation we shall use the singular when we wish to emphasize that a single physical system is involved, and the plural when we wish to emphasize the different experiences for the separate elements of the superposition. (e.g., "The observer performs an observation of the quantity A, after which each of the observers of the resulting superposition has perceived an eigenvalue.")

(2.4) $$\psi^{\sim S+O} = \sum_i a_i \phi_i \psi^O_{ii[\ldots,a_i,a_i]} .$$

Again, we see that each element of (2.4), $\phi_i \psi^O_{ii[\ldots,a_i,a_i]}$, describes a system eigenstate, but this time also describes the observer as having obobtained the *same result* for each of the two observations. Thus for every separate state of the observer in the final superposition, the result of the observation was repeatable, even though different for different states. This repeatability is, of course, a consequence of the fact that after an observation the *relative* system state for a particular observer state is the corresponding eigenstate.

Let us suppose now that an observer-system O, with initial state $\psi^O_{[\ldots]}$, measures the *same* quantity A in a number of separate identical systems which are initially in the same state, $\psi^{S_1} = \psi^{S_2} = \ldots = \psi^{S_n} = \sum_i a_i \phi_i$ (where the ϕ_i are, as usual, eigenfunctions of A). The initial total state function is then

(2.3) $$\psi_0^{S_1+S_2+\ldots+S_n+O} = \psi^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^O_{[\ldots]} .$$

We shall assume that the measurements are performed on the systems in the order $S_1, S_2, \ldots, S_n$. Then the total state after the first measurement will be, by *Rule* 1,

(2.4) $$\psi_1^{S_1+S_2+\ldots+S_n+O} = \sum_i a_i \phi_i^{S_1} \psi^{S_2}\ldots\psi^{S_n}\psi^O_{i[\ldots,a_i^1]}$$

(where a_i^1 refers to the first system, S_i) .

After the second measurement it will be, by *Rule* 2,

(2.5) $$\psi_2^{S_1+S_2+\ldots S_n+O} = \sum_{i,j} a_i a_j \phi_i^{S_1} \phi_j^{S_2} \psi^{S_3}\ldots\psi^{S_n}\psi^O_{ij[\ldots,a_i^1,a_j^2]}$$

and in general, after r measurements have taken place ($r \leqq n$) *Rule* 2 gives the result:

$$(2.6)\quad \psi_r = \sum_{i,j,\ldots,k} a_i a_j \ldots a_k \phi_i^{S_1} \phi_j^{S_2} \ldots \phi_k^{S_r} \psi^{S_{r+1}} \ldots \psi^{S_n} \psi^{O}_{ij\ldots k[\ldots,a_i^1,a_j^2,\ldots,a_k^r]} .$$

We can give this state, ψ_r, the following interpretation. It consists of a superposition of states:

$$(2.7)\quad \psi'_{ij\ldots k} = \phi_i^{S_1} \phi_j^{S_2} \ldots \phi_k^{S_r} \psi^{S_{r+1}} \ldots \psi^{S_n} \psi^{O}_{ij\ldots k[\ldots,a_i^1,a_j^2,\ldots,a_k^r]}$$

each of which describes the observer with a definite memory sequence $[\ldots,a_i^1,a_j^2,\ldots,a_k^r]$, and relative to whom the (observed system states are the corresponding eigenfunctions $\phi_i^{S_1},\phi_j^{S_2},\ldots,\phi_k^{S_r}$, the remaining systems, S_{r+1}, $\ldots S_n$, being unaltered.

In the language of subjective experience, the observer which is described by a typical element, $\psi'_{ij\ldots k}$, of the superposition has perceived an apparently random sequence of definite results for the observations. It is furthermore true, since in each element the system has been left in an eigenstate of the measurement, that if at this stage a redetermination of an earlier system observation (S_ℓ) takes place, every element of the resulting final superposition will describe the observer with a memory configuration of the form $[\ldots,a_i^1,\ldots,a_j^\ell,\ldots,a_k^r,a_j^\ell]$ in which the earlier memory coincides with the later – i.e., the memory states are *correlated.* It will thus *appear* to the observer which is described by a typical element of the superposition that each initial observation on a system caused the system to "jump" into an eigenstate in a random fashion and thereafter remain there for subsequent measurements on the same system. Therefore, qualitatively, at least, the probabilistic assertions of Process 1 *appear* to be valid to the observer described by a typical element of the final superposition.

In order to establish quantitative results, we must put some sort of measure (weighting) on the elements of a final superposition. This is

necessary to be able to make assertions which will hold for almost all of the observers described by elements of a superposition. In order to make quantitative statements about the relative frequencies of the different possible results of observation which are recorded in the memory of a typical observer we must have a method of selecting a *typical* observer.

Let us therefore consider the search for a general scheme for assigning a measure to the elements of a superposition of orthogonal states $\sum a_i \phi_i$. We require then a positive function $\mathfrak{M}$ of the complex coefficients of the elements of the superposition, so that $\mathfrak{M}(a_i)$ shall be the measure assigned to the element ϕ_i. In order that this general scheme shall be unambiguous we must first require that the states themselves always be normalized, so that we can distinguish the coefficients from the states. However, we can still only determine the *coefficients*, in distinction to the states, up to an arbitrary phase factor, and hence the function $\mathfrak{M}$ must be a function of the amplitudes of the coefficients alone, (i.e., $\mathfrak{M}(a_i) = \mathfrak{M}(\sqrt{a_i^* a_i})$), in order to avoid ambiguities.

If we now impose the additivity requirement that if we regard a subset of the superposition, say $\sum_{i=1}^{n} a_i \phi_i$, as a single element $\alpha \phi'$:

$$\alpha \phi' = \sum_{i=1}^{n} a_i \phi_i \ , \tag{2.8}$$

then the measure assigned to ϕ' shall be the sum of the measures assigned to the ϕ_i (i from 1 to n):

$$\mathfrak{M}(\alpha) = \sum_{i} \mathfrak{M}(a_i) \ , \tag{2.9}$$

then we have already restricted the choice of $\mathfrak{M}$ to the square amplitude alone. ($\mathfrak{M}(a_i) = a_i^* a_i$), apart from a multiplicative constant.)

To see this we note that the normality of ϕ' requires that $|\alpha| = \sqrt{\sum_{i=1}^{n} a_i^* a_i}$. From our remarks upon the dependence of $\mathfrak{M}$ upon the amplitude alone, we replace the a_i by their amplitudes $\mu_i = |a_i|$.

(2.9) then requires that

$$(2.10)\qquad \mathfrak{M}(\alpha) = \mathfrak{M}\left(\sqrt{\sum a_i^* a_i}\right) = \mathfrak{M}\left(\sqrt{\sum \mu_i^2}\right) = \sum \mathfrak{M}(\mu_i) = \sum \mathfrak{M}(\sqrt{\mu_i^2}) \ .$$

Defining a new function g(x):

$$(2.11)\qquad g(x) = \mathfrak{M}(\sqrt{x}) \ ,$$

we see that (2.10) requires that

$$(2.12)\qquad g\left(\sum \mu_i^2\right) = \sum g(\mu_i^2) \ ,$$

so that g is restricted to be linear and necessarily has the form:

$$(2.13)\qquad g(x) = cx \qquad (c \text{ constant}) \ .$$

Therefore $g(x^2) = cx^2 = \mathfrak{M}\sqrt{x^2} = \mathfrak{M}(x)$ and we have deduced that $\mathfrak{M}$ is restricted to the form

$$(2.14)\qquad \mathfrak{M}(a_i) = \mathfrak{M}(\mu_i) = c\mu_i^2 = ca_i^* a_i \ ,$$

and we have shown that the only choice of measure consistent with our additivity requirement is the square amplitude measure, apart from an arbitrary multiplicative constant which may be fixed, if desired, by normalization requirements. (The requirement that the total measure be unity implies that this constant is 1.)

The situation here is fully analogous to that of classical statistical mechanics, where one puts a measure on trajectories of systems in the phase space by placing a measure on the phase space itself, and then making assertions which hold for "almost all" trajectories (such as ergodicity, quasi-ergodicity, etc).[2] This notion of "almost all" depends here also upon the choice of measure, which is in this case taken to be Lebesgue measure on the phase space. One could, of course, contradict

---

2 See Khinchin [16].

the statements of classical statistical mechanics by choosing a measure for which only the exceptional trajectories had nonzero measure. Nevertheless the choice of Lebesgue measure on the phase space can be justified by the fact that it is the only choice for which the "conservation of probability" holds, (Liouville's theorem) and hence the only choice which makes possible any reasonable statistical deductions at all.

In our case, we wish to make statements about "trajectories" of observers. However, for us a trajectory is constantly branching (transforming from state to superposition) with each successive measurement. To have a requirement analogous to the "conservation of probability" in the classical case, we demand that the measure assigned to a trajectory at one time shall equal the sum of the measures of its separate branches at a later time. This is precisely the additivity requirement which we imposed and which leads uniquely to the choice of square-amplitude measure. Our procedure is therefore quite as justified as that of classical statistical mechanics.

Having deduced that there is a unique measure which will satisfy our requirements, the square-amplitude measure, we continue our deduction. This measure then assigns to the $i,j,\ldots,k^{th}$ element of the superposition (2.6),

$$\phi_i^{S_1}\phi_j^{S_2}\ldots\phi_k^{S_r}\psi^{S_{r+1}}\ldots\psi^{S_n}\psi^{O}_{ij\ldots k[\ldots,a_i^1,a_j^2,\ldots,a_k^r]} , \tag{2.15}$$

the measure (weight)

$$M_{ij\ldots k} = (a_i a_j\ldots a_k)^*(a_i a_j\ldots a_k) , \tag{2.16}$$

so that the observer state with memory configuration $[\ldots,a_i^1,a_j^2,\ldots,a_k^r]$ is assigned the measure $a_i^* a_i a_j^* a_j\ldots a_k^* a_k = M_{ij\ldots k}$. We see immediately that this is a product measure, namely

$$M_{ij\ldots k} = M_i M_j\ldots M_k , \tag{2.17}$$

where

$$M_\ell = a_\ell^* a_\ell ,$$

so that the measure assigned to a particular memory sequence $[\ldots,a_i^1,a_j^2,\ldots,a_k^r]$ is simply the product of the measures for the individual components of the memory sequence.

We notice now a direct correspondence of our measure structure to the probability theory of random sequences. Namely, *if we were to regard* the $M_{ij\ldots k}$ as probabilities for the sequences $[\ldots,a_i^1,a_j^2,\ldots,a_k^r]$, then the sequences are equivalent to the random sequences which are generated by ascribing to each term the *independent* probabilities $M_\ell = a_\ell^* a_\ell$. Now the probability theory is equivalent to measure theory mathematically, so that we can make use of it, while keeping in mind that all results should be translated back to measure theoretic language.

Thus, in particular, if we consider the sequences to become longer and longer (more and more observations performed) *each* memory sequence of the final superposition will satisfy any given criterion for a randomly generated sequence, generated by the independent probabilities $a_i^* a_i$, except for a set of total measure which tends toward zero as the number of observations becomes unlimited. Hence all averages of functions over *any* memory sequence, including the special case of frequencies, can be computed from the probabilities $a_i^* a_i$, except for a set of memory sequences of measure zero. We have therefore shown that the statistical assertions of Process 1 will appear to be valid to *almost all* observers described by separate elements of the superposition (2.6), in the limit as the number of observations goes to infinity.

While we have so far considered only sequences of observations of the same quantity upon identical systems, the result is equally true for arbitrary sequences of observations. For example, the sequence of observations of the quantities $A^1, A^2,\ldots, A^n,\ldots$ with (generally different) eigenfunction sets $\{\phi_i^1\}, \{\phi_j^2\},\ldots,\{\phi_k^n\},\ldots$ applied successively to the systems $S_1, S_2,\ldots, S_n,\ldots$, with (arbitrary) initial states $\psi^{S_1},\psi^{S_2},\ldots,\psi^{S_n},\ldots$ transforms the total initial state:

$$\psi^{S_1+\ldots+S_n+O} = \psi^{S_1}\psi^{S_2}\ldots\psi^{S_n}\psi^{O}_{[\ldots]} \tag{2.18}$$

by rules 1 and 2, into the final state:

$$(2.19)\qquad \psi'^{S_1+S_2+\ldots+S_n+O} = \sum_{i,j,\ldots,k} (\phi_i^1,\psi^{S_1})(\phi_j^2,\psi^{S_2})\ldots(\phi_k^n,\psi^{S_n}) \ldots\phi_i^1\phi_j^2\ldots\phi_k^n\ldots\psi^O_{[\ldots,a_i^1,a_j^2,\ldots,a_k^n,\ldots]},$$

where the memory sequence element a_ℓ^r characterizes the ℓ^{th} eigenfunction, ϕ_ℓ^r of the operator A^r. Again the square amplitude measure for each element of the superposition (2.19) reduces to the product measure of the individual memory element measures, $|(\phi_\ell^r,\psi^{S_r})|^2$ for the memory sequence element a_ℓ^r. Therefore, the memory sequence of a *typical* element of (2.19) has all the characteristics of a random sequence, with individual, independent (and now different), probabilities $|(\phi_\ell^r,\psi^{S_r})|^2$ for the r^{th} memory state.

