REVIEWS OF MODERN PHYSICS VOLUME 27, NUMBER 2 APRIL, 1955

Symmetry of Physical Laws. Part III. Prediction and Retrodiction

SATOSI WATANABE
U. S. Naval Postgraduate School Monterey, California

An attempt is made within the framework of the accepted quantum physics to achieve the maximum parallelism between prediction (inference of the future observational data from the present ones) and retrodiction (inference of the past observational data from the present ones). To implement this program, it is shown that the "retrodictive state function" (extrapolation of the present data to the past) can be just as useful as the ordinary "predictive state function" (extrapolation of the present data to the future). This leads to a formalism in which time-reversal becomes a linear transformation and double time-reversal becomes a c-number. In spite of this formal symmetry, it can be shown that the actual success of a retrodiction depends on the satisfaction of an additional condition which is not required in prediction, and which is not always fulfilled. From the same point of view, a logical loophole is pointed out in the indiscriminate application of the H-theorem to the past. The so-called irreversibility of observation is interpreted in terms of the decrease of "information" in the process of inference.

1. INTRODUCTION

IN accordance with its expected role in human activities, physical theory is pre-eminently a predictive instrument. Man is, however, not immune to temptation of the adventure of guessing with the same instrument what happened in the past outside the reach of his own observation. In the nonstatistical domain of classical physics, retrodiction must be in principle just as successful as prediction. However, in statistical applications of classical physics and in quantum physics, a careful study is needed to determine the confirmability of an attempted retrodiction. The present paper is intended to provide an answer to some of the rudimentary questions in this rather neglected field of intellectual interest. Although some new points of view and a new formalism are introduced, the content of this paper will remain perfectly faithful to the accepted premises of classical and quantum physics. It should be noted that retrodiction is a question defined differently from the so-called time-reversal, although it is related to this in a certain way which will become clear in our Sections 3 and 5.

There have been at least three circumstantial incentives which motivated undertaking this work. In the first place, it was emphasized by the author in a previous paper[1] that an essential difference between classical physics and quantum physics lies in the fact that in the latter the result of an observation can be used as the initial condition of the "state" immediately after the observation, but not as the final condition of the "state" immediately before the observation. Although this is in agreement with the customary usage of quantum physics, the conscious emphasis on this fact led the author himself to inquire whether one could not formulate quantum physics in such a way that the result of an observation can be used as the "retrodictive state" just before the observation.[2] This question will be answered in Section 5 of this paper. Although the answer is in the affirmative, the actual usefulness of such a retrodictive theory is extremely limited.

The second motive stemmed from an enlightening illustration that Dr. Keith Symon chose in a conversation to explain the reason why the H-theorem cannot be used for the past. He imagines that a man discovers on a desk two piles of playing cards, one in a perfect order and another in disorder. In spite of the fact that every permutation of cards has the same *a priori* probability, he would not guess that the well-ordered pile is a result of shuffling, but he would justifiably infer a selective human intervention in the past of the well-ordered pile. Keeping in mind that a "permutation of cards" corresponds to a quantum state, "well-ordered-ness" and "disordered-ness" to macroscopic cells, and "shuffling" to ergodic process, the reader will find that this pattern of inference is given a mathematical expression in our formulation of retrodiction in Section 4.

Thirdly, everyone familiar with the quantum theory of time-reversal[3] is rather disturbed by the fact that the operation of time-reversal is not a linear transformation, and also by the fact that the operation of double time-reversal does not become an identity transformation. One could expect that these esthetically unwelcome features of the theory may be avoided by a formulation which treats prediction and retrodiction on an equal footing. It will be shown in Section 5 that this expectation is justified.

The problem of retrodiction may be formulated in brief as follows: An observer B would like to guess from his own experimental data the result of another observer

[1] An article contributed by the author to the monograph, *Louis de Broglie, physicien et penseur* (Albin Michel, Paris, 1952), p. 385.

[2] It is the pleasure of the author to note with thanks that Dr. Adolf Grünbaum in a private communication encouraged undertaking clarification of this question. See also A. Grünbaum's article in the monograph, *Philosophy of Rudolf Carnap* (Tudor Publishing Company) (to be published).

[3] S. Watanabe, Phys. Rev. **84**, 1008 (1951). See also, S. Watanabe, Revs. Modern Phys. **27**, 26, 40 (1955).

A who observed the system some time before B and who has not confided his result to B. The main difficulty for retrodictor B arises from the fact that in his retrodictive inference he has to assume, apart from his own experimental finding, an *a priori* probability to each possible initial state (in which A might have found the system). There is in general no reason to assume an equal *a priori* probability for each quantum state except when the initial ensemble given to A can justifiably be assumed to be the result of an ergodic process. It will be explained in Section 3 by a simple example how easily a retrodictor can completely fail while a predictor cannot fail, naturally in the statistical sense of the word.

However, by assuming the uniform *a priori* initial probability, one can obtain an interesting formalism which exhibits on one hand a complete symmetry with respect to the two directions of "time," but which on the other manifests a definite one-way-ness of the direction of human "inference." In short, the present paper may be said to be an elaboration in the light of quantum physics of the following pregnant words due to W. Gibbs[4]:

"It should not be forgotten, when our ensembles are chosen to illustrate the probabilities of events in the real world, that while the probabilities of subsequent events may often be determined from the probabilities of prior events, it is rarely the case that probabilities of prior events can be determined from those of subsequent events, for we are rarely justified in excluding the consideration of the antecedent probability of the prior events."

