5 October 1998

PHYSICS LETTERS A

Physics Letters A 247 (1998) 9-13

Feasible histories, maximum entropy and a MinMax choice criterion

Itamar Pitowsky [1]

Department of Philosophy, The Hebrew University, Jerusalem, Israel

Received 17 June 1998; accepted for publication 21 July 1998
Communicated by P.R. Holland

Abstract

We consider the broadest possible consistency condition for a family of histories, which extends all previous proposals. A family that satisfies this condition is called *feasible*. On each feasible family of histories we choose a probability measure by maximizing entropy, while keeping the probabilities of commuting histories to their quantum mechanical values. This procedure is justified by the hypothesis that decoherence increases entropy. Finally, a criterion for identifying the nearly classical histories is proposed. © 1998 Published by Elsevier Science B.V.

1. Introduction

The central idea of the consistent histories approach to quantum mechanics is that it is sometimes possible to assign probabilities to sequences of events – or *histories* – and not just to a single quantum event. Moreover, this can be done in a meaningful way even for closed systems on which no external measurement was performed. The enterprise was developed by Griffiths [1], Omnès [2] and Gell-Mann and Hartle [3].

As noted by Goldstein and Page [4], the definition of "consistent family of histories", and the association of probability with each element in such a family, is not uniquely determined by quantum theory. The theory dictates the probability of a history only when all the projection operators comprising it commute in pairs. The probability of non-commuting histories can then be defined in a variety of ways, provided that it obeys the usual axioms of probability. This last requirement is just the consistency condition.

What I suggest is to take, as a starting point, the broadest possible consistency condition (which, obviously, includes all previous ones). The condition is called *feasibility condition*, and a family of histories that satisfy it is called *feasible*. This name is not accidental. It carries the technical meaning associated with it in the theory of optimization.

Given a feasible family, there is usually more than one probability measure on it that agrees with the prescriptions of quantum mechanics. To choose the right one, we impose a *maximum entropy* condition in the spirit of Janes [5]. This is justified since decoherence increases entropy. With this principle we obtain a probability distribution and a partition function for each feasible family. We prove that they can be represented in terms of a *decoherence functional* [6].

This still leaves us with the major problem of the consistent histories enterprise: The class of consistent (or feasible) families is too large. Here I agree with

[1] E-mail address: itamarp@vms.huji.ac.il.

PII S0375-9601(98)00564-7

the penetrating criticism of Dowker and Kent [7]. To amend the situation, and, in particular, to explain why our experience is "nearly classical", I suggest a possible choice principle in the last section.

2. Feasible families

For the sake of simplicity we assume that all the lowest dimensional projections in a family of histories are one-dimensional. (The more general case requires easy modifications.) Also, for simplicity, we shall deal with Hilbert spaces of finite dimension. I shall work in the Heisenberg picture and suppress explicit reference to time.

Thus, let H be a complex Hilbert space of finite dimension d. *A family of histories* F, is an ordered set of n orthonormal bases in H,

$$\begin{aligned}&|1,1\rangle,\ |1,2\rangle,\ \ldots,\ |1,d\rangle,\\&|2,1\rangle,\ |2,2\rangle,\ \ldots,\ |2,d\rangle,\quad \ldots\\&|n,1\rangle,\ |n,2\rangle,\ \ldots,\ |n,d\rangle.\end{aligned} \tag{1}$$

With each such family we associate *the index set* of the family, which is just $I = \{1,2,\ldots,d\}^n = \{1,2,\ldots,d\} \times \ldots \times \{1,2,\ldots,d\}$. A subset of the index set is a *cylinder* if it has the form $\alpha = \alpha_1 \times \alpha_2 \times \ldots \times \alpha_n$, where for each i, $\alpha_i \subseteq \{1,2,\ldots,d\}$. With each cylinder α we associate a *history*. It is just the sequence of projections $(P_{\alpha_1}, P_{\alpha_2}, \ldots, P_{\alpha_n})$ where $P_{\alpha_i} = \sum_{j\in\alpha_i} |i,j\rangle\langle i,j|$, $i = 1,2,\ldots,n$. The history is *commutative* if the P_{α_i}'s pairwise commute. Here we shall also say that the cylinder α is *commutative*.