Finally, we can generalize to the case where several observations are allowed to be performed upon the *same* system. For example, if we permit the observation of a new quantity B, (eigenfunctions η_m, memory characterization β_i) upon the system S_r for which A^r has already been observed, then the state (2.19):

$$(2.20)\qquad \psi' = \sum_{i,\ell,\ldots,k} (\phi_i^1,\psi^{S_1})\ldots(\phi_\ell^r,\psi^{S_r})\ldots(\phi_k^n,\psi^{S_n}) \phi_i^1\ldots\phi_\ell^r\ldots\phi_k^n\ldots\psi^O_{[\ldots,a_i^1,\ldots,a_\ell^r,\ldots,a_k^n,\ldots]}$$

is transformed by *Rule* 2 into the state:

$$(2.21)\quad \psi' = \sum_{i,\ldots,\ell,\ldots,k,\underline{m}} (\phi_i^1,\psi^{S_1})\ldots(\phi_\ell^r,\psi^{S_r})\ldots(\phi_k^n,\psi^{S_n})(\underline{\eta_m^r},\phi_\ell^r) \phi_i^1\ldots\phi_\mu^{r-1}\ldots\underline{\eta_m^r}\ldots\phi_\nu^{r+1}\ldots\phi_k^n\ldots\psi^O_{[\ldots,a_i^1,\ldots,a_\ell^r,\ldots,a_k^n,\ldots,\underline{\beta_m^r}\ldots]}.$$

The *relative* system states for S have been changed from the eigenstates of $A^r, \{\phi_i^r\}$, to the eigenstates of $B^r, \{\eta_m^r\}$. We notice further that, with respect to our measure on the superposition, the memory sequences still have the character of random sequences, but of random sequences for which the individual terms are no longer independent. The memory states β_m^r now depend upon the memory states a_ℓ^r which represent the result of the previous measurement upon the same system, S_r. The *joint* (normalized) measure for this pair of memory states, conditioned by fixed values for remaining memory states is:

$$(2.22) \qquad \mathrm{M}^{a_i^1 \ldots a_\mu^{r-1} a_\nu^{r+1} \ldots a_k^n}(a_\ell^r, \beta_m^r) = \frac{\mathrm{M}(a_i^1, \ldots, a_\ell^r, \ldots, a_k^n, \beta_m^r)}{\sum_{\ell,m} \mathrm{M}(a_i^2, \ldots, a_\ell^r, \ldots, a_k^n, \beta_m^r)}$$

$$= \frac{|(\phi_i^1, \psi^{S_1}) \ldots (\phi_\ell^r, \psi^{S_r}) \ldots (\phi_k^n, \psi^{S_n})(\eta_m^r, \phi_\ell^r)|^2}{\sum_{\ell,m} |(\phi_i^1, \psi^{S_1}) \ldots (\phi_\ell^r, \psi^{S_r}) \ldots (\phi_k^n, \psi^{S_n})(\eta_m^r, \phi_\ell^r)|^2}$$

$$= |(\phi_\ell^r, \psi^{S_r})|^2 \, |(\eta_m^r, \phi_\ell^r)|^2 .$$

The joint measure (2.15) is, first of all, independent of the memory states for the remaining systems ($S_1 \ldots S_n$ excluding S_r). Second, the dependence of β_m^r on a_ℓ^r is *equivalent*, measure theoretically, to that given by the *stochastic process*[3] which converts the states ϕ_ℓ^r into the states η_m^r with transition probabilities:

$$(2.23) \qquad T_{\ell m} = \text{Prob.}\,(\phi_\ell^r \to \eta_m^r) = |(\eta_m^r, \phi_\ell^r)|^2 .$$

3 Cf. Chapter II, §6.

If we were to allow yet another quantity C to be measured in S_r, the new memory states α_p^r corresponding to the eigenfunctions of C would have a similar dependence upon the previous states β_m^r, but *no direct dependence* on the still earlier states α_ℓ^r. This dependence upon only the previous result of observation is a consequence of the fact that the *relative* system states are completely determined by the last observation.

We can therefore summarize the situation for an arbitrary sequence of observations, upon the same or different systems in any order, and for which the number of observations of each quantity in each system is very large, with the following result:

Except for a set of memory sequences of measure nearly zero, the averages of any functions over a memory sequence can be calculated approximately by the use of the independent probabilities given by Process 1 for each initial observation, on a system, and by the use of the transition probabilities (2.23) for succeeding observations upon the same system. In the limit, as the number of all types of observations goes to infinity the *calculation is exact, and the exceptional set has measure zero.*

This prescription for the calculation of averages over memory sequences by probabilities assigned to individual elements is precisely that of the orthodox theory (Process 1). Therefore all predictions of the usual theory will appear to be valid to the observer in almost all observer states, since these predictions hold for almost all memory sequences.

In particular, the uncertainty principle is never violated, since, as above, the latest measurement upon a system supplies all possible information about the relative system state, so that there is no direct correlation between any earlier results of observation on the system, and the succeeding observation. Any observation of a quantity B, between two successive observations of quantity A (all on the same system) will destroy the one-one correspondence between the earlier and later memory states for the result of A. Thus for alternating observations of different quantities there are fundamental limitations upon the correlations between memory states for the same observed quantity, these limitations expressing the content of the uncertainty principle.

In conclusion, we have described in this section processes involving an idealized observer, processes which are entirely deterministic and continuous from the over-all viewpoint (the total state function is presumed to satisfy a wave equation at all times) but whose result is a superposition, each element of which describes the observer with a different memory state. We have seen that in almost all of these observer states it *appears* to the observer that the probabilistic aspects of the usual form of quantum theory are valid. We have thus seen how pure wave mechanics, without any initial probability assertions, can lead to these notions on a subjective level, as appearances to observers.

§3. *Several observers*

We shall now consider the consequences of our scheme when several observers are allowed to interact with the same systems, as well as with one another (communication). In the following discussion observers shall be denoted by $O_1, O_2, \ldots$, other systems by $S_1, S_2, \ldots$, and observables by operators A, B, C, with eigenfunctions $\{\phi_i\}$, $\{\eta_j\}$, $\{\xi_k\}$ respectively. The symbols α_i, β_j, γ_k, occurring in memory sequences shall refer to characteristics of the states ϕ_i, η_j, ξ_k, respectively. $(\psi^{O}_{i[\ldots,\alpha_i]}$ is interpreted as describing an observer, O_j, who has just observed the eigenvalue corresponding to ϕ_i, i.e., who is "aware" that the system is in state ϕ_i.)

We shall also wish to allow communication among the observers, which we view as an interaction by means of which the memory sequences of different observers become correlated. (For example, the transfer of impulses from the magnetic tape memory of one mechanical observer to that of another constitutes such a transfer of information.)[4] We shall regard these processes as observations made by one observer on another and shall use the notation that

[4] We assume that such transfers merely duplicate, but do not destroy, the original information.

$$\psi^{O_j}_{i[\ldots,a_i}{}^{O_k]}$$

represents a state function describing an observer O_j who has obtained the information a_i from another observer, O_k. Thus the obtaining of information about A from O_1 by O_2 will transform the state

$$\psi^{O_1}_{i[\ldots,a_i]}\psi^{O_2}_{[\ldots]}$$

into the state

$$\psi^{O_1}_{i[\ldots,a_i]}\psi^{O_2}_{i[\ldots,a_i}{}^{O_1]} \cdot \tag{3.1}$$

Rules 1 and 2 are, of course, equally applicable to these interactions. We shall now illustrate the possibilities for several observers, by considering several cases.

Case 1: We allow two observers to separately observe the same quantity in a system, and then compare results.

We suppose that first observer O_1 observes the quantity A for the system S. Then by *Rule* 1 the original state

$$\psi^{S+O_1+O_2} = \psi^S\psi^{O_1}_{[\ldots]}\psi^{O_2}_{[\ldots]}$$

is transformed into the state

$$\psi' = \sum_i (\phi_i^S,\psi^S)\phi_i^S\psi^{O_1}_{i[\ldots,a_i]}\psi^{O_2}_{[\ldots]} \cdot \tag{3.2}$$

We now suppose that O_2 observes A, and by *Rule* 2 the state becomes:

$$\psi'' = \sum_i (\phi_i^S,\psi^S)\phi_i^S\psi^{O_1}_{i[\ldots,a_i]}\psi^{O_2}_{i[\ldots,a_i]} \cdot \tag{3.3}$$

We now allow O_2 to "consult" O_1, which leads in the same fashion from (3.1) and *Rule* 2 to the final state

$$\psi'' = \sum_i (\phi_i^S, \psi^S) \phi_i^S \psi_{i[\ldots,a_i]}^{O_1} \psi_{ii[\ldots,a_i,a_i]}^{O_2} \cdot \tag{3.4}$$

Thus, for every element of the superposition the information obtained from O_1 agrees with that obtained directly from the system. This means that observers who have separately observed the same quantity will *always* agree with each other.

Furthermore, it is obvious at this point that the same result, (4.4), is obtained if O_2 *first* consults O_1, then performs the direct observation, except that the memory sequence for O_2 is reversed ($[\ldots,a_i^{O_1},a_i]$ instead of $[\ldots,a_i,a_i^{O_1}]$). There is still perfect agreement in every element of the superposition. Therefore, information obtained from another observer is always reliable, since subsequent direct observation will always verify it. We thus see the central role played by correlations in wave functions for the preservation of consistency in situations where several observers are allowed to consult one another. It is the transitivity of correlation in these cases (that if S_1 is correlated to S_2, and S_2 to S_3, then so is S_1 to S_2) which is responsible for this consistency.

Case 2: We allow two observers to measure separately two different, non-commuting quantities in the same system.

Assume that first O_1 observes A for the system, so that, as before, the initial state $\psi^S \psi^{O_1} \psi^{O_2}$ is transformed to:

$$\psi' = \sum_i (\phi_i, \psi^S) \phi_i \psi_{i[\ldots,a_i]}^{O_1} \psi_{[\ldots]}^{O_2} \cdot \tag{3.5}$$

Next let O_2 determine β for the system, where $\{\eta_j\}$ are the eigenfunctions of β. Then by application of *Rule* 2 the result is

$$(3.6) \qquad \psi'' = \sum_{i,j} (\phi_i, \psi^S)(\eta_j, \phi_i)(\eta_j \overset{O_1}{\psi_{i[\ldots,\alpha_i]}} \overset{O_2}{\psi_{j[\ldots,\beta_j]}}$$

O_2 is now perfectly correlated with the system, since a redetermination by him will lead to agreeing results. This is no longer the case for O_1, however, since a redetermination of A by him will result in (by *Rule* 2)

$$(3.7) \qquad \psi'' = \sum_{i,j,k} (\phi_i, \psi^S)(\eta_j, \phi_i)(\phi_k, \eta_j) \overset{S}{\phi_k} \overset{O_2}{\psi_{j[\ldots,\beta_j]}} \overset{O_1}{\psi_{ik[\ldots,\alpha_i,\alpha_k]}} \cdot$$

Hence the second measurement of O_1 does not in all cases agree with the first, and has been upset by the intervention of O_2.

We can deduce the statistical relation between O_1's first and second results (α_i and α_k) by our previous method of assigning a measure to the elements of the superposition (3.7). The measure assigned to the $(i, j, k)^{th}$ element is then:

$$(3.8) \qquad M_{ijk} = |(\phi_i, \psi^S)(\eta_j, \phi_i)(\phi_k, \eta_j)|^2 \cdot$$

This measure is equivalent, in this case, to the probabilities assigned by the orthodox theory (Process 1), where O_2's observation is regarded as having converted each state ϕ_i into a non-interfering mixture of states η_j, weighted with probabilities $|(\eta_j, \phi_i)|^2$, upon which O_1 makes his second observation.

Note, however, that this equivalence with the statistical results obtained by considering that O_2's observation changed the system state into a mixture, holds true *only so long as* O_1's *second observation is restricted to the system.* If he were to attempt to simultaneously determine a property of the system as well as of O_2, interference effects might become important. The description of the states relative to O_1, after O_2's observation, as non-interfering mixtures is therefore incomplete.

Case 3: We suppose that two systems S_1 and S_2 are correlated but no longer interacting, and that O_1 measures property A in S_1, and O_2 property β in S_2.

We wish to see whether O_2's intervention with S_2 can in any way affect O_1's results in S_1, so that perhaps signals might be sent by these means. We shall assume that the initial state for the system pair is

$$\psi^{S_1+S_2} = \sum_i a_i \phi_i^{S_1} \phi_i^{S_2} \quad . \tag{3.9}$$

We now allow O_1 to observe A in S_1, so that after this observation the total state becomes:

$$\psi'^{S_1+S_2+O_1+O_2} = \sum_i a_i \phi_i^{S_1} \phi_i^{S_2} \psi_{i[\ldots,a_i]}^{O_1} \psi_{[\ldots]}^{O_2} \quad . \tag{3.10}$$

O_1 can of course continue to repeat the determination, obtaining the same result each time.

We now suppose that O_2 determines β in S_2, which results in

$$\psi'' = \sum_{i,j} a_i(\eta_j^2, \phi_i^2) \phi_i^1 \eta_j^2 \psi_{i[\ldots,a_i]}^{O_1} \psi_{j[\ldots,\beta_j]}^{O_2} \quad . \tag{3.11}$$

However, in this case, as distinct from *Case* 2, we see that the intervention of O_2 in no way affects O_1's determinations, since O_1 is still perfectly correlated to the states $\phi_i^{S_1}$ of S_1, and any further observations by O_1 will lead to the same results as the earlier observations. Thus each memory sequence for O_1 continues without change due to O_2's observation, and such a scheme could not be used to send any signals.

Furthermore, we see that the result (3.11) is arrived at even in the case that O_2 should make his determination before that of O_1. Therefore any expectations for the outcome of O_1's first observation are in no way affected by whether or not O_2 performs his observation before that

of O_1. This is true because the expectation of the outcome for O_1 can be computed from (4.10), which is the same whether or not O_2 performs his measurement before or after O_1.

It is therefore seen that one observer's observation upon one system of a correlated, but non-interacting pair of systems, has no effect on the remote system, in the sense that the outcome or expected outcome of any experiments by another observer on the remote system are not affected. Paradoxes like that of Einstein-Rosen-Podolsky[5] which are concerned with such correlated, non-interacting, systems are thus easily understood in the present scheme.

Many further combinations of several observers and systems can be easily studied in the present framework, and all questions answered by first writing down the final state for the situation with the aid of the *Rules* 1 and 2, and then noticing the relations between the elements of the memory sequences.

5 Einstein [8].

V. SUPPLEMENTARY TOPICS

We have now completed the abstract treatment of measurement and observation, with the deduction that the statistical predictions of the usual form of quantum theory (Process 1) will appear to be valid to all observers. We have therefore succeeded in placing our theory in correspondence with experience, at least insofar as the ordinary theory correctly represents experience.