2. MICROSCOPIC RETRODICTION

Let $\mathbf{S}$ be a complete set of eigenstates,

$$S_1, S_2, S_3, \cdots, S_i, \cdots, \tag{2.1}$$

of a family of mutually commuting observables defined with respect to a certain physical system. We shall use the same symbol $\mathbf{S}$ also to designate this family of observables. The completeness of $\mathbf{S}$ implies that the probability p_i of the system being found in state S_i satisfies

$$\sum_i p_i = 1. \tag{2.2}$$

A family $\mathbf{T}$ of mutually commuting observables which do not commute with $\mathbf{S}$ will define another complete set $\mathbf{T}$ of eigenfunctions.

The probability that the system which was in state S_i at the initial instant will be found in state T_j after τ seconds will be denoted by

$$P(S_i \rightarrow T_j, \tau) = P(i \rightarrow j), \tag{2.3}$$

$$S_i \epsilon \mathbf{S}, \quad T_j \epsilon \mathbf{T}, \tag{2.4}$$

where $\mathbf{S}$ and $\mathbf{T}$ may or may not be the same complete set. On account of the assumed completeness, we have

$$\sum_j P(i \rightarrow j) = 1. \tag{2.5}$$

From the invariance of dynamical laws for time-reversal (reversibility) or from the invariance for space-and-time-inversion (inversibility) we can conclude the inverse normalization[3]:

$$\sum_i P(i \rightarrow j) = 1. \tag{2.6}$$

This can also be derived from the unitarity of transition matrix.

Suppose that an observer A observes the system at $t=0$ with the observable-family $\mathbf{S}$, and that observer B observes the same system at $t=\tau$ with the observable-family $\mathbf{T}$. "Prediction" consists in the following position of problem on the part of A. Knowing that observer B will observe with $\mathbf{T}$ at $t=\tau$, observer A proposes to guess the result of B on the basis of his own result. If observer A had the result S_i, then his prediction will be that the probability of B obtaining T_j will be $P(i \rightarrow j)$. This means that, if observer A prepares a large number N of the cases where the result at $t=0$ was S_i, then $NP(i \rightarrow j)$ will be the number of cases where observer B will obtain T_j at $t=\tau$. Observer A will be called predictor and observer B monitor.

"Retrodiction" is now to be defined in a close analogy to the previous problem, only interchanging the roles of A and B. Knowing that observer A observed the system with $\mathbf{S}$ at $t=0$, but not knowing what his result was, observer B proposes to infer the result of A from his own result that the system is found in T_j at $t=\tau$. B will be called retrodictor and A monitor.

This question does not have a unique answer unless retrodictor B assumes a certain statistical behavior of monitor A regarding selection of the initial states. Independently of his own result at $t=\tau$, retrodictor B may have some general information about A, on the basis of which he may assume that monitor A has the general habit of selecting (and handing over to B) states S_i with weight $w_i (\sum_i w_i = 1)$. If A prepared a large number N of cases at $t=0$, then Nw_i among them must have been in S_i, according to the assumption. At the receiving end, $Nw_iP(i \rightarrow j)$ among these Nw_i will turn out to be in T_j. The total number of cases which will land in T_j will then be

$$\sum_i Nw_iP(i \rightarrow j), \tag{2.7}$$

among which $Nw_iP(i \rightarrow j)$ have originated from S_i. Then retrodictor B will say that the probability $Q(i \leftarrow j)$ that a system which was found at $t=\tau$ to be in T_j had been found in S_i at $t=0$ is

$$Q(i \leftarrow j) = \frac{Nw_iP(i \rightarrow j)}{\sum_i Nw_iP(i \rightarrow j)} = \frac{w_iP(i \rightarrow j)}{\sum_i w_iP(i \rightarrow j)}. \tag{2.8}$$

We use the "left-to-right" order to indicate the chronological direction, and an arrow to indicate the direction of inference.

[4] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, New Haven, Connecticut, 1914), p. 150.

It should be clearly understood that the foregoing result does not mean at all that if A prepared an ensemble with the weight given by Eq. (2.8) for each S_i, then retrodictor B would obtain the result T_j. Indeed, if A started with the weight distribution given in Eq. (2.8), then B would obtain T_k with weight

$$\frac{\sum_i w_i P(i\to j)P(i\to k)}{\sum_i w_i P(i\to j)}.$$

In other words, Eq. (2.8) represents the weight of S_i in the subset of systems ending in T_j when the entire ensemble has the weight distribution w_i. Insofar as the estimation of w_i is correct, observer B's retrodiction based on Eq. (2.8) must be statistically successful in this subensemble. If the estimation of w_i is unreliable, all retrodiction is meaningless.

It is true that retrodictor B could in principle find out the w's from the observed values of the quantity (2.7) for all the T's, if the number of the S's equals (or is less than) the number of the T's, and if the determinant of $P(i\to j)$ does not vanish. This, however, implies that before any attempt at retrodiction starting from a particular final state T_j can be made, a very large number of observations has to be made for all T's. Furthermore, it must be guaranteed that the values of the w's are kept fixed during all the observations. Thus, determination of the w's by retrodictor B is seldom feasible in practice.