Definition 1. Let W be a state defined on H. A family of histories is feasible (with respect to W) if there is some probability measure μ on the subsets of the index set such that

$$\begin{aligned}\mu(\alpha) &= \mu(\alpha_1 \times \ldots \times \alpha_n)\\ &= \mathrm{tr}(W P_{\alpha_1} P_{\alpha_2} \ldots P_{\alpha_n}),\end{aligned} \tag{2}$$

for every commutative history $(P_{\alpha_1}, P_{\alpha_2}, \ldots, P_{\alpha_n})$.

Thus, in a feasible family one can define the probability of every history in a way that extends the quantum mechanical prescription (which applies to commutative histories and is given in formula (2)). The precise conditions of feasibility are known, though maybe hard to compute in practice. These are exactly the conditions for the existence of a "local hidden variable" model for quantum mechanical experiments; or more generally, for the possibility of interpolating statistical data by a probability measure [8] [2].

Example 1. Let H be a two-dimensional space and let $|+\rangle$, $|-\rangle$ be the vectors corresponding with "spin up" and "spin down" in the z-direction respectively. Let F be the family

$$\begin{aligned}&|1,1\rangle = |+\rangle,\quad |1,2\rangle = |-\rangle,\\&|2,1\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle),\quad |2,2\rangle = \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle),\\&|3,1\rangle = |+\rangle,\quad |3,2\rangle = |-\rangle,\end{aligned} \tag{3}$$

and let the initial state $W = |w\rangle\langle w|$ be pure, where $|w\rangle$ is given by $|w\rangle = (1/\sqrt{2})(\mathrm{e}^{\mathrm{i}\phi/2}|+\rangle + \mathrm{e}^{-\mathrm{i}\phi/2}|-\rangle)$ and ϕ is to be determined later. Denote $P_{ij} = |i,j\rangle\langle i,j|$, $i = 1,2,3$, $j = 1,2$, and by I the unit operator on H. The commutative histories – each followed by its probability – are given by

$$\begin{aligned}&(P_{11}, I, P_{31}), \tfrac{1}{2};\quad (P_{11}, I, P_{32}), 0;\quad (P_{12}, I, P_{31}), 0;\\&(P_{12}, I, P_{32}), \tfrac{1}{2};\quad (I, P_{21}, I), \cos^2 \tfrac{1}{2}\phi;\\&(I, P_{22}, I), \sin^2 \tfrac{1}{2}\phi;\quad (P_{11}, I, I), \tfrac{1}{2};\quad (P_{12}, I, I), \tfrac{1}{2};\\&(I, I, P_{31}), \tfrac{1}{2};\quad (I, I, P_{32}), \tfrac{1}{2};\quad (I, I, I), 1.\end{aligned}$$

This family of histories is feasible. To see that simply choose for the non-commuting, maximally refined histories (P_{1i}, P_{2j}, P_{3k}), $i,j,k = 1,2$ the probability

$$\begin{aligned}&\mu(\{i\} \times \{j\} \times \{k\})\\&\quad = \tfrac{1}{2}\delta_{ik}(\delta_{1j}\cos^2 \tfrac{1}{2}\phi + \delta_{2j}\sin^2 \tfrac{1}{2}\phi)\end{aligned} \tag{4}$$

and then extend μ by additivity to the rest of the algebra. In particular, one obtains the correct probabilities for the commutative histories. However, this family is not consistent according to the criterion of Refs. [1–3]. Indeed, we see that on the one hand

$$\begin{aligned}&\mathrm{tr}(P_{11}P_{21}P_{31}WP_{31}P_{21}P_{11})\\&\quad + \mathrm{tr}(P_{11}P_{21}P_{32}WP_{32}P_{21}P_{11}) = \tfrac{1}{4},\end{aligned}$$

[2] These conditions were already investigated by George Boole who called them "conditions of possible experience". See Ref. [9].

while, on the other hand, $\mathrm{tr}(P_{11}P_{21}WP_{21}P_{11}) = \frac{1}{2}\cos^2\frac{1}{2}\phi$, in violation of additivity for all $\phi \neq \frac{1}{2}\pi$. Similarly, the prescription of Ref. [4] does not work in this example, since $\mathrm{Re}\,\mathrm{tr}(P_{11}P_{21}P_{32}W) = \frac{1}{4}\cos\phi$, which is negative for $\frac{1}{2}\pi < \phi < \frac{3}{2}\pi$.