We should like to emphasize that this deduction was carried out by using only the principle of superposition, and the postulate that an observation has the property that *if* the observed variable has a definite value in the object-system then it will remain definite and the observer will perceive this value. This treatment is therefore valid for any possible quantum interpretation of observation processes, i.e., any way in which one can interpret wave functions as describing observers, as well as for any form of quantum mechanics for which the superposition principle for states is maintained. Our abstract discussion of observation is therefore logically complete, in the sense that our results for the subjective experience of observers are correct, if there are any observers at all describable by wave mechanics.[1]

In this chapter we shall consider a number of diverse topics from the point of view of our pure wave mechanics, in order to supplement the abstract discussion and give a feeling for the new viewpoint. Since we are now mainly interested in elucidating the reasonableness of the theory, we shall often restrict ourselves to plausibility arguments, rather than detailed proofs.

---

[1] They are, of course, vacuously correct otherwise.

§1. *Macroscopic objects and classical mechanics*

In the light of our knowledge about the atomic constitution of matter, any "object" of macroscopic size is composed of an enormous number of constituent particles. The wave function for such an object is then in a space of fantastically high dimension (3N, if N is the number of particles). Our present problem is to understand the existence of macroscopic objects, and to relate their ordinary (classical) behavior in the three dimensional world to the underlying wave mechanics in the higher dimensional space.

Let us begin by considering a relatively simple case. Suppose that we place in a box an electron and a proton, each in a definite momentum state, so that the position amplitude density of each is uniform over the whole box. After a time we would expect a hydrogen atom in the ground state to form, with ensuing radiation. We notice, however, that the position amplitude density of each particle is *still* uniform over the whole box. Nevertheless the amplitude distributions are now no longer independent, but correlated. In particular, the *conditional* amplitude density for the electron, conditioned by any definite proton (or centroid) position, is *not* uniform, but is given by the familiar ground state wave function for the hydrogen atom. What we mean by the statement, "a hydrogen atom has formed in the box," is just that this correlation has taken place – a correlation which insures that the *relative* configuration for the electron, for a definite proton position, conforms to the customary ground state configuration.

The wave function for the hydrogen atom can be represented as a product of a centroid wave function and a wave function over relative coordinates, where the centroid wave function obeys the wave equation for a particle with mass equal to the total mass of the proton-electron system. Therefore, if we now open our box, the centroid wave function will spread with time in the usual manner of wave packets, to eventually occupy a vast region of space. The *relative* configuration (described by the *relative coordinate* state function) has, however, a permanent nature, since

it represents a bound state, and it is this relative configuration which we usually think of as the object called the hydrogen atom. Therefore, no matter how indefinite the positions of the individual particles become in the total state function (due to the spreading of the centroid), this state can be regarded as giving (through the centroid wave function) an amplitude distribution over a comparatively definite object, the tightly bound electron-proton system. The general state, then, does not describe any single such definite object, but a superposition of such cases with the object located at different positions.

In a similar fashion larger and more complex objects can be built up through strong correlations which bind together the constituent particles. It is still true that the general state function for such a system may lead to marginal position densities for any single particle (or centroid) which extend over large regions of space. Nevertheless we can speak of the existence of a relatively definite object, since the specification of a single position for a particle, or the centroid, leads to the case where the *relative position densities of the remaining* particles are distributed closely about the specified one, in a manner forming the comparatively definite object spoken of.

Suppose, for example, we begin with a cannonball located at the origin, described by a state function:

$$\psi_{[c_j(0,0,0)]} ,$$

where the subscript indicates that the total state function ψ describes a system of particles bound together so as to form an object of the size and shape of a cannonball, whose centroid is located (approximately) at the origin, say in the form of a real gaussian wave packet of small dimensions, with variance σ_0^2 for each dimension.

If we now allow a long lapse of time, the centroid of the system will spread in the usual manner to occupy a large region of space. (The spread in each dimension after time t will be given by $\sigma_t^2 = \sigma_0^2 + (\hbar^2 t^2 / 4\sigma_0^2 m^2)$,

where m is the mass.) Nevertheless, for any *specified* centroid position, the particles, since they remain in bound states, have distributions which again correspond to the fairly well defined size and shape of the cannonball. Thus the total state can be regarded as a (continuous) superposition of states

$$\psi = \int a_{xyz}\, \psi_{[c_j(x,y,z)]}\, dx\, dy\, dz \ ,$$

each of which $(\psi_{[c_j(x,y,z)]})$ describes a cannonball at the position (x, y, z). The coefficients a_{xyz} of the superposition then correspond to the centroid distribution.

It is *not* true that each individual particle spreads independently of the rest, in which case we would have a final state which is a grand superposition of states in which the particles are located independently everywhere. The fact that they are in bound states restricts our final state to a superposition of "cannonball" states. The wave function for the centroid can therefore be taken as a representative wave function for the whole object.

It is thus in this sense of correlations between constituent particles that definite macroscopic objects can exist within the framework of pure wave mechanics. The building up of correlations in a complex system supplies us with a mechanism which also allows us to understand how condensation phenomena (the formation of spatial boundaries which separate phases of different physical or chemical properties) can be controlled by the wave equation, answering a point raised by Schrödinger

Classical mechanics, also, enters our scheme in the form of correlation laws. Let us consider a system of objects (in the previous sense), such that the centroid of each object has initially a fairly well defined position and momentum (e.g., let the wave function for the centroids consist of a product of gaussian wave packets). As time progresses, the

centers of the square amplitude distributions for the objects will move in a manner approximately obeying the laws of motion of classical mechanics, with the degree of approximation depending upon the masses and the length of time considered, as is well known. (Note that we do not mean to imply that the wave packets of the individual objects remain independent if they are interacting. They do not. The motion that we refer to is that of the centers of the *marginal* distributions for the centroids of the bodies.)

The general state of a system of macroscopic objects does not, however, ascribe any nearly definite positions and momenta to the individual bodies. Nevertheless, any general state can at any instant be analyzed into a *superposition* of states each of which *does* represent the bodies with fairly well defined positions and momenta.[2] Each of these states then propagates approximately according to classical laws, so that the general state can be viewed as a superposition of quasi-classical states propagating according to nearly classical trajectories. In other words, if the masses are large or the time short, there will be strong correlations between the initial (approximate) positions and momenta and those at a later time, with the dependence being given approximately by classical mechanics.

Since large scale objects obeying classical laws have a place in our theory of pure wave mechanics, we have justified the introduction of

---

[2] For any ε one can construct a complete orthonormal set of (one particle) states $\phi_{\mu,\nu}$, where the double index μ,ν refers to the approximate position and momentum, and for which the expected position and momentum values run independently through sets of approximately uniform density, such that the position and momentum uncertainties, σ_x and σ_p, satisfy $\sigma_x \leqq C\varepsilon$ and $\sigma_p \leqq C\frac{\hbar}{2\varepsilon}$ for each $\phi_{\mu,\nu}$, where C is a constant ~ 60. The uncertainty product then satisfies $\sigma_x\sigma_p \leqq C^2\frac{\hbar}{2}$, about 3,600 times the minimum allowable, but still sufficiently low for macroscopic objects. This set can then be used as a basis for our decomposition into states where every body has a roughly defined position and momentum. For a more complete discussion of this set see von Neumann [17], pp. 406-407.

models for observers consisting of classically describable, automatically functioning machinery, and the treatment of observation of Chapter IV is non-vacuous.

Let us now consider the result of an observation (considered along the lines of Chapter IV) performed upon a system of macroscopic bodies in a general state. The observer will *not* become aware of the fact that the state does not correspond to definite positions and momenta (i.e., he will not see the objects as "smeared out" over large regions of space) but will himself simply become correlated with the system – after the observation the composite system of objects + observer will be in a superposition of states, each element of which describes an observer who has perceived that the objects have nearly definite positions and momenta, and for whom the relative system state is a quasi-classical state in the previous sense, and furthermore to whom the system will appear to behave according to classical mechanics if his observation is continued. We see, therefore, how the classical appearance of the macroscopic world to us can be explained in the wave theory.

§2. *Amplification processes*

In Chapter III and IV we discussed abstract measuring processes, which were considered to be simply a direct coupling between two systems, the object-system and the apparatus (or observer). There is, however, in actuality a whole chain of intervening systems linking a microscopic system to a macroscopic observer. Each link in the chain of intervening systems becomes correlated to its predecessor, so that the result is an amplification of effects from the microscopic object-system to a macroscopic apparatus, and then to the observer.

The amplification process depends upon the ability of the state of one micro-system (particle, for example) to become correlated with the states of an enormous number of other microscopic systems, the totality of which we shall call a detection system. For example, the totality of gas atoms in a Geiger counter, or the water molecules in a cloud chamber, constitute such a detection system.

The amplification is accomplished by arranging the condition of the detection system so that the states of the individual micro-systems of the detector are *metastable*, in a way that if one micro-system should fall from its metastable state it would influence the reduction of others. This type of arrangement leaves the entire detection system metastable against chain reactions which involve a large number of its constituent systems. In a Geiger counter, for example, the presence of a strong electric field leaves the gas atoms metastable against ionization. Furthermore, the products of the ionization of one gas atom in a Geiger counter can cause further ionizations, in a cascading process. The operation of cloud chambers and photographic films is also due to metastability against such chain reactions.

The chain reactions cause large numbers of the micro-systems of the detector to behave as a unit, all remaining in the metastable state, or all discharging. In this manner the states of a sufficiently large number of micro-systems are correlated, so that one can speak of the whole ensemble *being in a state of discharge, or not.*

For example, there are essentially only two macroscopically distinguishable states for a Geiger counter; discharged or undischarged. The correlation of large numbers of gas atoms, due to the chain reaction effect, implies that either very few, or else very many of the gas atoms are ionized at a given time. Consider the complete state function ψ^G of a Geiger counter, which is a function of all the coordinates of all of the constituent particles. Because of the correlation of the behavior of a large number of the constituent gas atoms, the total state ψ^G can always be written as a superposition of two states

$$\psi^G = a_1 \psi^1_{[U]} + a_2 \psi^2_{[D]} \,, \tag{2.1}$$

where $\psi^1_{[U]}$ signifies a state where only a small number of gas atoms are ionized, and $\psi^2_{[D]}$ a state for which a large number are ionized.

To see that the decomposition (2.1) is valid, expand ψ^G in terms of individual gas atom stationary states:

$$\psi^G = \sum_{i,j,\ldots,k} a_{ij\ldots k} \psi_i^{S_1} \psi_j^{S_2} \ldots \psi_k^{S_n} , \tag{2.2}$$

where $\psi_\ell^{S_r}$ is the ℓ^{th} state of atom r. Each element of the superposition (2.2)

$$\psi_i^{S_1} \psi_j^{S_2} \ldots \psi_k^{S_n} \tag{2.3}$$

must contain either a very large number of atoms in ionized states, or else a very small number, because of the chain reaction effect. By choosing some medium-sized number as a dividing line, each element of (2.2) can be placed in one of the two categories, high number of low number of ionized atoms. If we then carry out the sum (2.2) over only those elements of the first category, we get a state (and coefficient)

$$a_1 \psi_{[D]}^1 = \sum_{ij\ldots k}' a_{ij\ldots k} \psi_i^{S_1} \psi_j^{S_2} \ldots \psi_k^{S_n} . \tag{2.4}$$

The state $\psi_{[D]}^1$ is then a state where a large number of particles are ionized. The subscript [D] indicates that it describes a Geiger counter which has discharged. If we carry out the sum over the remaining terms of (2.2) we get in a similar fashion:

$$a_2 \psi_{[U]}^2 = \sum_{ij\ldots k}'' a_{ij\ldots k} \psi_i^{S_1} \psi_j^{S_2} \ldots \psi_k^{S_n} \tag{2.5}$$

where [U] indicates the undischarged condition. Combining (2.4) and (2.5) we arrive at the desired relation (2.1). So far, this method of decomposition can be applied to any system, whether or not it has the chain reaction property. However, in our case, more is implied, namely that the spread of the number of ionized atoms in both $\psi_{[D]}$ and $\psi_{[U]}$ will be small compared to the separation of their averages, due to the fact that

the existence of the chain reactions means that either many or else few atoms will be ionized, with the middle ground virtually excluded.

This type of decomposition is also applicable to all other detection devices which are based upon this chain reaction principle (such as cloud chambers, photo plates, etc.).

We consider now the coupling of such a detection device to another micro-system (object-system) for the purpose of measurement. If it is true that the initial object-system state ϕ_1 will at some time t trigger the chain reaction, so that the state of the counter becomes $\psi^1_{[D]}$, while the object-system state ϕ_2 will not, then it is still true that the initial object-system state $a_1\phi_1 + a_2\phi_2$ will result in the superposition

$$a_1\phi'_1\psi^1_{[D]} + a_2\phi'_2\psi^2_{[U]} \tag{2.6}$$

at time t.

For example, let us suppose that a particle whose state is a wave packet ϕ, of linear extension greater than that of our Geiger counter, approaches the counter. Just before it reaches the counter, it can be decomposed into a superposition $\phi = a_1\phi_1 + a_2\phi_2$ (ϕ_1, ϕ_2 orthogonal) where ϕ_1 has non-zero amplitude only in the region before the counter and ϕ_2 has non-zero amplitude elsewhere (so that ϕ_1 is a packet which will entirely pass through the counter while ϕ_2 will entirely miss the counter). The initial total state for the system particle + counter is then:

$$\phi\psi_{[U]} = (a_1\phi_1 + a_2\phi_2)\psi_{[U]} ,$$

where $\psi_{[U]}$ is the initial (assumed to be undischarged) state of the counter.

But at a slightly later time ϕ_1 is changed to ϕ'_1, after traversing the counter and causing it to go into a discharged state $\psi^1_{[D]}$, while ϕ_2 passes by into a state ϕ'_2 leaving the counter in an undischarged state $\psi^2_{[U]}$. Superposing these results, the total state at the later time is

(2.7) $$a_1\phi'_1\psi^1_{[D]} + a_2\phi'_2\psi^2_{[U]}$$

in accordance with (2.6). Furthermore, the relative particle state for $\psi^1_{[D]}$, ϕ'_1, is a wave packet emanating from the counter, while the relative state for $\psi^2_{[U]}$ is a wave with a "shadow" cast by the counter. The counter therefore serves as an apparatus which performs an approximate position measurement on the particle.

No matter what the complexity or exact mechanism of a measuring process, the general superposition principle as stated in Chapter III, §3, remains valid, and our abstract discussion is unaffected. It is a vain hope that somewhere embedded in the intricacy of the amplification process is a mechanism which will somehow prevent the macroscopic apparatus state from reflecting the same indefiniteness as its object-system.