If retrodictor B does not have any preliminary knowledge about the habit of A, the only thing he can do is to resort to the principle of ignorance and to assume that the *a priori* probability w_i is equal for each quantum state S_i. This attitude of B will be successful (verifiable by repetition) only if A prepares an ensemble with equal weight for all quantum states (similar to the microcanonical ensemble on an energy shell), and if B picks up only those cases which have landed in T_j and then classifies them according to various possible initial states. According to this simplifying assumption, Eq. (2.8) will become

$$Q_0(i\leftarrow j)=\frac{P(i\to j)}{\sum_i P(i\to j)} \tag{2.9}$$

and, further with the help of the inverse normalization (Eq. 2.6),

$$Q_0(i\leftarrow j)=P(i\to j). \tag{2.10}$$

It should be well noted that Eqs. (2.9) and (2.10) are based on a specific assumption that w_i is uniform. In fact, retrodictor B can very easily be "fooled" by monitor A. Suppose for instance that there are only two possible states (1) and (2) and $P(1\to 1)=P(1\to 2)=P(2\to 1)=P(2\to 2)=\frac{1}{2}$. No matter what ratio w_1/w_2 monitor A may choose, retrodictor B will find one half of the cases in state (1) and the other half in state (2). Conversely, observer B's retrodiction based on the equal distribution that one half of the cases must have originated from (1) and the other half from (2) may be completely wrong; monitor A may have handed over to retrodictor B only those systems which were found by A to be in (1) at $t=0$. The best way to avoid this deception on the part of B would be to impose on A, as a rule of the game, that he should pick up cases at random from the "microcanonical ensemble." Then Eq. (2.9) or Eq. (2.10) will have a meaning in a subensemble which lands in T_j. We shall hereinafter refer to the retrodiction based on the uniform w's as a "blind retrodiction."

Prediction, in contrast to retrodiction, has a simpler rule of game: the monitor (posterior observer here) is required to show all his results. Then, the prediction based on Eq. (2.3) will always be statistically successful. It should be emphasized that this asymmetry between prediction and retrodiction originates from the asymmetry of the "rules of game." In prediction, the predictor has the right to prepare the ensemble, while in retrodiction, the monitor has the right. We can easily change the rules to make prediction just as unreliable as retrodiction. Suppose monitor B referring to a prediction has the tendency to forget to record some of the cases in such a way that the chance of state T_j being recorded by him is proportional to w_j. Then, the prediction by A will be that the systems registered by him as S_i will be recorded by monitor B with the distribution given by

$$\frac{P(i\to j)w_j}{\sum_j P(i\to j)w_j}, \tag{2.11}$$

which offers a nice parallelism to Eq. (2.8). We shall however seldom have to deal with such a "forgetful" observer. The "rules of game" must be chosen in each case in such a way that they correspond faithfully to the nature of the actual description of physical phenomena under consideration. In this sense, $P(i\to j)$ given in Eq. (2.3) may be used for prediction, but for retrodiction we have to use $Q(i\leftarrow j)$ given in Eq. (2.8) with indeterminate w_i in general cases.

Exception has to be made to the entire consideration of this section either if (1) $\mathbf{S}$ and $\mathbf{T}$ are the same set and it commutes with the Hamiltonian of the system or if (2) $\mathbf{S}$ and $\mathbf{T}$ are the same and the time duration τ is zero. In this case, $P(i\to j)=\delta_{ij}$ and Eq. (2.10) follows automatically from Eq. (2.8) irrespective of the w's:

$$Q(i\leftarrow j)=P(i\to j)=\delta_{ij}. \tag{2.12}$$

Retrodiction is perfectly successful in this special case.

The situation in classical physics may be included in Eq. (2.10), if the "state" is determined as precisely as possible in principle, i.e., if the system is located at a point in the phase space.

3. MACROSCOPIC RETRODICTION

We shall now introduce the concept of macroscopic cells in our consideration. The macroscopic observations are known to be compatible with one another, therefore we can think of a family **S** of microscopic observables which are compatible with all these macroscopic observations. In general, this **S** will not be commutable with the exact Hamiltonian of the system.[5] Suppose further that the eigenstates, S_1, S_2, etc., are grouped into macroscopic cells which are labeled by $\mu=1, 2, \cdots$, in such a way that cell $\mu=1$ contains n_1 eigenstates of **S**, cell $\mu=2$ contains n_2 eigenstates of **S**, etc.

A macroscopic prediction consists in inferring the probability of finding the system in cell ν at $t=\tau$ when it is known that the system was found in cell μ at $t=0$. The answer will be, in terms of the microscopic transition probabilities,*

$$P(\mu\rightarrow\nu)=\frac{1}{n_\mu}\sum_i^{(\mu)}\sum_j^{(\nu)}P(i\rightarrow j), \qquad (3.1)$$

where $\sum_i^{(\mu)}$ means that i should run over all the eigenstates contained in cell μ. It should be noted that this answer is based on the equal weight of S_i within cell μ, and complete disorder of phase among these states S_i. In other words, we are taking as the initial state a density matrix (statistical ensemble) which corresponds to the Hilbert subspace μ. Writing $\mathcal{P}[\varphi]$ for the projection operator for quantum state φ, we can express our initial ensemble by

$$\mathcal{G}=\frac{1}{n_\mu}\sum_i^{(\mu)}\mathcal{P}[\varphi_i]. \qquad (3.2)$$

This is the best we can do under the given information that the system was found in μ at $t=0$. The quantity given in Eq. (3.1) is of course normalized with regard to ν, i.e., $\sum_\nu P(\mu\rightarrow\nu)=1$.

