BRYCE SELIGMAN DEWITT
NEILL GRAHAM

The Many-Worlds Interpretation of Quantum Mechanics

PRINCETON LEGACY LIBRARY

The Many-Worlds Interpretation of Quantum Mechanics

Princeton Series in Physics

edited by Arthur S. Wightman
and John J. Hopfield

Quantum Mechanics for Hamiltonians Defined as Quadratic Forms *by Barry Simon*

Lectures on Current Algebra and Its Applications *by Sam B. Treiman, Roman Jackiw, and David J. Gross*

Physical Cosmology *by P. J. E. Peebles*

The Many-Worlds Interpretation of Quantum Mechanics *edited by B. S. DeWitt and N. Graham*

The Many-Worlds Interpretation of Quantum Mechanics

A Fundamental Exposition by
HUGH EVERETT, III, with Papers by
J. A. WHEELER, B. S. DEWITT,
L. N. COOPER and D. VAN VECHTEN,
and N. GRAHAM

Edited by
BRYCE S. DEWITT and NEILL GRAHAM

Princeton Series in Physics

Princeton University Press
Princeton, New Jersey, 1973

LC Card: 72-12116

ISBN: 0-691-08126-3 (hard cover edition)
ISBN: 0-691-88131-X (paperback edition)

Library of Congress Cataloguing in Publication data will be found on the last printed page of this book.

Printed in the United States of America
by Princeton University Press

PREFACE

In 1957, in his Princeton doctoral dissertation, Hugh Everett, III, proposed a new interpretation of quantum mechanics that denies the existence of a separate classical realm and asserts that it makes sense to talk about a state vector for the whole universe. This state vector never collapses, and hence reality as a whole is rigorously deterministic. This reality, which is described *jointly* by the dynamical variables and the state vector, is not the reality we customarily think of, but is a reality composed of many worlds. By virtue of the temporal development of the dynamical variables the state vector decomposes naturally into orthogonal vectors, reflecting a continual splitting of the universe into a multitude of mutually unobservable but equally real worlds, in each of which every good measurement has yielded a definite result and in most of which the familiar statistical quantum laws hold.

In addition to his short thesis Everett wrote a much larger exposition of his ideas, which was never published. The present volume contains both of these works, together with a handful of papers by others on the same theme. Looked at in one way, Everett's interpretation calls for a return to naive realism and the old fashioned idea that there can be a direct correspondence between formalism and reality. Because physicists have become more sophisticated than this, and above all because the implications of his approach appear to them so bizarre, few have taken Everett seriously. Nevertheless his basic premise provides such a stimulating framework for discussions of the quantum theory of measurement that this volume should be on every quantum theoretician's shelf.

"... a picture, incomplete yet not false, of the universe as Ts'ui Pên conceived it to be. Differing from Newton and Schopenhauer,... [he] did not think of time as absolute and uniform. He believed in an infinite series of times, in a dizzily growing, ever spreading network of diverging, converging and parallel times. This web of time – the strands of which approach one another, bifurcate, intersect or ignore each other through the centuries – embraces *every* possibility. We do not exist in most of them. In some you exist and not I, while in others I do, and you do not, and in yet others both of us exist. In this one, in which chance has favored me, you have come to my gate. In another, you, crossing the garden, have found me dead. In yet another, I say these very same words, but am an error, a phantom."

Jorge Luis Borges, *The Garden of Forking Paths*

"Actualities seem to float in a wider sea of possibilities from out of which they were chosen; and *somewhere*, indeterminism says, such possibilities exist, and form part of the truth."

William James

CONTENTS

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 $\mathrm{Exp}\,[F]$, is defined to be:

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

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

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

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

---

[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:

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

and we note the following easily verified rules for expectations:

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

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

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

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].

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

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 $I_{XY\ldots Z}$:

$$\begin{aligned} I_{XY\ldots Z} &= \sum_{i,j,\ldots,k} P(x_i, y_j, \ldots, z_k) \ln P(x_i, y_j, \ldots, z_k) \\ &= \text{Exp}\,[\ln P(x_i, y_j, \ldots, z_k)] \,, \end{aligned} \tag{2.2}$$

---

[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.)