§3. *Reversibility and irreversibility*

Let us return, for the moment, to the probabilistic interpretation of quantum mechanics based on Process 1 as well as Process 2. Suppose that we have a large number of identical systems (ensemble), and that the j^{th} system is in the state ψ^j. Then for purposes of calculating expectation values for operators over the ensemble, the ensemble is represented by the mixture of states ψ^j weighted with $1/N$, where N is the number of systems, for which the density operator[3] is:

(3.1) $$\rho = \frac{1}{N}\sum_j [\psi^j] \ ,$$

where $[\psi^j]$ denotes the projection operator on ψ^j. This density operator, in turn, is equivalent to a density operator which is a sum of projections on orthogonal states (the eigenstates of ρ):[4]

---

[3] Cf. Chapter III, §1.

[4] See Chapter III, §2, particularly footnote 6, p. 46.

(3.2) $$\rho = \sum_i P_i [\eta_i] \ , \quad (\eta_i, \eta_j) = \delta_{ij}, \ \sum_i P_i = 1 \ ,$$

so that any ensemble is always equivalent to a mixture of orthogonal states, which representation we shall henceforth assume.

Suppose that a quantity A, with (non-degenerate) eigenstates $\{\phi_j\}$ is measured in each system of the ensemble. This measurement has the effect of transforming each state η_i into the state ϕ_j, with probability $|(\phi_j, \eta_i)|^2$; i.e., it will transform a large ensemble of systems in the state η_i into an ensemble represented by the mixture whose density operator is $\sum_j |(\phi_j, \eta_i)|^2 [\phi_j]$. Extending this result to the case where the original ensemble is a mixture of the η_i weighted by P_i ((3.2)), we find that the density operator ρ is transformed by the measurement of A into the new density operator ρ':

(3.3) $$\rho' = \sum_i P_i \sum_j |(\eta_i, \phi_j)|^2 [\phi_j] = \sum_j \left(\sum_i P_i (\phi_j, (\eta_i, \phi_j) \eta_i) \right) [\phi_j]$$

$$= \sum_j \left(\phi_j, \sum_i P_i [\eta_i] \phi_j \right) [\phi_j] = \sum_j (\phi_j, \rho \phi_j) [\phi_j] \ .$$

This is the general law by which mixtures change through Process 1.

However, even when no measurements are taking place, the states of an ensemble are changing according to Process 2, so that after a time interval t each state ψ will be transformed into a state $\psi' = U_t \psi$, where U_t is a unitary operator. This natural motion has the consequence that each mixture $\rho = \sum_i P_i [\eta_i]$ is carried into the mixture $\rho' = \sum_i P_i [U_t \eta_i]$ after a time t. But for every state ξ,

(3.4) $$\rho' \xi = \sum_i P_i [U_t \eta_i] \xi = \sum_i P_i (U_t \eta_i, \xi) U_t \eta_i$$

$$= U_t \sum_i P_i (\eta_i, U_t^{-1} \xi) \eta_i = U_t \sum_i P_i [\eta_i] (U_t^{-1} \xi)$$

$$= (U_t \rho U_t^{-1}) \xi \ .$$

Therefore

$$\rho' = U_t \rho U_t^{-1} \quad , \tag{3.5}$$

which is the general law for the change of a mixture according to Process 2.

We are now interested in whether or not we get from any mixture to another by means of these two processes, i.e., if for any pair ρ, ρ', there exist quantities A which can be measured and unitary (time dependence) operators U such that ρ can be transformed into ρ' by suitable applications of Processes 1 and 2. We shall see that this is not always possible, and that Process 1 can cause irreversible changes in mixtures.

For each mixture ρ we define a quantity I_ρ:

$$I_\rho = \text{Trace } (\rho \ln \rho) \quad . \tag{3.6}$$

This number, I_ρ, has the character of information. If $\rho = \sum_i P_i [\eta_i]$, a mixture of orthogonal states η_i weighted with P_i, then I_ρ is simply the information of the distribution P_i over the eigenstates of ρ (relative to the uniform measure). (Trace $(\rho \ln \rho)$ is a unitary invariant and is proportional to the negative of the entropy of the mixture, as discussed in Chapter III, §2.)

Process 2 therefore has the property that it leaves I_ρ unchanged, because

$$\begin{aligned} I_{\rho'} &= \text{Trace } (\rho' \ln \rho') = \text{Trace } (U_t \rho U_t^{-1} \ln U_t \rho U_t^{-1}) \\ &= \text{Trace } (U_t \rho \ln \rho U_t^{-1}) = \text{Trace } (\rho \ln \rho) = I_\rho \quad . \end{aligned} \tag{3.7}$$

Process 1, on the other hand, can decrease I_ρ but never increase it. According to (3.3):

$$\rho' = \sum_j (\phi_j, \rho \phi_j) [\phi_j] = \sum_{i,j} P_i \, |(\eta_i, \phi_j)|^2 [\phi_j] = \sum_j P'_j [\phi_j] \quad , \tag{3.8}$$

where $\rho'_j \quad \sum_i P_i T_{ij}$ and $T_{ij} = |(\eta_i, \phi_j)|^2$ is a doubly-stochastic matrix.[5] But $I_{\rho'} = \sum_j P'_j \ln P'_j$ and $I_\rho = \sum_i P_i \ln P_i$, with the P_i, P'_j connected by T_{ij}, implies, by the theorem of information decrease for stochastic processes (II-§6), that:

$$I_{\rho'} \leqq I_\rho \ . \tag{3.9}$$

Moreover, it can easily be shown by a slight strengthening of the theorems of Chapter II, §6 that *strict* inequality must hold unless (for each i such that $\rho_i > 0$) $T_{ij} = 1$ for one j and 0 for the rest ($T_{ij} = \delta_{ikj}$). This means that $|(\eta_i, \phi_j)|^2 = \delta_{ikj}$, which implies that the original mixture was already a mixture of eigenstates of the measurement.

We have answered our question, and it is *not* possible to get from any mixture to another by means of Processes 1 and 2. There is an essential irreversibility to Process 1, since it corresponds to a stochastic process, which cannot be compensated by Process 2, which is reversible, like classical mechanics.[6]

Our theory of pure wave mechanics, to which we now return, must give equivalent results on the subjective level, since it leads to Process 1 there. Therefore, measuring processes will appear to be irreversible to any observers (even though the composite system including the observer changes its state reversibly).

---

[5] Since $\sum_i T_{ij} = \sum_i |(\eta_i, \phi_j)|^2 = \sum_i (\phi_j, [\eta_i]\phi_j) = (\phi_j, \sum_i [\eta_i]\phi_j) = (\phi_j, I\phi_j) = 1$, and similarly $\sum_j T_{ij} = 1$ because T_{ij} is symmetric.

[6] For another, more complete, discussion of this topic in the probabilistic interpretation see von Neumann [17], Chapter V, §4.

There is another way of looking at this apparent irreversibility within our theory which recognizes only Process 2. When an observer performs an observation the result is a superposition, each element of which describes an observer who has perceived a particular value. From this time forward there is no interaction between the separate elements of the superposition (which describe the observer as having perceived different results), since each element separately continues to obey the wave equation. Each observer described by a particular element of the superposition behaves in the future completely independently of any events in the remaining elements, and he can no longer obtain any information whatsoever concerning these other elements (they are completely unobservable to him).

The irreversibility of the measuring process is therefore, within our framework, simply a subjective manifestation reflecting the fact that in observation processes the state of the observer is transformed into a superposition of observer states, each element of which describes an observer who is irrevocably cut off from the remaining elements. While it is conceivable that some outside agency could reverse the total wave function, such a change cannot be brought about by any observer which is represented by a single element of a superposition, since he is entirely powerless to have any influence on any other elements.

There are, therefore, fundamental restrictions to the knowledge that an observer can obtain about the state of the universe. It is impossible for any observer to discover the total state function of any physical system, since the process of observation itself leaves no independent state for the system or the observer, but only a composite system state in which the object-system states are inextricably bound up with the observer states. As soon as the observation is performed, the composite state is split into a superposition for which each element describes a different object-system state and an observer with (different) knowledge of it. Only the totality of these observer states, with their diverse knowledge, contains complete information about the original object-system state – but there is no possible communication between the observers described by these separate

states. Any single observer can therefore possess knowledge only of the relative state function (relative to his state) of any systems, which is in any case all that is of any importance to him.

We conclude this section by commenting on another question which might be raised concerning irreversible processes: Is it necessary for the existence of measuring apparata, which can be correlated to other systems, to have frictional processes which involve systems of a large number of degrees of freedom? Are such thermodynamically irreversible processes possible in the framework of pure wave mechanics with a reversible wave equation, and if so, does this circumstance pose any difficulties for our treatment of measuring processes?

In the first place, it is certainly not necessary for dissipative processes involving additional degrees of freedom to be present before an interaction which correlates an apparatus to an object-system can take place. The counter-example is supplied by the simplified measuring process of III-§3, which involves only a system of one coordinate and an apparatus of one coordinate and no further degrees of freedom.

To the question whether such processes are possible within reversible wave mechanics, we answer *yes*, in the same sense that they are present in classical mechanics, where the microscopic equations of motion are also reversible. This type of irreversibility, which might be called *macroscopic irreversibility*, arises from a failure to separate "macroscopically indistinguishable" states into "true" microscopic states.[7] It has a fundamentally different character from the irreversibility of Process 1, which applies to micro-states as well and is peculiar to quantum mechanics. Macroscopically irreversible phenomena are common to both classical and quantum mechanics, since they arise from our incomplete information concerning a system, not from any intrinsic behavior of the system.[8]

---

[7] See any textbook on statistical mechanics, such as ter Haar [11], Appendix I.

[8] Cf. the discussion of Chapter II, §7. See also von Neumann [17], Chapter V, §4.

Finally, even when such frictional processes are involved, they present no new difficulties for the treatment of measuring and observation processes given here. We imposed no restrictions on the complexity or number of degrees of freedom of measuring apparatus or observers, and if any of these processes are present (such as heat reservoirs, etc.) then these systems are to be simply included as part of the apparatus or observer.

§4. *Approximate measurement*

A phenomenon which is difficult to understand within the framework of the probabilistic interpretation of quantum mechanics is the result of an approximate measurement. In the abstract formulation of the usual theory there are two fundamental processes; the discontinuous, probabilistic Process 1 corresponding to precise measurement, and the continuous, deterministic Process 2 corresponding to absence of any measurement. What mixture of probability and causality are we to apply to the case where only an approximate measurement is effected (i.e., where the apparatus or observer interacts only weakly and for a finite time with the object-system)?

In the case of approximate measurement, we need to be supplied with rules which will tell us, for any initial object-system state, first, with what probability can we expect the various possible apparatus readings, and second, what new state to ascribe to the system after the value has been observed. We shall see that it is generally impossible to give these rules within a framework which considers the apparatus or observer as performing an (abstract) observation subject to Process 1, and that it is necessary, in order to give a full account of approximate measurements, to treat the entire system, including apparatus or observer, wave mechanically.

The position that an approximate measurement results in the situation that the object-system state is changed into an eigenstate of the exact measurement, but for which particular one the observer has only imprecise

information, is manifestly false. It is a fact that we can make successive approximate position measurements of particles (in cloud chambers, for example) and use the results for somewhat reliable predictions of future positions. However, if either of these measurements left the particle in an "eigenstate" of position (δ function), even though the particular one remained unknown, the momentum would have such a variance that no such prediction would be possible. (The possibility of such predictions lies in the correlations between position and momentum at one time with position and momentum at a later time for wave packets[9] – correlations which are totally destroyed by precise measurements of either quantity.)

Instead of continuing the discussion of the inadequacy of the probabilistic formulation, let us first investigate what actually happens in approximate measurements, from the viewpoint of pure wave mechanics. An approximate measurement consists of an interaction, for a finite time, which only imperfectly correlates the apparatus (or observer) with the object-system. We can deduce the desired rules in any particular case by the following method: For fixed interaction and initial apparatus state and for any initial object-system state we solve the wave equation for the time of interaction in question. The result will be a superposition of apparatus (observer) states and relative object-system states. Then (according to the method of Chapter IV for assigning a measure to a superposition) we assign a probability to each observed result equal to the square-amplitude of the coefficient of the element which contains the apparatus (observer) state representing the registering of that result. Finally, the object-system is assigned the new state which is its *relative state* in that element.

For example, let us consider the measuring process described in Chapter III-§3, which is an excellent model for an approximate measurement. After the interaction, the total state was found to be (III-(3.12)):

---

[9] See Bohm [1], p. 202.

$$\psi_t^{S+A} = \int \frac{1}{N_{r'}} \xi^{r'}(q)\, \delta(r-r')\, dr' \quad . \tag{4.1}$$

Then, according to our prescription, we assign the probability density $P(r')$ to the observation of the apparatus coordinate r'

$$P(r') = \left|\frac{1}{N_{r'}}\right|^2 = \int \phi^*\phi(q)\,\eta^*\eta(r'-qt)\,dq \quad , \tag{4.2}$$

which is the square amplitude of the coefficient $\left(\frac{1}{N_{r'}}\right)$ of the element $\xi^{r'}(q)\delta(r-r')$ of the superposition (4.1) in which the apparatus coordinate has the value $r = r'$. Then, depending upon the observed apparatus coordinate r', we assign the object-system the new state

$$\xi^{r'}(q) = N_{r'}\phi(q)\,\eta(r'-qt) \tag{4.3}$$

(where $\phi(q)$ is the old state, and $\eta(r)$ is the initial apparatus state) which is the relative object-system state in (4.1) for apparatus coordinate r'.