Now the retrodiction consists in inferring the probability that the system had been found in macroscopic state μ at $t=0$, when it is known that the system was found to be in macroscopic state ν at $t=\tau$. Again introducing the *a priori* probability w_μ for each quantum state in cell μ, one will answer that the probability in question is given by

$$Q(\mu\leftarrow\nu)=\frac{w_\mu n_\mu P(\mu\rightarrow\nu)}{\sum_\mu w_\mu n_\mu P(\mu\rightarrow\nu)}$$

$$=\frac{w_\mu\sum_i^{(\mu)}\sum_j^{(\nu)}P(i\rightarrow j)}{\sum_\mu w_\mu\sum_i^{(\mu)}\sum_j^{(\nu)}P(i\rightarrow j)}. \qquad (3.3)$$

If we can assume that the background ensemble of A was a microcanonical ensemble, i.e., if all the w's are equal, then we can simplify Eq. (3.3) to the form:

$$Q_0(\mu\leftarrow\nu)=(n_\mu/n_\nu)P(\mu\rightarrow\nu), \qquad (3.4)$$

where the inverse normalization given in Eq. (2.4) has been utilized. The probability given in Eq. (3.3) or (3.4) satisfies the normalization condition with regard to μ: $\sum_\mu Q(\mu\leftarrow\nu)=0$.

It might appear as if prediction in the macroscopic case were equally unreliable as retrodiction since we have to use the assumption of equal probability (within cell μ) also for prediction here. However, it should not be forgotten that it is observer-predictor A himself who prepares the initial ensemble, therefore unless he puts uneven, selective weights to various states within cell μ he can succeed. On the other hand, the supposed even weight all over the energy shell assumed in Eq. (3.4) is not in the control of observer-retrodictor B, who therefore can very easily fail in his retrodiction.

Let us next examine the consequences of reversibility (invariance for time-reversal) and inversibility (invariance for space-and-time inversion) on our problem. The reversed state S' of a state S means the one in which all the particles have the same positions as in S but the equal and opposite velocities to those in S.[6] The inverted state S' of a state S means the one in which all the particles have the same velocities as in S and the space-inverted positions as compared with S.[6] Reversibility and inversibility, which hold in the basic processes in quantum mechanics, then mean:

$$P(S\rightarrow T)=P(T'\rightarrow S'). \qquad (3.5)$$

The macroscopic observations usually[7] cannot distinguish a state from its reversed or inverted state. In other words, a cell μ contains the reversed, as well as inverted, state S' of a state S if it contains S. Then from Eq. (3.1), we obtain in virtue of Eq. (3.5),

$$P(\mu\rightarrow\nu)=\frac{1}{n_\mu}\sum_{i'}^{(\mu)}\sum_{j'}^{(\nu)}P(j'\rightarrow i')=\frac{n_\nu}{n_\mu}P(\nu\rightarrow\mu). \qquad (3.6)$$

Using this relation, we can write Eq. (3.3) in the form:

$$Q(\mu\leftarrow\nu)=\frac{w_\mu P(\nu\rightarrow\mu)}{\sum_\mu w_\mu P(\nu\rightarrow\mu)}, \qquad (3.7)$$

and with the assumption of uniform weight,

$$Q_0(\mu\leftarrow\nu)=P(\nu\rightarrow\mu) \qquad (3.8)$$

which has a striking simplicity. Equations (3.4) and (3.8) are applicable only to the "blind" macroscopic retrodiction.

What has been developed in this section also applies to classical, statistical considerations if we replace the number of quantum states by the volume in the phase space.

[5] The exact Hamiltonian may commute with the "macroscopic energy" but not with the other macroscopic quantities. J. v. Neumann, Z. Physik **57**, 30 (1929).

* *Note added in proof.*—The P on the left hand side of Eq. (3.1) is what was denoted by W in Part I. (See reference 6.)

[6] For the precise definition of time-reversal and space-and-time inversion, see Sections 3, 4, Part I, and Section 3, Part II, S. Watanabe, Revs. Modern Phys. **27**, 26–76 (1955).

[7] This is certainly the case if we limit the macroscopic quantities to a certain category, for instance to the thermodynamical variables.

4. APPLICATION OF THE *H*-THEOREM TO THE PAST

The probability $P(\nu\rightarrow\mu)$ depends naturally on the length of the interval τ between the two observations. The ergodic H-theorem, in essence, states that if the exact Hamiltonian does not commute with the observable-family **S** which is compatible with the macroscopic observations,[8] then the probability $P(\nu\rightarrow\mu)$, averaged over possible values of τ,[9] is proportional to the size of the final cell μ:

$$\langle P(\nu\rightarrow\mu)\rangle = n_\mu/N, \tag{4.1}$$

where N is the total number of quantum states on the energy shell.

Equation (4.1) shows that if we take a value of τ arbitrarily from its possible domain, $0<\tau<\infty$, the probability $P(\nu\rightarrow\mu)$ of finding the system in a large cell μ is large. Invoking now the relation (3.8), we can say that on the assumption of blind retrodiction, the probability $Q_0(\mu\leftarrow\nu)$ that the system had been found at $t=-\tau$ in a large cell μ is also large if $-\tau$ is arbitrarily taken from its possible domain $-\infty<-\tau<0$. If we use the Boltzmann entropy $\mathcal{S}_B$,

$$\mathcal{S}_B = \log n_\mu. \tag{4.2}$$

We can say on the basis of blind retrodiction that if $\mathcal{S}_B$ at $t=0$ has a certain nonmaximum value, then it is just as probable to have a larger entropy value in the future as in the past. This is the well-known conclusion of a formal application of the H-theorem to the past. We could also use the Gibbs entropy $\mathcal{S}_G$,

$$\mathcal{S}_G(\tau) = -\sum_\mu P(\nu\rightarrow\mu)\log[P(\nu\rightarrow\mu)/n_\mu], \tag{4.3}$$

$$\mathcal{S}_G(-\tau) = -\sum_\mu Q(\mu\leftarrow\nu)\log[Q(\mu\leftarrow\nu)/n_\mu], \tag{4.4}$$

with $\tau>0$, but it may be easier to visualize the situation with the help of the Boltzmannian entropy.