Example 2. Let a, b, c be three coplanar directions which are 120 degrees apart, and let $|+a\rangle$, $|-a\rangle$ be the states "spin up" and "spin down" in the a direction, and similarly for b and c. Let H be the four-dimensional space associated with two spin-$\frac{1}{2}$ particles. Consider the following family,

$$|+a\rangle|+b\rangle,\ |+a\rangle|-b\rangle,\ |-a\rangle|+b\rangle,\ |-a\rangle|-b\rangle,$$
$$|+a\rangle|+c\rangle,\ |+a\rangle|-c\rangle,\ |-a\rangle|+c\rangle,\ |-a\rangle|-c\rangle,$$
$$|+b\rangle|+c\rangle,\ |+b\rangle|-c\rangle,\ |-b\rangle|+c\rangle,\ |-b\rangle|-c\rangle,$$

and let W be the singlet spin state. We shall show that this family is not feasible. Put $X = \{1,2,3,4\}$, then our index set is X^4. Assume, by negation, that there is a probability measure μ on the subsets of X^4 which recovers the quantum mechanical probabilities for commutative cylinders.

The cylinders $A_1 = \{1,2\} \times X \times X$, $A_2 = X \times X \times \{1,2\}$, $A_3 = X \times \{1,3\} \times X$, $A_4 = \{1,3\} \times X \times X$ are commutative and each has probability $\frac{1}{2}$ (for example, A_1 corresponds to the history $(|+a\rangle\langle+a| \otimes I, I, I))$. Similarly, the cylinders $A_1 \cap A_3$, $A_1 \cap A_4$, $A_2 \cap A_3$, $A_2 \cap A_4$ are commutative and have probabilities $\frac{3}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, 0 respectively (for example the cylinder $A_1 \cap A_3$ corresponds with the history $(|+a\rangle\langle+a| \otimes I, I \otimes |+c\rangle\langle+c|, I))$, But these probabilities violate the Clauser–Horne inequality

$$\mu(A_1 \cap A_3) + \mu(A_2 \cap A_3) + \mu(A_1 \cap A_4) - \mu(A_2 \cap A_4) - \mu(A_1) - \mu(A_3) \leqslant 0,$$

which is valid in every probability space [10,8]. Therefore, the family is not feasible.

3. The maximum entropy principle

All that is required from a family of histories to count as feasible is that *some* probability measure that satisfy (2) exists. But if there is one such measure there are usually many. How are we to choose between them and determine the real probability of a history? Since we want to model decoherence, and since decoherence increases entropy we choose the following:

The maximum entropy principle. Given a feasible set F, choose the measure μ on the algebra of subsets of the index set, such that

$$H(\mu, F) = -\sum_{(j_1,j_2,\dots,j_n)\in I} \mu\{(j_1,j_2,\dots,j_n)\} \times \ln\mu\{j_1,j_2,\dots,j_n)\} \tag{5}$$

is maximized under the constraints $\mu(\alpha) = \mu(\alpha_1 \times \dots \times \alpha_n) = \mathrm{tr}(WP_{\alpha_1}P_{\alpha_2}\dots P_{\alpha_n})$, for every commutative cylinder α.

Condition (2) is therefore the feasibility condition (in the technical sense of the theory of optimization) of the convex program (5). If certain additional conditions are satisfied [11] the solution to the optimization problem can be written in the form

$$\mu\{(j_1,j_2,\dots,j_n)\} = Z^{-1}\exp\left(-\frac{1}{k} \times \sum_{\alpha\ \text{commutative}} \lambda(\alpha)\,\chi_\alpha(j_1,j_2,\dots,j_n)\right). \tag{6}$$

Here $k = k_F$ is a constant that may depend on the family F, and $\lambda(\alpha)$ and the (real) Lagrange multipliers (there is one for every commutative cylinder α). The function χ_α is the indicator of α, so that $\chi_\alpha(j_1,\dots,j_n) = 1$ if $(j_1,\dots,j_n) \in \alpha$ and $= 0$ otherwise, and Z is the partition function,

$$Z = \sum_{(j_1,j_2,\dots,j_n)\in I} \exp\left(-\frac{1}{k} \times \sum_{\alpha\ \text{commutative}} \lambda(\alpha)\,\chi_\alpha(j_1,j_2,\dots,j_n)\right). \tag{7}$$

Take as an example the feasible family F in (3). The probability measure given in (4) maximizes entropy. The reason is that μ is obtained by taking the commutative cylinders as *independent*. When this can be done in a consistent manner we obtain the required optimum.