This example supplies the counter-example to another conceivable method of dealing with approximate measurement within the framework of Process 1. This is the position that when an approximate measurement of a quantity Q is performed, in actuality another quantity Q' is precisely measured, where the eigenstates of Q' correspond to fairly well-defined (i.e., sharply peaked distributions for) Q values.[10] However, any such scheme based on Process 1 always has the prescription that after the measurement, the (unnormalized) new state function results from the old by a projection (on an eigenstate or eigenspace), which depends upon the observed value. If this is true, then in the above example the new state $\xi^{r'}(q)$ must result from the old, $\phi(q)$, by a projection E:

$$\xi^{r'}(q) = N\,E\,\phi(q) = N_{r'}\phi(q)\,\eta(r'-qt) \tag{4.4}$$

[10] Cf. von Neumann [17], Chapter IV, §4.

where $N, N_{r'}$ are normalization constants). But E is only a projection if $E^2 = E$. Applying the operation (4.4) twice, we get:

$$(4.5) \qquad E(N E\phi(q)) = N E^2\phi(q) = N'\phi(q)\eta^2(r'-qt) \Rightarrow E^2\phi(q)$$

$$= \frac{N'}{N}\phi(q)\eta^2(r'-qt) ,$$

and we see that E cannot be a projection unless $\eta(q) = \eta^2(q)$ for all q (i.e., $\eta(q) = 0$ or 1 for all q) and we have arrived at a contradiction to the assumption that in all cases the changes of states for approximate measurements are governed by projections. (In certain special cases, such as approximate position measurements with slits or Geiger counters,[11] the new functions arise from the old by multiplication by sharp cutoff functions which are 1 over the slit or counter and 0 elsewhere, so that these measurements *can* be handled by projections.)

One cannot, therefore, account for approximate measurements by any scheme based on Process 1, and it is necessary to investigate these processes entirely wave-mechanically. Our viewpoint constitutes a framework in which it is possible to make precise deductions about such measurements and observations, since we can follow in detail the interaction of an observer or apparatus with an object-system.

§5. *Discussion of a spin measurement example*

We shall conclude this chapter with a discussion of an instructive example of Bohm.[12] Bohm considers the measurement of the z component of the angular momentum of an atom, whose total angular momentum is $\frac{\hbar}{2}$, which is brought about by a Stern-Gerlach experiment. The measurement

---

[11] Cf. §2, this chapter.

[12] Bohm [1], p. 593.

is accomplished by passing an atomic beam through an inhomogeneous magnetic field, which has the effect of giving the particle a momentum which is directed up or down depending upon whether the spin was up or down.

The measurement is treated as impulsive, so that during the time that the atom passes through the field the Hamiltonian is taken to be simply the interaction:

$$H_I = \mu(\vec{\delta} \cdot \vec{\mathcal{H}}) , \quad \mu = -\frac{e\hbar}{2mc} \tag{5.1}$$

where $\mathcal{H}$ is the magnetic field and $\vec{\delta}$ the spin operator for the atom. The particle is presumed to pass through a region of the field where the field is in the z direction, so that during the time of transit the field is approximately $\mathcal{H}_z \cong \mathcal{H}_0 + z\mathcal{H}_0' \left(\mathcal{H}_0 = (\mathcal{H}_z)_{z=0}\right.$ and $\left.\mathcal{H}_0' = \left(\frac{\partial \mathcal{H}_z}{\partial z}\right)_{z=0}\right)$, and hence the interaction is approximately:

$$H_I \cong \mu(\mathcal{H}_0 + z\mathcal{H}_0')S_z , \tag{5.2}$$

where S_z denotes the operator for the z component of the spin.

It is assumed that the state of the atom, just prior to entry into the field, is a wave packet of the form:

$$\psi_0 = f_0(z)(c_+ v_+ + c_- v_-) \tag{5.3}$$

where v_+ and v_- are the spin functions for $S_z = 1$ and -1 respectively. Solving the Schrödinger equation for the Hamiltonian (5.2) and initial condition (5.3) yields the state for a later time t:

$$\psi = f_0(z)\left(c_+ e^{-i\mu(\mathcal{H}_0 + z\mathcal{H}_0')t/\hbar} v_+ + c_- e^{+i\mu(\mathcal{H}_0 + z\mathcal{H}_0')t/\hbar} v_-\right) . \tag{5.4}$$

Therefore, if Δt is the time that it takes the atom to traverse the field,[13] each component of the wave packet has been multiplied by a phase factor

---

[13] This time is, strictly speaking, not well defined. The results, however, do not depend critically upon it.

$e^{\pm i\mu(\mathcal{H}_0+z\mathcal{H}_0')\Delta t/\hbar}$, i.e., has had its mean momentum in the z direction changed by an amount $\pm\mathcal{H}_0'\mu\Delta t$, depending upon the spin direction. Thus the initial wave packet (with mean momentum zero) is split into a superposition of two packets, one with mean z-momentum $+\mathcal{H}_0'\mu\Delta t$ and spin up, and the other with spin down and mean z-momentum $-\mathcal{H}_0'\mu\Delta t$.

The interaction (5.2) has therefore served to correlate the spin with the momentum in the z-direction. These two packets of the resulting superposition now move in opposite z-directions, so that after a short time they become widely separated (provided that the momentum changes $\pm\mathcal{H}_0'\mu\Delta t$ are large compared to the momentum spread of the original packet), and the z-coordinate is itself then correlated with the spin – representing the "apparatus" coordinate in this case. The Stern-Gerlach apparatus therefore splits an incoming wave packet into a superposition of two diverging packets, corresponding to the two spin values.

We take this opportunity to caution against a certain viewpoint which can lead to difficulties. This is the idea that, after an apparatus has interacted with a system, in "actuality" one or another of the elements of the resultant superposition described by the composite state-function has been realized to the exclusion of the rest, the existing one simply being unknown to an external observer (i.e., that instead of the superposition there is a genuine mixture). This position must be erroneous since there is always the possibility for the external observer to make use of interference properties between the elements of the superposition.

In the present example, for instance, it is in principle possible to deflect the two beams back toward one another with magnetic fields and recombine them in another inhomogeneous field, which duplicates the first, in such a manner that the original spin state (before entering the apparatus) is restored.[14] This would not be possible if the original Stern-Gerlach apparatus performed the function of converting the original wave

[14] As pointed out by Bohm [1], p. 604.

packet into a non-interfering mixture of packets for the two spin cases. Therefore the position that after the atom has passed through the inhomogeneous field it is "really" in one or the other beam with the corresponding spin, although we are ignorant of which one, is incorrect.

After two systems have interacted and become correlated it is true that marginal expectations for *subsystem* operators can be calculated correctly when the composite system is represented by a certain non-interfering mixture of states. Thus if the composite system state is $\psi^{S_1+S_2} = \sum_i a_i \phi_i^{S_1} \eta_i^{S_2}$, where the $\{\eta_i\}$ are orthogonal, then for purposes of calculating the expectations of operators on S_1 the state $\psi^{S_1+S_2}$ is equivalent to the non-interfering mixture of states $\phi_i^{S_1} \eta_i^{S_2}$ weighted by $P_i = a_i^* a_i$, and one can take the picture that one or another of the cases $\phi_i^{S_1} \eta_i^{S_2}$ has been realized to the exclusion of the rest, with probabilities P_i.[15]

However, this representation by a mixture must be regarded as only a mathematical artifice which, although useful in many cases, is an *incomplete description* because it ignores phase relations between the separate elements which actually exist, and which become important in any interactions which involve more than just a subsystem.

In the present example, the "composite system" is made of the "subsystems" spin value (object-system) and z-coordinate (apparatus), and the superposition of the two diverging wave packets is the state after interaction. It is only correct to regard this state as a mixture so long as any contemplated future interactions or measurements will involve only the spin value or only the z-coordinate, but not both simultaneously. As we saw, phase relations between the two packets are present and become important when they are deflected back and recombined in another inhomogeneous field – a process involving the spin values and z-coordinate simultaneously.

---

[15] See Chapter III, §1.

It is therefore improper to attribute any less validity or "reality" to any element of a superposition than any other element, due to this ever present possibility of obtaining interference effects between the elements. All elements of a superposition must be regarded as simultaneously existing.

At this time we should like to add a few remarks concerning the notion of *transition probabilities* in quantum mechanics. Often one considers a system, with Hamiltonian H and stationary states $\{\phi_i\}$, to be perturbed for a time by a time-dependent addition to the Hamiltonian, $H_I(t)$. Then under the action of the perturbed Hamiltonian $H' = H + H_I(t)$ the states $\{\phi_i\}$ are generally no longer stationary but change after time t into new states $\{\psi_i(t)\}$:

$$\phi_i \to \psi_i(t) = \sum_j (\phi_j, \psi_i(t))\phi_j = \sum_j a_{ij}(t)\phi_j \ , \tag{5.5}$$

which can be represented as a superposition of the old stationary states with time-dependent coefficients $a_{ij}(t)$.

If at time τ a measurement with eigenstates ϕ_j is performed, such as an energy measurement (whose operator is the original H), then according to the probabilistic interpretation the probability for finding the state ϕ_j, given that the state was originally ϕ_i, is $P_{ij}(\tau) = |a_{ij}(\tau)|^2$. The quantities $|a_{ij}(\tau)|^2$ are often referred to as *transition probabilities*. In this case, however, the name is a misnomer, since it carries the connotation that the original state ϕ_i is transformed into a *mixture* (of the ϕ_j weighted by $P_{ij}(\tau)$), and gives the erroneous impression that the quantum formalism itself implies the existence of quantum-jumps (stochastic processes) independent of acts of observation. This is incorrect since there is still a pure state $\sum_j a_{ij}(\tau)\phi_j$ with phase relations between the ϕ_j, and expectations of operators other than the energy *must* be calculated from the superposition and not the mixture.

There is another case, however, the one usually encountered in fact, where the transition probability concept is somewhat more justified. This

is the case in which the perturbation is due to interaction of the system s_1 with another system s_2, and not simply a time dependence of s_1's Hamiltonian as in the case just considered. In this situation the interaction produces a *composite system state,* for which there are in general no independent subsystem states. However, as we have seen, for purposes of calculating expectations of operators on s_1 alone, we can regard s_1 as being represented by a certain mixture. According to this picture the states of subsystem s_1 are gradually converted into mixtures by the interaction with s_2 and the concept of transition probability makes some sense. Of course, it must be remembered that this picture is only justified so long as further measurements on s_1 alone are contemplated, and any attempt to make a simultaneous determination in s_1 and s_2 involves the composite state where interference properties may be important.

An example is a hydrogen atom interacting with the electromagnetic field. After a time of interaction we can picture the atom as being in a mixture of its states, so long as we consider future measurements on the atom only. But in actuality the state of the atom is dependent upon (correlated with) the state of the field, and some process involving both atom and field could conceivably depend on interference effects between the states of the alleged mixture. With these restrictions, however, the concept of transition probability is quite useful and justified.

VI. DISCUSSION

We have shown that our theory based on pure wave mechanics, which takes as the basic description of physical systems the state function – supposed to be an *objective* description (i.e., in one-one, rather than statistical, correspondence to the behavior of the system) – can be put in satisfactory correspondence with experience. We saw that the probabilistic assertions of the usual interpretation of quantum mechanics can be *deduced from this theory, in a manner analogous to the methods* of classical statistical mechanics, as subjective appearances to observers – observers which were regarded simply as physical systems subject to the same type of description and laws as any other systems, and having no preferred position. The theory is therefore capable of supplying us with a complete conceptual model of the universe, consistent with the assumption that it contains more than one observer.

Because the theory gives us an objective description, it constitutes a framework in which a number of puzzling subjects (such as classical level phenomena, the measuring process itself, the inter-relationship of several observers, questions of reversibility and irreversibility, etc.) can be investigated in detail in a logically consistent manner. It supplies a new way of viewing processes, which clarifies many apparent paradoxes of the usual interpretation[1] – indeed, *it constitutes an objective framework in* which it is possible to understand the general consistency of the ordinary view.

---

[1] Such as that of Einstein, Rosen, and Podolsky [8], as well as the paradox of the introduction.

We shall now resume our discussion of alternative interpretations. There has been expressed lately a great deal of dissatisfaction with the present form of quantum theory by a number of authors, and a wide variety of new interpretations have sprung into existence. We shall now attempt to classify briefly a number of these interpretations, and comment upon them.

a. *The "popular" interpretation.* This is the scheme alluded to in the introduction, where ψ is regarded as objectively characterizing the single system, obeying a deterministic wave equation when the system is isolated but changing probabilistically and discontinuously under observation.

In its unrestricted form this view can lead to paradoxes like that mentioned in the introduction, and is therefore untenable. However, this view *is* consistent so long as it is assumed that there is only one observer in the universe (the solipsist position – Alternative 1 of the Introduction). This consistency is most easily understood from the viewpoint of our own theory, where we were able to show that all phenomena will *seem* to follow the predictions of this scheme to any observer. Our theory therefore justifies the personal adoption of this probabilistic interpretation, for purposes of making practical predictions, from a more satisfactory framework.

b. *The Copenhagen interpretation.* This is the interpretation developed by Bohr. The ψ function is not regarded as an objective description of a physical system (i.e., it is in no sense a conceptual model), but is regarded as merely a mathematical artifice which enables one to make statistical predictions, albeit the best predictions which it is possible to make. This interpretation in fact denies the very possibility of a single conceptual model applicable to the quantum realm, and asserts that the totality of phenomena can only be understood by the use of different, mutually exclusive (i.e., "complementary") models in different situations. All state-

ments about microscopic phenomena are regarded as meaningless unless accompanied by a complete description (classical) of an experimental arrangement.

While undoubtedly safe from contradiction, due to its extreme conservatism, it is perhaps overcautious. We do not believe that the primary purpose of theoretical physics is to construct "safe" theories at severe cost in the applicability of their concepts, which is a sterile occupation, but to make useful models which serve for a time and are replaced as they are outworn.[2]

Another objectionable feature of this position is its strong reliance upon the classical level from the outset, which precludes any possibility of explaining this level on the basis of an underlying quantum theory. (The deduction of classical phenomena from quantum theory is impossible simply because no meaningful statements can be made without pre-existing classical apparatus to serve as a reference frame.) This interpretation suffers from the dualism of adhering to a "reality" concept (i.e., the possibility of objective description) on the classical level but renouncing the same in the quantum domain.

c. *The "hidden variables" interpretation.* This is the position (Alternative 4 of the Introduction) that ψ is not a complete description of a single system. It is assumed that the correct complete description, which would involve further (hidden) parameters, would lead to a deterministic theory, from which the probabilistic aspects arise as a result of our ignorance of these extra parameters in the same manner as in classical statistical mechanics.