The foregoing argument is based on the premise of blind retrodiction which may be the only possible basis of inference if it is perfectly certain that the system had been isolated from the exterior system (except a possible prior observer who does not perform any kind of selection) and if we have absolutely no other information about the system than that it was found in ν at $t=0$. However, such conditions are seldom satisfied in the actual circumstances. A sounder inference than the mere blind retrodiction, in line with Symon's idea explained in our Section 1, would be somewhat as follows:

Consider two cells μ and ν such that $n_\mu \gg n_\nu$. Then according to Eq. (3.6), we have

$$\frac{P(\mu\rightarrow\nu)}{P(\nu\rightarrow\mu)} = \frac{n_\nu}{n_\mu} \ll 1, \tag{4.5}$$

and according to Eq. (4.1) we have also

$$\frac{\langle P(\kappa\rightarrow\nu)\rangle}{\langle P(\lambda\rightarrow\mu)\rangle} = \frac{n_\nu}{n_\mu} \ll 1 \quad (\kappa,\ \lambda \text{ arbitrary}). \tag{4.6}$$

Now let us assume that we find a system at $t=0$ in cell ν. Seeing from Eq. (4.5) that it is extremely improbable for a system starting from a large cell μ to reach a small cell ν, we suspect that such was not the actual history behind the system we have just found in ν. This inference is a direct contradiction to the aforementioned result based on the uniform w. Thus we are led to modify the assumption of uniform w in such a manner as to give less weight w_μ to larger cells μ and larger weight to smaller cells. Such an assumption of nonuniform w is perfectly allowable according to our theory. In fact, if there is any possible doubt about the isolation of the system in the past, for instance, there is no reason to adopt the hypothesis of blind retrodiction. Then the result of observation that the system was found in a small cell at present can very well reflect itself in our estimation of the w's.

Once we have abandoned the assumption of uniform w, we cannot use Eq. (3.8) any longer and have to go back to Eq. (3.7). In spite of the fact that $P(\nu\rightarrow\mu)$ may be large, $Q(\mu\leftarrow\nu)$ can be small if w_μ is small in Eq. (3.7). And the probability of the system having originated from a small cell can become quite large. Thus the entropy value $\mathcal{S}_B$ at $t=-\tau$ may probably have been smaller. Our formalism is flexible enough to incorporate this very reasonable inference.

The foregoing argument can be applied also to the classical, statistical mechanics. It is interesting to note how our argument can stand the famous objection due to Loschmidt. It is true that in cell ν there are just as many microstates headed for larger values of entropy $\mathcal{S}_B$ in the future as those which have originated from larger values of entropy in the past. If w is uniform, then each microstate inside cell ν will be occupied by the same weight on account of the permanence of the microcanonical ensemble.[10] Then, Loschmidt's argument becomes valid, and we have to conclude larger values of entropy for the future as well as for the past. But if w is not necessarily uniform, then we need not assume equal weight for each microstate inside the cell for the purpose of extrapolation towards the past. Then Loschmidt's objection does not hold any longer. For a realistic macroscopic retrodiction, we should not use the uniform weight within the macroscopic cell ν, while it may be assumed for prediction.

It is interesting to note that the blind application of the ergodic H-theorem to the past does not actually

[8] Although the noncommutability of the exact Hamiltonian with **S** is the main hypothesis, we need some more auxiliary conditions to derive this result. For the two versions of these conditions, see J. von Neumann, Z. Physik **57**, 30 (1929); and W. Pauli and M. Fierz, Z. Physik **106**, 572 (1937).

[9] The time-average in v. Neumann's proof can be only for the positive values of τ. It should be noted also that the H-theorem considered here refers to one initial observation ($t=0$) and one final observation ($t=\tau$) and is different from the consideration based on repeated observations. See Section 7, Part I, S. Watanabe, Revs. Modern Phys. **27**, 26 (1955).

[10] This is a consequence of the inverse normalization, Eq. (2.6).

yield any newer information than what one has put in as the assumption. The combination of Eqs. (4.1) and (3.8) gives

$$\langle Q_0(\mu\leftarrow\nu)\rangle = n_\mu/N, \tag{4.7}$$

which is nothing but an expression of the uniform probability, an assumption which has been used in deriving Eq. (3.8).

5. RETRODICTIVE QUANTUM MECHANICS

Our basic equation (2.8) for retrodiction can be written as

$$Q(i\leftarrow j) = w_i Q_0(i\leftarrow j)/\sum_i w_i Q_0(i\leftarrow j), \tag{5.1}$$

with the help of Eq. (2.10):

$$Q_0(i\leftarrow j) = P(i\rightarrow j), \tag{5.2}$$

where the w_i depends on the over-all judgment of the retrodictor. Only when the system has been isolated in the past and there is no other clue to its past history than the observational fact that the system is found at present in state j, will the retrodictor use the uniform value of w_i for various i's and $Q(i\leftarrow j)$ will reduce to $Q_0(i\leftarrow j)$. What follows mainly concerns the blind retrodiction represented by Q_0, but Q can be derived from Q_0 by the use of Eq. (5.1) if there is any way of estimating w_i.