We shall now express (6), (7) in terms of the Hilbert space formalism. Let $H^{(n)} = H \otimes H \otimes \dots \otimes H$ be the n-times tensor product of H by itself. For a cylinder $\beta = \beta_1 \times \beta_2 \times \dots \times \beta_n$ (not necessarily a commutative one) consider the projection operator on $H^{(n)}$

$$P_\beta = P_{\beta_1} \otimes P_{\beta_2} \otimes \ldots \otimes P_{\beta_n},$$
$$P_{\beta_i} = \sum_{j\in\beta_i} |i,j\rangle\langle i,j| . \tag{8}$$

Now define

$$A = \sum_{\alpha \text{ commutative}} \lambda(\alpha)\, P_\alpha . \tag{9}$$

Then A is a Hermitian operator in $H^{(n)}$ and we have

$$Z = \mathrm{tr}(\mathrm{e}^{-A/k}),$$
$$\mu(\beta) = \frac{\mathrm{tr}(P_\beta \mathrm{e}^{-A/k})}{\mathrm{tr}(\mathrm{e}^{-A/k})},$$
$$\text{for } \beta = \beta_1 \times \beta_2 \times \ldots \times \beta_n . \tag{10}$$

The traces are computed on $H^{(n)}$. We can therefore define the *decoherence functional* [6] to be

$$\mathrm{d}(\beta,\gamma) = \frac{\mathrm{tr}(P_\beta \mathrm{e}^{-A/k} P_\gamma^\dagger)}{\mathrm{tr}(\mathrm{e}^{-A/k})} . \tag{11}$$

Now, note that for any trace class operators $A_1, A_2, \ldots, A_n$ on H we have $\mathrm{tr}(A_1, A_2, \ldots, A_n) = \mathrm{tr}(A_1 \otimes A_2 \otimes \ldots \otimes A_n S)$, where the first trace is taken on H and the second trace on $H^{(n)}$. Here S is the operator on $H^{(n)}$ defined on product vectors by $S(x_1 \otimes x_2 \otimes \ldots \otimes x_n) = x_2 \otimes x_3 \otimes \ldots \otimes x_n \otimes x_1$ and extended by linearity to the rest of $H^{(n)}$. (See Ref. [6].) If we define

$$\bar{W} = \frac{1}{n}(W \otimes I \otimes \ldots \otimes I + I \otimes W \otimes \ldots \otimes I + \ldots$$
$$+ I \otimes I \otimes \ldots \otimes W) , \tag{12}$$

then $\bar{W}S = \bar{W}$, and $\bar{W}$ is a state on $H^{(n)}$. Moreover, if *α is commutative* we have

$$\begin{aligned}\mathrm{tr}(P_\alpha \bar{W}) &= \mathrm{tr}(P_\alpha \bar{W} S)\\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{tr}(P_{\alpha_1} \otimes \ldots \otimes P_{\alpha_i} W \otimes \ldots \otimes P_{\alpha_n} S)\\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{tr}(P_{\alpha_1} \ldots P_{\alpha_i} W P_{\alpha_{i+1}} \ldots P_{\alpha_n})\\ &= \mathrm{tr}(W P_{\alpha_1} P_{\alpha_2} \ldots P_{\alpha_n}) .\end{aligned}$$

Therefore, $\mathrm{tr}(\bar{W}A) = \sum_{\alpha \text{ commutative}} \lambda(\alpha)\, \mathrm{tr}(W P_{\alpha_1} \times P_{\alpha_2} \ldots P_{\alpha_n})$. If we substitute (6) into (5) and use this fact, we obtain another expression for the entropy,

$$H(\lambda, F) = k \ln[\mathrm{tr}(\mathrm{e}^{-A/k})] + \mathrm{tr}(\bar{W}A) . \tag{13}$$

Using the duality theorem for convex programming (the Kuhn–Tucker theorem [12]) we see that maximizing entropy in (5) is equivalent to the search for the Lagrange multipliers λ – which determine the Hermitian operator A in (9) – such that (13) attains its *minimum*. (To see that take $\partial H/\partial\lambda(\alpha) = 0$ and compare with (6).) The constraints on the λ's are those that guarantee that the minimum exists. Namely, that the matrix $\partial^2 H/\partial\lambda(\alpha)\,\partial\lambda(\alpha')$ is positive definite at the point where $\partial H/\partial\lambda(\alpha) = 0$.