---

2 Cf. Appendix II.

The ψ-function is therefore regarded as a description of an *ensemble* of systems rather than a single system. Proponents of this interpretation include Einstein,[3] Bohm,[4] Wiener and Siegal.[5]

Einstein hopes that a theory along the lines of his general relativity, where all of physics is reduced to the geometry of space-time could satisfactorily explain quantum effects. In such a theory a particle is no longer a simple object but possesses an enormous amount of structure (i.e., it is thought of as a region of space-time of high curvature). It is conceivable that the interactions of such "particles" would depend in a sensitive way upon the details of this structure, which would then play the role of the "hidden variables."[6] However, these theories are non-linear and it is enormously difficult to obtain any conclusive results. Nevertheless, the possibility cannot be discounted.

Bohm considers ψ to be a real force field acting on a particle which always has a well-defined position and momentum (which are the hidden variables of this theory). The ψ-field satisfying Schrödinger's equation is pictured as somewhat analogous to the electromagnetic field satisfying Maxwell's equations, although for systems of n particles the ψ-field is in a 3n-dimensional space. With this theory Bohm succeeds in showing that in all actual cases of measurement the best predictions that can be made are those of the usual theory, so that no experiments could ever rule out his interpretation in favor of the ordinary theory. Our main criticism of this view is on the grounds of simplicity – if one desires to hold the view that ψ is a real field then the associated particle is superfluous since, as we have endeavored to illustrate, the pure wave theory is itself satisfactory.

---

3 Einstein [7].

4 Bohm [2].

5 Wiener and Siegal [20].

6 For an example of this type of theory see Einstein and Rosen [9].

Wiener and Siegal have developed a theory which is more closely tied to the formalism of quantum mechanics. From the set N of all non-degenerate linear Hermitian operators for a system having a complete set of eigenstates, a subset I is chosen such that no two members of I commute and every element outside I commutes with at least one element of I. The set I therefore contains precisely one operator for every orientation of the principal axes of the Hilbert space for the system. It is postulated that each of the operators of I corresponds to an independent observable which can take any of the real numerical values of the spectrum of the operator. This theory, in its present form, is a theory of infinitely[7] many "hidden variables," since a system is pictured as possessing (at each instant) a value for every one of these "observables" simultaneously, with the changes in these values obeying precise (deterministic) dynamical laws. However, the change of any one of these variables with time depends upon the entire set of observables, so that it is impossible ever to discover by measurement the complete set of values for a system (since only one "observable" at a time can be observed). Therefore, statistical ensembles are introduced, in which the values of all of the observables are related to points in a "differential space," which is a Hilbert space containing a measure for which each (differential space) coordinate has an independent normal distribution. It is then shown that the resulting statistical dynamics is in accord with the usual form of quantum theory.

It cannot be disputed that these theories are often appealing, and might conceivably become important should future discoveries indicate serious inadequacies in the present scheme (i.e., they might be more easily modified to encompass new experience). But from our viewpoint they are usually more cumbersome than the conceptually simpler theory based on pure wave mechanics. Nevertheless, these theories are of great theoretical importance because they provide us with examples that "hidden variables" theories are indeed possible.

---

[7] A non-denumerable infinity, in fact, since the set I is uncountable!

d. *The stochastic process interpretation.* This is the point of view which holds that the fundamental processes of nature are stochastic (i.e., probabilistic) processes. According to this picture physical systems are supposed to exist at all times in definite states, but the states are continually undergoing probabilistic changes. The discontinuous probabilistic "quantum-jumps" are not associated with acts of observation, but are fundamental to the systems themselves.

A stochastic theory which emphasizes the particle, rather than wave, aspects of quantum theory has been investigated by Bopp.[8] The particles do not obey deterministic laws of motion, but rather probabilistic laws, and by developing a general "correlation statistics" Bopp shows that his quantum scheme is a special case which gives results in accord with the usual theory. (This accord is only approximate and in principle one could decide between the theories. The approximation is so close, however, that it is hardly conceivable that a decision would be practically feasible.)

Bopp's theory seems to stem from a desire to have a theory founded upon particles rather than waves, since it is this particle aspect (highly localized phenomena) which is most frequently encountered in present day high-energy experiments (cloud chamber tracks, etc.). However, it seems to us to be much easier to understand particle aspects from a wave picture (concentrated wave packets) than it is to understand wave aspects (diffraction, interference, etc.) from a particle picture.

Nevertheless, there can be no fundamental objection to the idea of a stochastic theory, except on grounds of a naked prejudice for determinism. The question of determinism or indeterminism in nature is obviously forever undecidable in physics, since for any current deterministic [probabilistic] theory one could always postulate that a refinement of the theory

---

8 Bopp [5].

would disclose a probabilistic [deterministic] substructure, and that the current deterministic [probabilistic] theory is to be explained in terms of the refined theory on the basis of the law of large numbers [ignorance of *hidden variables*]. *However, it is quite another matter to object to a mixture of the two where the probabilistic processes occur only with acts of observation.*

e. *The wave interpretation.* This is the position proposed in the present thesis, in which the wave function *itself is held to be the* fundamental entity, obeying at all times a deterministic wave equation.

This view also corresponds most closely with that held by Schrödinger.[9] However, this picture only makes sense when observation processes themselves are treated within the theory. It is only in this manner that the *apparent* existence of definite macroscopic objects, as well as localized phenomena, such as tracks in cloud chambers, can be satisfactorily ex*plained in a wave theory where the waves are continually diffusing.* With the deduction in this theory that phenomena will appear to observers to be subject to Process 1, Heisenberg's criticism[10] of Schrödinger's opinion – that continuous wave mechanics could not seem to explain the discontinuities which are everywhere observed – is effectively met. The "quantum-jumps" exist in our theory as *relative* phenomena (i.e., the states of an object-system relative to chosen observer states show this effect), while the absolute states change quite continuously.

The wave theory is definitely tenable and forms, we believe, the simplest complete, self-consistent theory.

---

9 Schrodinger [18].

10 Heisenberg [14].

We should like now to comment on some views expressed by Einstein. Einstein's[11] criticism of quantum theory (which is actually directed more against what we have called the "popular" view than Bohr's interpretation) is mainly concerned with the drastic changes of state brought about by simple acts of observation (i.e., the infinitely rapid collapse of wave functions), particularly in connection with correlated systems which are widely separated so as to be mechanically uncoupled at the time of observation.[12] At another time he put his feeling colorfully by stating that he could not believe that a mouse could bring about drastic changes in the universe simply by looking at it.[13]

However, from the standpoint of our theory, *it is not so much the system which is affected by an observation as the observer, who becomes correlated to the system.*

In the case of observation of one system of a pair of spatially separated, correlated systems, nothing happens to the remote system to make any of its states more "real" than the rest. It had no independent states to begin with, but a number of states occurring in a superposition with corresponding states for the other (near) system. Observation of the near system simply correlates the observer to this system, a purely local process – but a process which also entails automatic correlation with the remote system. Each state of the remote system still exists with the same amplitude in a superposition, but now a superposition for which element contains, in addition to a remote system state and correlated near system state, an observer state which describes an observer who perceives the state of the near system.[14] From the present viewpoint all elements of

---

11 Einstein [7].

12 For example, the paradox of Einstein, Rosen, and Podolsky [8].

13 Address delivered at Palmer Physical Laboratory, Princeton, Spring, 1954.

14 See in this connection Chapter IV, particularly pp. 82, 83.

this superposition are equally "real." Only the observer state has changed, so as to become correlated with the state of the near system and hence naturally with that of the remote system also. The mouse does not affect the universe – only the mouse is affected.

Our theory in a certain sense bridges the positions of Einstein and Bohr, since the complete theory is quite objective and deterministic ("God does not play dice with the universe"), and yet on the subjective level, of assertions relative to observer states, it is probabilistic in the *strong sense* that there is no way for observers to make any predictions better than the limitations imposed by the uncertainty principle.[15]

In conclusion, we have seen that if we wish to adhere to objective descriptions then the principle of the psycho-physical parallelism requires that we should be able to consider some mechanical devices as representing observers. The situation is then that such devices must either cause the probabilistic discontinuities of Process 1, or must be transformed into the superpositions we have discussed. We are forced to abandon the former possibility since it leads to the situation that some physical systems would obey different laws from the rest, with no clear means for distinguishing between these two types of systems. We are thus led to our present theory which results from the complete abandonment of Process 1 as a basic process. Nevertheless, within the context of this theory, which is objectively deterministic, it develops that the probabilistic aspects of Process 1 reappear at the subjective level, as relative phenomena to observers.

One is thus free to build a conceptual model of the universe, which postulates only the existence of a universal wave function which obeys a linear wave equation. One then investigates the internal correlations in this wave function with the aim of deducing laws of physics, which are

---

15 Cf. Chapter V, §2.

statements that take the form: Under the conditions C the property A of a subsystem of the universe (subset of the total collection of coordinates for the wave function) is correlated with the property B of another subsystem (with the manner of correlation being specified). For example, the classical mechanics of a system of massive particles becomes a law which expresses the correlation between the positions and momenta (approximate) of the particles at one time with those at another time.[16] All statements about subsystems then become *relative* statements, i.e., statements about the subsystem relative to a prescribed state for the remainder (since this is generally the only way a subsystem even possesses a unique state), and all laws are correlation laws.

The theory based on pure wave mechanics is a conceptually simple causal theory, which fully maintains the principle of the psycho-physical parallelism. It therefore forms a framework in which it is possible to discuss (in addition to ordinary phenomena) observation processes themselves, including the inter-relationships of several observers, in a logical, unambiguous fashion. In addition, all of the correlation paradoxes, like that of Einstein, Rosen, and Podolsky,[17] find easy explanation.

While our theory justifies the personal use of the probabilistic interpretation as an aid to making practical predictions, it forms a broader frame in which to understand the consistency of that interpretation. It transcends the probabilistic theory, however, in its ability to deal logically with questions of imperfect observation and approximate measurement.

Since this viewpoint will be applicable to all forms of quantum mechanics which maintain the superposition principle, it may prove a fruitful framework for the interpretation of new quantum formalisms. Field theories, particularly any which might be relativistic in the sense of general rela-

---

[16] Cf. Chapter V, §2.

[17] Einstein, Rosen, and Podolsky [8].

tivity, might benefit from this position, since one is free to construct formal (non-probabilistic) theories, and supply any possible statistical interpretations later. (This viewpoint avoids the necessity of considering anomalous probabilistic jumps scattered about space-time, and one can assert that field equations are satisfied everywhere and everywhen, then *deduce* any statistical assertions by the present method.)

By focusing attention upon questions of correlations, one may be able to deduce useful relations (correlation laws analogous to those of classical mechanics) for theories which at present do not possess known classical counterparts. Quantized fields do not generally possess pointwise independent field values, the values at one point of space-time being correlated with those at neighboring points of space-time in a manner, it is to be expected, approximating the behavior of their classical counterparts. If correlations are important in systems with only a finite number of degrees of freedom, how much more important they must be for systems of infinitely many coordinates.

Finally, aside from any possible practical advantages of the theory, it remains a matter of intellectual interest that the statistical assertions of the usual interpretation do not have the status of independent hypotheses, but are deducible (in the present sense) from the pure wave mechanics, which results from their omission.

APPENDIX I

We shall now supply the proofs of a number of assertions which have been made in the text.

§1. *Proof of Theorem* 1

We now show that $\{X,Y,\ldots,Z\} > 0$ unless $X,Y,\ldots,Z$ are independent random variables. Abbreviate $P(x_i,y_j,\ldots,z_k)$ by $P_{ij\ldots k}$, and let

$$Q_{ij\ldots k} = \begin{cases} \dfrac{P_{ij\ldots k}}{P_iP_j\ldots P_k} & \text{if } P_iP_j\ldots P_k > 0 \\ 1 & \text{if } P_iP_j\ldots P_k = 0 \end{cases} \tag{1.1}$$

(Note that $P_iP_j\ldots P_k = 0$ implies that also $P_{ij\ldots k} = 0$.) Then always

$$P_{ij\ldots k} = Q_{ij\ldots k}P_iP_j\ldots P_k\ , \tag{1.2}$$

and we have

$$\{X,Y,\ldots,Z\} = \operatorname{Exp}\left[\ln \frac{P_{ij\ldots k}}{P_iP_j\ldots P_k}\right] = \operatorname{Exp}\left[\ln Q_{ij\ldots k}\right] \tag{1.3}$$

$$= \sum_{ij\ldots k} P_iP_j\ldots P_k\, Q_{ij\ldots k} \ln Q_{ij\ldots k}\ .$$

Applying the inequality for $x \geqq 0$:

$$x \ln x > x - 1 \qquad \text{(except for } x = 1) \tag{1.4}$$

(which is easily established by calculating the minimum of $x \ln x - (x-1)$) to (1.3) we have:

$$(1.5) \qquad P_i P_j \ldots P_k \, Q_{ij\ldots k} \ln Q_{ij\ldots k} > P_i P_j \ldots P_k (Q_{ij\ldots k} - 1)$$

$$(\text{unless } Q_{ij\ldots k} = 1) \ .$$

Therefore we have for the sum:

$$(1.6) \sum_{ij\ldots k} P_i P_j \ldots P_k Q_{ij\ldots k} \ln Q_{ij\ldots k} > \sum_{ij\ldots k} P_i P_j \ldots P_k Q_{ij\ldots k} - \sum_{ij\ldots k} P_i P_j \ldots P_k \ ,$$

unless *all* $Q_{ij\ldots k} = 1$. But $\sum_{ij\ldots k} P_i P_j \ldots P_k \, Q_{ij\ldots k} = \sum_{ij\ldots k} P_{ij\ldots k} = 1$, and $\sum_{ij\ldots k} P_i P_j \ldots P_k = 1$, so that the right side of (1.6) vanishes. The left side is, by (1.3) the correlation $\{X,Y,\ldots,Z\}$, and the condition that all of the $Q_{ij\ldots k}$ equal one is precisely the independence condition that $P_{ij\ldots k} = P_i P_j \ldots P_k$ for all $i,j,\ldots,k$. We have therefore proved that

$$(1.7) \qquad \{X,Y,\ldots,Z\} > 0$$

unless $X,Y,\ldots,Z$ are mutually independent.