In this section, we shall first show that the quantity given in Eq. (5.2) can be calculated in two ways: either solving the Schrödinger equation with the initial condition S_i, or solving the same equation with the final condition T_j. Although the resulting values of probability are the same, the first method agrees better with the idea suggested by the right hand side of Eq. (5.2), while the second method reflects more faithfully the idea suggested by the left hand side. Since the first method is the customary one, we shall only show how the second method can be used to evaluate the same probability.

Let the eigenfunctions of **S** be called $\varphi_1, \varphi_2, \cdots, \varphi_i, \cdots$ and those of **T** $\psi_1, \psi_2, \cdots, \psi_j, \cdots$. Further, let the solution of the Schrödinger equation,

$$\partial\Psi(t)/\partial t = -iH(t)\Psi(t), \tag{5.3}$$

satisfying the final condition,

$$\Psi(\tau) = \psi_j, \tag{5.4}$$

be denoted by $\Psi_r(t)$. Expanding $\Psi_r(0)$ according to the φ_i:

$$\Psi(0) = \sum_i a_i \varphi_i \tag{5.5}$$

we can easily show that $a_i^* a_i$ represents the probability (5.2).

Consider the transition matrix $U(t_1,t_2)$ defined by:

$$\begin{aligned} \partial U(t_1,t_2)/\partial t_1 &= -iH(t_1)U(t_1,t_2), \\ \partial U(t_1,t_2)/\partial t_2 &= +iU(t_1,t_2)H(t_2), \end{aligned} \tag{5.6}$$

$$U(t_1,t_1) = 1, \quad U(t_2,t_1) = U^{-1}(t_1,t_2) = \bar{U}(t_1,t_2). \tag{5.7}$$

Then according to the customary theory, the probability (5.2) is given by

$$P(i\rightarrow j) = |(\psi_j, U(\tau,0)\varphi_i)|^2. \tag{5.8}$$

On the other hand, $\Psi_r(t)$ considered in the foregoing is

$$\Psi_r(t) = U(t,\tau)\psi_j, \tag{5.9}$$

and the coefficients a_i are

$$a_i = (\varphi_i, U(0,\tau)\psi_j). \tag{5.10}$$

On account of the unitarity of U and of the relation $U(0,\tau) = U^{-1}(\tau,0)$, as in Eq. (5.7), we obtain

$$\begin{aligned} a_i &= (\bar{U}(0,\tau)\varphi_i, \psi_j) \\ &= (U(\tau,0)\varphi_i, \psi_j) = (\psi_j, U(\tau,0)\varphi_i)^*. \end{aligned} \tag{5.11}$$

Hence, in view of Eq. (5.8),

$$a_i^* a_i = P(i\rightarrow j) = Q_0(i\leftarrow j). \tag{5.12}$$

This situation suggests a new picture of the "state" of a system between two observations, one at $t=0$ and the other at $t=\tau$: There exist simultaneously two states, one being a predictive state $\Psi_p(t)$ which complies with the initial condition at $t=0$, and the other a retrodictive state $\Psi_r(t)$ which complies with the final condition at $t=\tau$. Both $\Psi_p(t)$ and $\Psi_r(t)$ obey the same Schrödinger equation. This picture, though redundant in practical applications, offers certain intellectual interest, for it provides a complete symmetry between two consecutive observations. Ψ_r can be given just as much, or just as little, "reality" as Ψ_p.

Now, we should like to look upon the same situation from a slightly different point of view, namely we attempt to establish a time-symmetry, not with regard to two observations, but with regard to the future and past referring to a single observation at hand. Suppose we make an observation at $t=0$ and obtain a result φ_i. Then our inference will develop towards the past just as well as towards the future. Let us introduce a new variable s, called inference parameter, which coincides with t when it refers to prediction, and which is equal to minus t when it refers to retrodiction. s is then always positive.

The development of a retrodictive state $\Psi_r(t)$ starting backward from φ_i at $t=0$ is nothing but the extrapolation of the predictive state and obeys

$$\partial\Psi_r(t)/\partial t = -iH(t)\Psi_r(t), \quad \Psi_r(0) = \varphi_i, \tag{5.13}$$

or in terms of s,

$$\partial\Psi_r(-s)/\partial s = +iH(-s)\Psi_r(-s). \tag{5.14}$$

The Hamiltonian being Hermitian, the complex conjugate of Eq. (5.14) becomes

$$\partial\Psi_r^*(-s)/\partial s = -i\Psi_r^*(-s)H(-s). \tag{5.15}$$

Introducing a time-independent unitary operator R, called reversion operator,[6] such that

$$(R^{-1}H(-s)R)^T = H(s), \tag{5.16}$$

we can rewrite Eq. (5.15) in the form:

$$\partial R^T\Psi_r^*(-s)/\partial s=-iH(s)R^T\Psi_r^*(-s). \quad (5.17)$$

This means that $\Phi(s)$ defined by

$$\Phi(s)=R^T\Psi_r^*(-s) \quad \text{or} \quad \Psi_r(-s)=R\Phi^*(s) \quad (5.18)$$

satisfies the same equation as the predictive state,

$$\partial\Phi(s)/\partial s=-iH(s)\Phi(s). \quad (5.19)$$

The only difference is that $\Phi(s)$ satisfies the initial condition:

$$\Phi(0)=R^T\varphi_i^*, \quad (5.20)$$

while the predictive state $\Psi_p(s)$ satisfies

$$\Psi_p(0)=\varphi_i. \quad (5.21)$$

The probability of finding this system at $t=s>0$ in state ψ_j will be

$$P(i\rightarrow j)=|(\psi_j,\Psi_p(s))|^2, \quad (5.22)$$

while the blind probability that the system had been found at $t=-s<0$ in state ψ_j will be

$$\begin{aligned} Q_0(j\leftarrow i)&=|(\psi_j,\Psi_r(-s))|^2\\ &=|(\psi_j,R\Phi^*(s))|^2=|(R^T\psi_j^*,\Phi(s))|^2. \end{aligned} \quad (5.23)$$