To sum up: With every feasible family of histories there corresponds a Hermitian operator A on the tensor product of the Hilbert space by itself n times. The probability of a history is given by (10) and more generally the decoherence functional by (11). A is the operator that minimizes (13), and the conditions under which that minimum exists entail, in particular, the conditions of feasibility.

4. The MinMax principle

The concept of *feasible family of histories* is broader than all the proposals made so far. We thus allow for interference between two histories in a family, if it is not too big. On the other hand, we would like to maintain decoherence at its maximal possible level. Consequently we impose on any feasible family of histories the probability measure that maximizes entropy (while, of course, maintaining the probabilities dictated by quantum mechanics).

There is really no point in *enlarging* the set of possible families unless we have a way of choosing one (or at any rate, a few) which will represent the one and only real world. Here we agree with the view expressed by Dowker and Kent [7]. But how are we to choose? Presumably we have to explain why the actual universe is nearly classical and populated by IGUSes[3] of a particular kind. We take the key from Wiener [13]:

[3] The term IGUS stands for "Information Gathering and Utilizing System" and was introduced by Gell-Mann and Hartle [3].

... In the world with which we are immediately concerned there are stages which, though they occupy an insignificant fraction of eternity, are of great significance for our purposes, for in them entropy does not increase and organization and its correlative, information, are being built up.

Thus we conjecture that what distinguishes the near classical family of histories from the other feasible families is its *low* entropy. To "find" it, we propose to look for a family F that is feasible with respect to the initial state of the universe, compatible with its dynamical laws, and for which the entropy is *minimal*. Combining this idea with (5) we are trying to solve

$$\min_{F} \max_{\mu} H(F, \mu) , \tag{14}$$

where F is ranging over all feasible families of histories, as above, and for each F, μ is ranging over all measures that are satisfying the feasibility constraints (2).

So far there is no evidence for the conjecture except for the observation that the "world with which we are immediately concerned" has low entropy. To see whether (14) yields such a world (given a realistic initial state) requires further research, including large scale simulations.

Acknowledgement

I would like to thank Meir Hemmo for fruitful discussions and for his encouragement. This work was supported in part by the Edelstein Center for the History and Philosophy of Science at the Hebrew University.

References

[1] R.B. Griffiths, J. Stat. Phys. 36 (1984) 219.

[2] R. Omnès, J. Stat. Phys. 53 (1988) 893, 933, 957; 57 (1989) 357; Ann. Phys. (NY) 201 (1990) 354; Rev. Mod. Phys. 64 (1992) 339.

[3] M. Gell-Mann, J.B. Hartle, in: Complexity, entropy, and the physics of information, Santa Fe Institute Studies in the sciences of Complexity, Vol. III, ed. W.H. Zurek (Adisson-Wesley, Redwood City, 1990) p. 425; Phys. Rev. D 47 (1993) 3345.

[4] S. Goldstein, D.N. Page, Phys. Rev. Lett. 74 (1995) 3715.

[5] E.T. Janes, Papers on probability, statistics and statistical physics (Reidel, Dordrecht, 1983).

[6] C.J. Isham, N. Linden, S. Schreckenberg, J. Math. Phys. 35 (1994) 6360.

[7] F. Dowker, A. Kent, J. Stat. Phys. 82 (1996) 1575.

[8] I. Pitowsky, Quantum probability quantum logic, Lecture Notes in Physics 321 (Springer, Berlin, 1989); Math. Programming A 50 (1991) 359.

[9] I. Pitowsky, Brit. J. Phil. Sci. 45 (1994) 95.

[10] J.F. Clauser, M.A. Horne, Phys. Rev. D 10 (1974) 526.

[11] J.N. Kapur, H.K. Kesaven, Entropy optimization principles with applications (Academic Press, Boston, 1992).

[12] A.I. Peressini, F.E. Sullivan, J.J. Uhl Jr., The mathematics of nonlinear programming (Springer, Berlin, 1980).

[13] N. Wiener, The human use of human beings (Doubleday, Garden City, NY, 1954) p. 31.