§2. *Convex function inequalities*

We shall now establish some basic inequalities which follow from the convexity of the function $x \ln x$.

LEMMA 1.

$$x_i \geqq 0, \quad P_i \geqq 0, \quad \sum_i P_i = 1$$

$$\Rightarrow \left(\sum_i P_i x_i\right) \ln \left(\sum_i P_i x_i\right) \leqq \sum_i P_i x_i \ln x_i \ .$$

This property is usually taken as the definition of a convex function,[1] but follows from the fact that the second derivative of $x \ln x$ is positive for all positive x, which is the elementary notion of convexity. There is also an immediate corollary for the continuous case:

[1] See Hardy, Littlewood, and Pólya [13], p. 70.

COROLLARY 1. $g(x) \geqq 0, \quad P(x) \geqq 0, \quad \int P(x)\,dx = 1$

$$\Rightarrow \left[\int P(x)\,g(x)\,dx\right] \ln \left[\int P(x)\,g(x)\,dx\right] \leqq \int P(x)\,g(x) \ln g(x)\,dx .$$

We can now derive a more general and very useful inequality from Lemma 1:

LEMMA 2. $x_i \geqq 0, \quad a_i \geqq 0 \quad (\text{all } i)$

$$\Rightarrow \left(\sum_i x_i\right) \ln\left(\frac{\sum_i x_i}{\sum_i a_i}\right) \leqq \sum_i x_i \ln\left(\frac{x_i}{a_i}\right) .$$

Proof: Let $P_i = a_i / \sum_i a_i$, so that $P_i \geqq 0$ and $\sum_i P_i = 1$. Then by Lemma 1:

$$(2.1) \qquad \left[\sum_i P_i\left(\frac{x_i}{a_i}\right)\right] \ln\left[\sum_i P_i\left(\frac{x_i}{a_i}\right)\right] \leqq \sum_i P_i\left(\frac{x_i}{a_i}\right) \ln\left(\frac{x_i}{a_i}\right) .$$

Substitution for P_i yields:

$$(2.2) \left[\sum_i \frac{a_i}{\left(\sum_i a_i\right)}\left(\frac{x_i}{a_i}\right)\right] \ln\left[\sum_i \frac{a_i}{\left(\sum_i a_i\right)}\left(\frac{x_i}{a_i}\right)\right] \leqq \sum_i \frac{a_i}{\left(\sum_i a_i\right)}\left(\frac{x_i}{a_i}\right) \ln\left(\frac{x_i}{a_i}\right),$$

which reduces to

$$(2.3) \qquad \left(\sum_i x_i\right) \ln\left(\frac{\sum_i x_i}{\sum_i a_i}\right) \leqq \sum_i x_i \ln\left(\frac{x_i}{a_i}\right) ,$$

and we have proved the lemma.

We also mention the analogous result for the continuous case:

COROLLARY 2. $f(x) \geqq 0, \quad g(x) \geqq 0 \quad (\text{all } x)$

$$\Rightarrow \left[\int f(x)\,dx\right] \ln \left[\frac{\int f(x)\,dx}{\int g(x)\,dx}\right] \leqq \int f(x) \ln\left(\frac{f(x)}{g(x)}\right) dx .$$

§3. *Refinement theorems*

We now supply the proof for Theorems 2 and 4 of Chapter II, which concern the behavior of correlation and information upon refinement of the distributions. We suppose that the original (unrefined) distribution is $P_{ij\ldots k} = P(x_i, y_j, \ldots, z_k)$, and that the *refined* distribution is $P'^{\mu_i,\nu_j,\ldots,\eta_k}_{ij\ldots k}$, where the original value x_i for X has been resolved into a number of values $x_i^{\mu_i}$, and similarly for Y,...,Z. Then:

$$(3.1) \qquad P_{ij\ldots k} = \sum_{\mu_i,\nu_j,\ldots,\eta_k} P'^{\mu_i,\nu_j,\ldots,\eta_k}_{ij\ldots k}, \quad P_i = \sum_{\mu_i} P_i'^{\mu_i}, \quad \text{etc.}$$

Computing the new correlation $\{X,Y,\ldots,Z\}'$ for the refined distribution $P'^{\mu_i,\nu_j,\ldots,\eta_k}_{ij\ldots k}$ we find:

$$(3.2) \quad \{X,Y,\ldots,Z\}' = \sum_{ij\ldots k} \sum_{\mu_i,\nu_j,\ldots,\eta_k} P'^{\mu_i,\nu_j,\ldots,\eta_k}_{ij\ldots k} \ln\left(\frac{P'^{\mu_i,\nu_j,\ldots,\eta_k}_{ij\ldots k}}{P_i'^{\mu_i} P_j'^{\nu_j} \cdots P_k'^{\eta_k}}\right).$$

However, by Lemma 2, §2:

$$(3.3) \qquad \left(\sum_{\mu_i\cdots\eta_k} P'^{\mu_i\cdots\eta_k}_{i\ldots k}\right) \ln\left(\frac{\sum_{\mu_i\cdots\eta_k} P'^{\mu_i\cdots\eta_k}_{i\ldots k}}{\sum_{\mu_i\cdots\eta_k} P_i'^{\mu_i} P_j'^{\nu_j} \cdots P_k'^{\eta_k}}\right)$$

$$\leqq \sum_{\mu_i\cdots\eta_k} P'^{\mu_i\cdots\eta_k}_{i\ldots k} \ln\left(\frac{P'^{\mu_i\cdots\eta_k}_{i\ldots k}}{P_i'^{\mu_i} P_j'^{\nu_j} \cdots P_k'^{\eta_k}}\right).$$

Substitution of (3.3) into (3.2), noting that $\sum_{\mu_i \cdots \eta_k} P'^{\mu_i}_i, P'^{\nu_j}_j, \ldots, P'^{\eta_k}_k$ is equal to $\left(\sum_{\mu_i} P'^{\mu_i}_i\right)\left(\sum_{\nu_j} P'^{\nu_j}_j\right)\cdots\left(\sum_{\eta_k} P'^{\eta_k}_k\right)$, leads to:

(3.4)

$$\{X,Y,\ldots,Z\}' \geqq \left(\sum_{ij\ldots k} \sum_{\mu_i \cdots \eta_k} P'^{\mu_i \cdots \eta_k}_{ij\ldots k}\right) \ln \left[\frac{\sum_{\mu_i \cdots \eta_k} P'^{\mu_i \cdots \eta_k}_{ij\ldots k}}{\left(\sum_{\mu_i} P'^{\mu_i}_i\right)\left(\sum_{\nu_j} P'^{\nu_j}_j\right)\cdots\left(\sum_{\eta_k} P'^{\eta_k}_k\right)}\right]$$

$$= \sum_{ij\ldots k} P_{ij\ldots k} \ln \frac{P_{ij\ldots k}}{P_i P_j \cdots P_k} = \{X,Y,\ldots,Z\} ,$$

and we have completed the proof of Theorem 2 (Chapter II), which asserts that refinement never decreases the correlation.[2]

We now consider the effect of refinement upon the relative information. We shall use the previous notation, and further assume that $a'^{\mu_i}_i, b'^{\nu_j}_j, \ldots, c'^{\eta_k}_k$ are the information measures for which we wish to compute the relative information of $P'^{\mu_i,\nu_j,\ldots,\eta_k}_{ij\ldots k}$ and of $P_{ij\ldots k}$. The information measures for the unrefined distribution $P_{ij\ldots k}$ then satisfy the relations:

$$a_i = \sum_{\mu_i} a'^{\mu_i}_i , \quad b_j = \sum_{\nu_j} b'^{\nu_j}_j , \ldots . \tag{3.5}$$

The relative information of the refined distribution is

$$I'_{XY\ldots Z} = \sum_{i\ldots j} \sum_{\mu_i \cdots \eta_k} P'^{\mu_i \cdots \eta_k}_{ij\ldots k} \ln \left[\frac{P'^{\mu_i \cdots \eta_k}_{ij\ldots k}}{a'^{\mu_i}_i, b'^{\nu_j}_j, \ldots, c'^{\eta_k}_k}\right] , \tag{3.6}$$

and by exactly the same procedure as we have just used for the correlation we arrive at the result:

---

2 Cf. Shannon [19], Appendix 7, where a quite similar theorem is proved.

$$I'_{XY\ldots Z} \geqq \sum_{i\ldots k} P_{ij\ldots k} \ln \frac{P_{ij\ldots k}}{a_i b_j \ldots c_k} = I_{XY\ldots Z} \quad , \tag{3.7}$$

and we have proved that refinement never decreases the relative information (Theorem 4, Chapter II).

It is interesting to note that the relation (3.4) for the behavior of correlation under refinement can be deduced from the behavior of relative information, (3.7). This deduction is an immediate consequence of the fact that the correlation is a relative information – the information of the *joint distribution* relative to the product measure of the *marginal distributions.*

§4. *Monotone decrease of information for stochastic processes*

We consider a sequence of transition-probability matrices T^n_{ij} ($\sum_j T^n_{ij} = 1$ for all n, i, and $0 \leqq T^n_{ij} \leqq 1$ for all n, i, j), and a sequence of measures a^n_i ($a^n_i \geqq 0$) having the property that

$$a^{n+1}_j = \sum_i a^n_i T^n_{ij} \quad . \tag{4.1}$$

We further suppose that we have a sequence of probability distributions, P^n_i, such that

$$P^{n+1}_j = \sum_i P^n_i T^n_{ij} \quad . \tag{4.2}$$

For each of these probability distributions the relative information I^n (relative to the a^n_i measure) is defined:

$$I^n = \sum_i P^n_i \ln \left(\frac{P^n_i}{a^n_i} \right) \quad . \tag{4.3}$$

Under these circumstances we have the following theorem:

THEOREM. $$I^{n+1} \leqq I^n \quad .$$

Proof: Expanding I^{n+1} we get:

$$(4.4) \qquad I^{n+1} = \sum_j P_j^{n+1} \ln\left(\frac{P_j^{n+1}}{a_j^{n+1}}\right) = \sum_j \left(\sum_i P_i^n T_{ij}^n\right) \ln \frac{\left(\sum_i P_i^n T_{ij}^n\right)}{\left(\sum_i a_i^n T_{ij}^n\right)} .$$

However, by Lemma 2 (§2, Appendix I) we have the inequality

$$(4.5) \qquad \left(\sum_i P_i^n T_{ij}^n\right) \ln \frac{\left(\sum_i P_i^n T_{ij}^n\right)}{\left(\sum_i a_i^n T_{ij}^n\right)} \leqq \sum_i P_i^n T_{ij}^n \ln \frac{P_i^n T_{ij}^n}{a_i^n T_{ij}^n} .$$

Substitution of (4.5) into (4.4) yields:

$$(4.6) \qquad I^{n+1} \leqq \sum_j \left(\sum_i P_i^n T_{ij}^n \ln \frac{P_i^n}{a_i^n}\right) = \sum_i P_i^n \left(\sum_j T_{ij}^n\right) \ln\left(\frac{P_i^n}{a_i^n}\right)$$

$$= \sum_i P_i^n \ln\left(\frac{P_i^n}{a_i^n}\right) = I^n ,$$

and the proof is completed.

This proof can be successively specialized to the case where T is stationary ($T_{ij}^n = T_{ij}$ for all n) and then to the case where T is doubly-stochastic ($\sum_i T_{ij} = 1$ for all j):

COROLLARY 1. T_{ij}^n *is stationary* ($T_{ij}^n = T_{ij}$, *all* n), *and the measure* a_i *is a stationary measure* ($a_j = \sum_i a_i T_{ij}$), *imply that the information,* $I^n = \sum_i P_i^n \ln (P_i^n/a_i^n)$, *is monotone decreasing. (As before,* $P_j^{n+1} = \sum_i P_i^n T_{ij}^n$.)

Proof: Immediate consequence of preceding theorem.

COROLLARY 2. T_{ij} *is doubly-stochastic* ($\sum_i T_{ij} = 1$, *all* j) *implies that the information relative to the uniform measure* ($a_i = 1$, *all* i), $I^n = \sum_i P_i^n \ln P_i^n$, *is monotone decreasing.*

Proof: For $a_i = 1$ (all i) we have that $\sum_i a_i T_{ij} = \sum_i T_{ij} = 1 = a_j$. Therefore the uniform measure is stationary in this case and the result follows from Corollary 1.

These results hold for the continuous case also, and may be easily verified by replacing the above summations by integrations, and by replacing Lemma 2 by its corollary.

§5. *Proof of special inequality for Chapter IV* (1.7)

LEMMA. *Given probability densities* $P(r)$, $P_1(x)$, $P_2(r)$, *with* $P(r) = \int P_1(x) P_2(r-x\tau)\,dx$. *Then* $I_R \leqq I_X - \ln\tau$, *where* $I_X = \int P_1(x) \ln P_1(x)\,dx$ *and* $I_R = \int P(r) \ln P(r)\,dr$.

Proof: We first note that:

$$\int P_2(r-x\tau)\,dx = \int P_2(\omega)\,\frac{d\omega}{\tau} = \frac{1}{\tau} \qquad \text{(all r)} \tag{5.1}$$

and that furthermore

$$\int P_2(r-x\tau)\,dr = \int P_2(\omega)\,d\omega = 1 \qquad \text{(all x)}\,. \tag{5.2}$$

We now define the density $\widetilde{P}^r(x)$:

$$\widetilde{P}^r(x) = \tau P_2(r-x\tau)\,, \tag{5.3}$$

which is normalized, by (5.1). Then, according to §2, Corollary 1 Appendix I), we have the relation:

$$(5.4)\quad \left(\int \tilde{P}^r(x)P_1(x)\,dx\right) \ln \left(\int \tilde{P}^r(x)P_1(x)\,dx\right) \leqq \int \tilde{P}^r(x)P_1(x)\,dx \ .$$

Substitution from (5.3) gives

$$(5.5)\quad \left(r\int P_2(r-xr)P_1(x)dx\right) \ln \left(r\int P_2(r-xr)P_1(x)dx\right) \leqq r\int P_2(r-xr)P_1(x)\ln P_1(x)\,dx \ .$$

The relation $P(r) = \int P_1(x)P_2(r-xr)dx$, together with (5.5) then implies

$$(5.6)\quad P(r)\ln rP(r) \leqq \int P_2(r-xr)P_1(x)\ln P_1(x)\,dx \ ,$$

which is the same as:

$$(5.7)\quad P(r)\ln P(r) \leqq \int P_2(r-xr)P_1(x)\ln P_1(x)\,dx - P(r)\ln r \ .$$

Integrating with respect to r, and interchanging the order of integration on the right side gives:

$$(5.8)\quad I_R = \int P(r)\ln P(r)\,dr \leqq \int\left[\int P_2(r-xr)\,dr\right]P_1(x)\ln P_1(x)\,dx - (\ln r)\int P(r)\,dr \ .$$

But using (5.2) and the fact that $\int P(r)\,dr = 1$ this means that

$$(5.9)\quad I_R \leqq \int P_1(x)\ln P_1(x)\,dx - \ln r = I_X - \ln r \ ,$$

and the proof of the lemma is completed.