In brief, the two inferential states $\Psi_p(s)$ and $\Phi(s)$ can be treated in a parallel fashion, only using $R^T\phi^*$ for $\Phi(s)$ wherever we would use ϕ for $\Psi_p(s)$. Compare Eqs. (5.20) and (5.23), respectively, with Eqs. (5.21) and (5.22). It would then be a tempting idea to introduce a quantity which comprises both Ψ_p and Φ on the same footing. A "double inferential state" composed of two components is defined by

$$\mathbf{\Psi}(t)=\begin{pmatrix}\Psi_p(s)\\ \Phi(s)\end{pmatrix}=\begin{pmatrix}\Psi_p(t)\\ \Phi(-t)\end{pmatrix}. \quad (5.24)$$

where s is now freed from the condition $s>0$. This function will obey

$$\partial\mathbf{\Psi}(t)/\partial t=-i\mathbf{H}(t)\mathbf{\Psi}(t), \quad (5.25)$$

with

$$\mathbf{H}(t)=\begin{pmatrix}H(t) & 0\\ 0 & -H(-t)\end{pmatrix} =\begin{pmatrix}H(t) & 0\\ 0 & -(R^{-1}H(t)R)^T\end{pmatrix}. \quad (5.26)$$

The initial condition of $\mathbf{\Psi}(t)$ is

$$\mathbf{\Psi}(0)=\begin{pmatrix}\varphi_i\\ R^T\varphi_i^*\end{pmatrix}, \quad (5.27)$$

and the solution of Eq. (5.25) at an arbitrary value of t will then have the form:

$$\mathbf{\Psi}(t)=\begin{pmatrix}\Psi(t)\\ R^T\Psi^*(t)\end{pmatrix}. \quad (5.28)$$

We can however liberalize the relationship between the two components of Eq. (5.28) without affecting their physical meaning. Namely, taking any unitary operator W[11] which commutes with all the known physical quantities, we can write, instead of Eq. (5.28),

$$\mathbf{\Psi}(t)=\begin{pmatrix}\Psi(t)\\ R^TW^*\Psi^*(t)\end{pmatrix}. \quad (5.29)$$

This amounts to replacing $\Psi(t)$ by $W\Psi(t)$, which of course does not change the meaning of a state function. The initial condition of Eq. (5.29) is then

$$\mathbf{\Psi}(0)=\begin{pmatrix}\varphi_i\\ R^TW^*\varphi_i^*\end{pmatrix}. \quad (5.30)$$

To make our discussion more concrete let us take as W

$$W=\Delta^n, \quad (5.31)$$

where n in any arbitrary integer and Δ is given by[11]

$$\Delta=\Delta^{-1}=\Delta^T=\Delta^*=R^TR^{-1}=\prod_i^{\text{all}}(-1)^{N_i}, \quad (5.32)$$

in which N_i is the occupation number operator for the spinor eigenstate labeled i. Δ is known to commute with the reversion operator, R. Then the general pattern of a double inferential function is

$$\mathbf{\Psi}(t)=\begin{pmatrix}\Psi(t)\\ \Delta^nR^T\Psi^*(t)\end{pmatrix}, \quad (5.33)$$

with arbitrary n. Δ^n is unity when n is even.

We can now introduce the "reversed" inferential function $\mathbf{\Psi}$ of Eq. (5.33) by

$$\mathbf{\Psi}_R(t)=\begin{pmatrix}\Delta^nR^T\Psi^*(t)\\ \Psi(t)\end{pmatrix}, \quad (5.34)$$

which certainly falls in the supposed general pattern of an inferential function (5.33), only the arbitrary number n being replaced by $n+1$, for

$$\Psi(t)=\Delta^{n+1}R^T(\Delta^nR^T\Psi^*(t))^*. \quad (5.35)$$

Furthermore, at each value of t, the first and the second components of Eq. (5.34) represent respectively the so-called "reversed states" of the first and the second components of Eq. (5.33).[6] Indeed, for a given state $\phi(t)$, its reversed state can be expressed by $WR^T\phi^*(t)$. The transformation from Eq. (5.33) to Eq. (5.34) can be written

$$\mathbf{\Psi}_R(t)=\mathbf{R}\mathbf{\Psi}(t), \quad \mathbf{\Psi}(t)=\mathbf{R}\mathbf{\Psi}_R(t), \quad (5.36)$$

[11] See Section 12, Part II, S. Watanabe, Revs. Modern Phys. 27, 40 (1955).

with

$$\mathbf{R}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}. \tag{5.37}$$

The formalism presented here has no practical advantage over the current quantum theory, but it has a formal advantage in that time-reversal is represented here by a linear transformation (5.37), and double time-reversal[6] becomes an identity transformation:

$$\mathbf{R}^2=1. \tag{5.38}$$

6. IRREVERSIBILITY OF INFERENCE AND INFORMATION

Suppose observer A prepares a large number N of systems which were found at $t=0$ in state φ_i. After τ seconds, each system will become $U(\tau,0)\varphi_i$. If observer B performs at $t=\tau$ an observation with the complete set $\mathbf{T}(\psi_1,\psi_2,\cdots)$, then $N|(\psi_j,U(\tau,0)\varphi_i)|^2$ systems will turn out to be in state ψ_j. With the help of projection operators $\mathcal{P}[\phi]$ we can write this process in the following schema:

$$\begin{aligned}\mathcal{G}_1=\mathcal{P}[\varphi_i]\rightarrow\mathcal{G}_2=\mathcal{P}[U(\tau,0)\varphi_i]\rightarrow\\ \mathcal{G}_3=\textstyle\sum_j|(\psi_j,U(\tau,0)\varphi_i)|^2\mathcal{P}[\psi_j].\end{aligned} \tag{6.1}$$

The amount of "information"[12] carried by the knowledge about the system represented by $\mathcal{G}$ is

$$I=\mathrm{Spur}(\mathcal{G}\log\mathcal{G})+\mathrm{const.} \tag{6.2}$$

This quantity does not change in the first step of transition in Eq. (6.1), but does decrease in the second step. This is the famous irreversibility of observation pointed out by von Neumann.[12] It should be noted that $\mathcal{G}_3$ in Eq. (6.1) does not represent the knowledge obtained by observer B in individual cases, for in each case observer B knows perfectly well in which one of the ψ's the system is found. $\mathcal{G}_3$ can be considered as a global description of the entire ensemble after the observation, or as the prediction of the result of B in each case.

Next, suppose that observer A prepares at $t=0$ a large number of systems with equal weight in all possible φ's ($\epsilon\mathbf{S}$). Observer B performs at $t=\tau$ an observation with $\mathbf{T}$, and a certain large number N of systems is found to be in state ψ_j. He considers now only those systems ending in ψ_j, and ask what percent of them had been registered as φ_i by the previous observer A. Then, he extrapolates ψ_j backward by the Schrödinger equation from $t=\tau$ to $t=0$, and calculates $|(\varphi_i,U(0,\tau)\psi_j)|^2$. His inference will then be that, among N systems that he found in ψ_j, $N|(\varphi_i,U(0,\tau)\psi_j)^2|$ systems must have been found by A to be in φ_i. Schematically, this inference can be denoted by

$$\begin{aligned}\mathcal{G}_3'=\textstyle\sum_i|(\varphi_i,U(0,\tau)\psi_j)|^2\mathcal{P}[\varphi_i]\leftarrow\\ \mathcal{G}_2'=\mathcal{P}[U(0,\tau)\psi_j]\leftarrow\mathcal{G}_1'=\mathcal{P}[\psi_j],\end{aligned} \tag{6.3}$$

which exhibits a parallelism to (6.1). $\mathcal{G}_3'$ in (6.3) represents a partial ensemble immersed in the uniform ensemble:

$$\mathcal{G}_0=\lambda\textstyle\sum_i\mathcal{P}[\varphi_i] \tag{6.4}$$

prepared by A. If A would have started with $\mathcal{G}_3'$, then B would not obtain $\mathcal{G}_1'$. Nonetheless, $\mathcal{G}_3'$ represents the legitimate inference made by B based on the blind retrodiction hypothesis with regard to the results that A had obtained in the systems which were later found by B in ψ_j.

If B has any further source of judgment about the initial ensemble, he will modify the assumption of uniform weight, and attach a reappraised *a priori* probability w_i to each φ_i. In this case, $\mathcal{G}_3'$ will become

$$\begin{aligned}\mathcal{G}_3'=\{\textstyle\sum_i w_i|(\varphi_i,U(0,\tau)\psi_j)|^2\mathcal{P}[\varphi_i]\}/\\ \{\textstyle\sum_i w_i|(\varphi_i,U(0,\tau)\psi_j)|^2\}\end{aligned} \tag{6.5}$$

in accordance with Eqs. (2.8) or (5.1).

It is evident that, no matter whether one uses $\mathcal{G}_3'$ of Eqs. (6.3) or that of Eq. (6.5), the amount of information carried by $\mathcal{G}_3'$ is smaller than that carried by $\mathcal{G}_1'$, i.e., the decrease of information here takes place in the backward direction of time. Both the case of prediction, Eq. (6.1), and the case of retrodiction, Eqs. (6.3) or (6.5), can, however, be included in the statement that the amount of information decreases in the direction of inference, i.e., in the positive direction of the inference parameter of the last section. This last result is in a good agreement with the common sense, for an inference cannot contain more information than the fact from which the inference is drawn.

In the statements in the foregoing, the phrase "information decreases" must be replaced by "information remains constant" in the following two cases: When (1) $\mathbf{S}$ and $\mathbf{T}$ are the same set and the elapse of time τ is zero, or (2) $\mathbf{S}$ and $\mathbf{T}$ are the same set and commute with the Hamiltonian of the system.

[12] C. Shannon and W. Weaver, *Mathematical Theory of Communication* (University of Illinois Press, Urbana, Illinois, 1949). We do not indulge here in the discussion regarding the sign before the Spur and regarding the constant in Eq. (6.2). The quantity (6.2) was first used by von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Verlag Julius Springer, Berlin, Germany, 1932). See also L. Szilard, Z. Physik **53**, 840 (1929). For an early application of the quantity (6.2) to a concrete physical problem, see S. Watanabe, Z. Physik **113**, 482 (1939).