§6. *Stationary point of* $I_K + I_X$

We shall show that the information sum:

$$(6.1)\quad I_K + I_X = \int_{-\infty}^{\infty} \phi^*\phi(k)\ln\phi^*\phi(k)\,dk + \int_{-\infty}^{\infty} \psi^*\psi(x)\ln\psi^*\psi(x)\,dx \ ,$$

where

$$\phi(k) = (1/\sqrt{2\pi})\int_{-\infty}^{\infty} e^{-ikx}\,\psi(x)\,dx$$

is *stationary* for the functions:

$$\psi_0(x) = (1/2\pi\sigma_x^2)^{\frac{1}{4}}\, e^{-x^2/4\sigma^2 x}\,, \quad \phi_0(k) = (2\sigma_x^2/\pi)^{\frac{1}{4}}\, e^{-k^2\sigma^2 x}\,, \tag{6.2}$$

with respect to variations of ψ, $\delta\psi$, which preserve the normalization:

$$\int_{-\infty}^{\infty} \delta(\psi^*\psi)\,dx = 0\,. \tag{6.3}$$

The variation $\delta\psi$ gives rise to a variation $\delta\phi$ of $\phi(k)$:

$$\delta\phi = (1/\sqrt{2\pi})\int_{-\infty}^{\infty} e^{-ikx}\delta\psi\,dx\,. \tag{6.4}$$

To avoid duplication of effort we first calculate the variation δI_ξ for an arbitrary wave function $u(\xi)$. By definition,

$$I_\xi = \int_{-\infty}^{\infty} u^*(\xi)\,u(\xi)\,\ln u^*(\xi)\,u(\xi)\,d\xi\,, \tag{6.5}$$

so that

$$\begin{aligned}\delta I_\xi &= \int_{-\infty}^{\infty} [u^*u\,\delta(\ln u^*u) + \delta(u^*u)\ln u^*u]\,d\xi \\ &= \int_{-\infty}^{\infty} (1 + \ln u^*u)(u^*\delta u\; u\delta u^*)\,d\xi\,.\end{aligned} \tag{6.6}$$

We now suppose that u has the *real* form:

$$u(\xi) = a\, e^{-b\xi^2} = u^*(\xi)\,, \tag{6.7}$$

and from (6.6) we get

$$\delta I_\xi = \int_{-\infty}^{\infty} (1 + \ln a^2 - 2b\xi^2)\,a\,e^{-b\xi^2}(\delta u)\,d\xi + \text{complex conjugate.} \tag{6.8}$$

We now compute δI_K for ϕ_0 using (6.8), (6.2), and (6.4):

$$\delta I_K\Big|_{\phi_0} = \int_{-\infty}^{\infty} (1 + \ln a'^2 - 2b'k^2)a'e^{-b'k^2}\,\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ikx}\delta\psi\,dx\,dk + \text{c.c.}\,, \tag{6.9}$$

where

$$a = (2\sigma_x^2/\pi)^{\frac{1}{4}}, \quad b' = \sigma_x^2 .$$

Interchanging the order of integration and performing the definite integration over k we get:

$$(6.10) \qquad \delta I_K\Big|_{\phi_0} = \int_{-\infty}^{\infty} \frac{a'}{\sqrt{2b'}} \left(\ln a'^2 + \frac{x^2}{2b'}\right) e^{-(x^2/4b')} \delta\psi(x)\,dx + \text{c.c.} ,$$

while application of (6.8) to ψ_0 gives

$$(6.11) \qquad \delta I_X\Big|_{\psi_0} = \int_{-\infty}^{\infty} (1 + \ln a''^2 - 2b''x^2)\, a'' e^{-b''x^2} \delta\psi(x)\,dx + \text{c.c.} ,$$

where

$$a'' = (1/2\pi\sigma_x^2)^{\frac{1}{4}}, \quad b'' = (1/4\sigma_x^2) .$$

Adding (6.10) and (6.11), and substituting for a', b', a'', b'', yields:

$$(6.12) \quad \delta(I_K + I_X)\Big|_{\psi_0} = (1-\ln\pi) \int_{-\infty}^{\infty} (1/2\pi\sigma_x^2)^{\frac{1}{4}} e^{-(x^2/4\sigma_x^2)} \delta\psi(x)\,dx + \text{c.c.} .$$

But the integrand of (6.12) is simply $\psi_0(x)\delta\psi(x)$, so that

$$(6.13) \qquad \delta(I_K + I_X)\Big|_{\psi_0} = (1-\ln\pi) \int_{-\infty}^{\infty} \psi_0 \delta\psi\,dx + \text{c.c.} .$$

Since ψ_0 is real, $\psi_0\delta\psi + \text{c.c.} = \psi_0^*\delta\psi + \text{c.c.} = \psi_0^*\delta\psi + \psi_0\delta\psi^* = \delta(\psi^*\psi)$, so that

$$(6.14) \qquad \delta(I_K + I_X)\Big|_{\psi_0} = (1-\ln\pi) \int_{-\infty}^{\infty} \delta(\psi^*\psi)\,dx = 0 ,$$

due to the normality restriction (6.3), and the proof is completed.

APPENDIX II

REMARKS ON THE ROLE OF THEORETICAL PHYSICS

There have been lately a number of new interpretations of quantum mechanics, most of which are equivalent in the sense that they predict the same results for all physical experiments. Since there is therefore no hope of deciding among them on the basis of physical experiments, we must turn elsewhere, and inquire into the fundamental question of the nature and purpose of physical theories in general. Only after we have investigated and come to some sort of agreement upon these general questions, i.e., of the role of theories themselves, will we be able to put these alternative interpretations in their proper perspective.

Every theory can be divided into two separate parts, the formal part, and the interpretive part. The formal part consists of a purely logico-mathematical structure, i.e., a collection of symbols together with rules for their manipulation, while the interpretive part consists of a set of "associations," which are rules which put some of the elements of the formal part into correspondence with the perceived world. The essential point of a theory, then, is that it is a *mathematical model*, together with an *isomorphism*[1] between the model and the world of experience (i.e., the sense perceptions of the individual, or the "real world" – depending upon one's choice of epistemology).

---

[1] By isomorphism we mean a mapping of some elements of the model into elements of the perceived world which has the property that the model is faithful, that is, if in the model a symbol A implies a symbol B, and A corresponds to the happening of an event in the perceived world, then the event corresponding to B must also obtain. The word homomorphism would be technically more correct, since there may not be a one-one correspondence between the model and the external world.

The model nature is quite apparent in the newest theories, as in nuclear physics, and particularly in those fields outside of physics proper, such as the Theory of Games, various economic models, etc., where the degree of applicability of the models is still a matter of considerable doubt. However, when a theory is highly successful and becomes firmly established, the model tends to become identified with "reality" itself, and the model nature of the theory becomes obscured. The rise of classical physics offers an excellent example of this process. The constructs of classical physics are just as much fictions of our own minds as those of any other theory we simply have a great deal more confidence in them. It must be deemed a mistake, therefore, to attribute any more "reality" here than elsewhere.

Once we have granted that any physical theory is essentially only a model for the world of experience, we must renounce all hope of finding anything like "*the* correct theory." There is nothing which prevents any number of quite distinct models from being in correspondence with experience (i.e., all "correct"), and furthermore no way of ever verifying that any model is completely correct, simply because the totality of all experience is never accessible to us.

Two types of prediction can be distinguished; the prediction of phenomena already understood, in which the theory plays simply the role of a device for compactly summarizing known results (the aspect of most interest to the engineer), and the prediction of new phenomena and effects, unsuspected before the formulation of the theory. Our experience has shown that a theory often transcends the restricted field in which it was formulated. It is this phenomenon (which might be called the "inertia" of theories) which is of most interest to the theoretical physicist, and supplies a greater motive to theory construction than that of aiding the engineer.

From the viewpoint of the first type of prediction we would say that the "best" theory is the one from which the most accurate predictions can be most easily deduced – two not necessarily compatible ideals.

Classical physics, for example, permits deductions with far greater ease than the more accurate theories of relativity and quantum mechanics, and in such a case we must retain them all. It would be the worst sort of folly to advocate that the study of classical physics be completely dropped in favor of the newer theories. It can even happen that several quite distinct models can exist which are completely equivalent in their predictions, such that different ones are most applicable in different cases, a situation which seems to be realized in quantum mechanics today. It would seem foolish to attempt to reject all but one in such a situation, where it might be profitable to retain them all.

Nevertheless, we have a strong desire to construct a single all-embracing theory which would be applicable to the entire universe. From what stems this desire? The answer lies in the second type of prediction – the discovery of new phenomena – and involves the consideration of inductive inference and the factors which influence our *confidence* in a given theory (to be applicable outside of the field of its formulation). This is a difficult subject, and one which is only beginning to be studied seriously. Certain main points are clear, however, for example, that our confidence increases with the number of successes of a theory. If a new theory replaces several older theories which deal with separate phenomena, i.e., a comprehensive theory of the previously diverse fields, then our confidence in the new theory is very much greater than the confidence in either of the older theories, since the range of success of the new theory is much greater than any of the older ones. It is therefore this factor of confidence which seems to be at the root of the desire for comprehensive theories.

A closely related criterion is *simplicity* – by which we refer to conceptual simplicity rather than ease in use, which is of paramount interest to the engineer. A good example of the distinction is the theory of general relativity which is conceptually quite simple, while enormously cumbersome in actual calculations. Conceptual simplicity, like comprehensiveness, has the property of increasing confidence in a theory. A theory

containing many *ad hoc* constants and restrictions, or many independent hypotheses, in no way impresses us as much as one which is largely free of arbitrariness.

It is necessary to say a few words about a view which is sometimes expressed, the idea that a physical theory should contain no elements which do not correspond directly to observables. This position seems to be founded on the notion that the only purpose of a theory is to serve as a summary of known data, and overlooks the second major purpose, the discovery of totally new phenomena. The major motivation of this viewpoint appears to be the desire to construct perfectly "safe" theories which will never be open to contradiction. Strict adherence to such a philosophy would probably seriously stifle the progress of physics.

The critical examination of just what quantities are observable in a theory does, however, play a useful role, since it gives an insight into ways of modification of a theory when it becomes necessary. A good example of this process is the development of Special Relativity. Such successes of the positivist viewpoint, when used merely as a tool for deciding which modifications of a theory are possible, in no way justify its universal adoption as a general principle which all theories must satisfy.

In summary, a physical theory is a logical construct (model), consisting of symbols and rules for their manipulation, some of whose elements are associated with elements of the perceived world. The fundamental requirements of a theory are logical consistency and correctness. There is no reason why there cannot be any number of different theories satisfying these requirements, and further criteria such as usefulness, simplicity, comprehensiveness, pictorability, etc., must be resorted to in such cases to further restrict the number. Even so, it may be impossible to give a total ordering of the theories according to "goodness," since different ones may rate highest according to the different criteria, and it may be most advantageous to retain more than one.

As a final note, we might comment upon the concept of *causality*. It should be clearly recognized that causality is a property of a model, and

not a property of the world of experience. The concept of causality only makes sense with reference to a theory, in which there are logical dependences among the elements. A theory contains relations of the form "A implies B," which can be read as "A causes B," while our experience, uninterpreted by any theory, gives nothing of the sort, but only a *correlation* between the event corresponding to B and that corresponding to A.

REFERENCES

[1] D. Bohm, *Quantum Theory*. Prentice-Hall, New York: 1951.

[2] D. Bohm, *Phys. Rev.* 84, 166, 1952 and 85, 180, 1952.

[3] N. Bohr, in *Albert Einstein, Philosopher-Scientist*. The Library of Living Philosophers, Inc., Vol. 7, p. 199. Evanston: 1949.

[4] N. Bohr, *Atomic Theory and the Description of Nature.*

[5] F. Bopp, *Z. Naturforsch.* 2a(4), 202, 1947; 7a 82, 1952; 8a, 6, 1953.

[6] J. L. Doob, *Stochastic Processes*. Wiley, New York: 1953.

[7] A. Einstein, in *Albert Einstein, Philosopher-Scientist*. The Library of Living Philosophers, Inc., Vol. 7, p. 665. Evanston: 1949.

[8] A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.* 47, 777, 1935.

[9] A. Einstein, N. Rosen, *Phys. Rev.* 48, 73, 1935.

[10] W. Feller, *An Introduction to Probability Theory and its Applications*. Wiley, New York: 1950.

[11] D. ter Haar, *Elements of Statistical Mechanics*. Rinehart, New York, 1954.

[12] P. R. Halmos, *Measure Theory*. Van Nostrand, New York: 1950.

[13] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*. Cambridge University Press: 1952.

[14] W. Heisenberg, in *Niels Bohr and the Development of Physics*. McGraw-Hill, p. 12. New York: 1955.

[15] J. Kelley, *General Topology*. Van Nostrand, New York: 1955.

[16] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*. (Translated by George Gamow) Dover, New York: 1949.

[17] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*. (Translated by R. T. Beyer) Princeton University Press: 1955.

[18] E. Schrödinger, *Brit. J. Phil. Sci.* 3, 109, 233, 1952.

[19] C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press: 1949.

[20] N. Wiener, I. E. Siegal, *Nuovo Cimento Suppl.* 2, 982 (1955).

[21] P. M. Woodward, *Probability and Information Theory, with Applications to Radar*. McGraw-Hill, New York: 1953.