Journal of Statistical Physics, Vol. 82, Nos. 5/6, 1996

On the Consistent Histories Approach to Quantum Mechanics

Fay Dowker[1,2] and Adrian Kent[3]

Received September 1, 1994; final June 9, 1995

We review the consistent histories formulations of quantum mechanics developed by Griffiths, Omnès, and Gell-Mann and Hartle, and describe the classification of consistent sets. We illustrate some general features of consistent sets by a few simple lemmas and examples. We consider various interpretations of the formalism, and examine the new problems which arise in reconstructing the past and predicting the future. It is shown that Omnès' characterization of true statements—statements which can be deduced unconditionally in his interpretation—is incorrect. We examine critically Gell-Mann and Hartle's interpretation of the formalism, and in particular their discussions of communication, prediction, and retrodiction, and conclude that their explanation of the apparent persistence of quasiclassicality relies on assumptions about an as-yet-unknown theory of experience. Our overall conclusion is that the consistent histories approach illustrates the need to supplement quantum mechanics by some selection principle in order to produce a fundamental theory capable of unconditional predictions.

KEY WORDS: Quantum mechanics; quantum cosmology; consistent histories; decoherence.

1. INTRODUCTION

Quantum theory has been so successful that it is natural to wonder whether it can be applied to the universe as a whole. It has long been clear that to do any such thing we would need a new way of thinking since, while the standard Copenhagen interpretation works well in the laboratory,

[1] Physics Department, University of California Santa Barbara, Santa Barbara, California 93106.

[2] Isaac Newton Institute for Mathematical Sciences, Cambridge CB3 0EH, U.K.

[3] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, U.K.

it cannot be applied to closed systems. Recently Griffiths, [1,2] Omnès,[3,4] and Gell-Mann and Hartle[5–9] have indeed set out a new way of looking at quantum mechanics and quantum field theory, in which the fundamental objects are consistent sets of histories. So, can the consistent histories formalism do the job required? Could it be the correct language in which to give an intrinsically quantum description of cosmology? Is it complete? Our aim here is to examine these questions. We shall attempt to clarify and extend the arguments in the literature, to make explicit the premises on which they rely, and to clarify the present status of interpretations of the consistent histories formalism as scientific theories. We briefly sketch our approach in the rest of this introduction.

We should say, first, that the present article contains a much greater word-to-equation ratio than is usual in a scientific paper. We have found this, unfortunately, to be unavoidable since, while some of the questions that arise in the area are technical, the deepest controversies involve the interpretation of the formalism and are not to be resolved simply by calculating. Many contradictory arguments have been advanced, and we have tried comprehensively to deal with the main proposals. A short letter[10] illustrating the main features of the consistent histories formalism may be of some help to the reader.

By a quantum cosmological theory we mean some precise, observer-independent formulation of quantum theory (such as, perhaps, a consistent histories formulation) together with some theory of the boundary conditions. Without a quantum theory of gravity, the best one can hope for is an effective theory. However, given a fixed spacetime with preferred timelike directions, and given some theory about the density matrix on some very early spacelike surface, the question remains: what should, or could, we ask from such a theory? It is surely too much to hope that it will predict in detail the world we experience. However, one might at least hope to predict the gross features: the existence and persistence of (what appears to us to be) a classical world with very large scale structure; the fact that matter clumps and fields in this classical world generally obey classical equations of motion to a very good degree of approximation; the fact that microscopic particles, when sporadically interacting with the classical world, follow the probabilistic laws of Copenhagen quantum theory.

Obtaining predictions such as these from a formulation of quantum theory is notoriously tricky. It clearly is not good enough to show that by some procedure one can obtain from quantum theory a picture which gives a satisfactory account of cosmology, if one can equally well extract by similar procedures many other pictures which give completely different, unsatisfactory accounts. So we must first set out a way of obtaining cosmological data from quantum theory. Given any such procedure, we must then characterize all the sets of data which it gives. Only then can we tell whether we have a sensible and accurate quantum theory of cosmology, a potentially useful theory with an unresolved selection problem, or something less coherent. The theory, in other words, is more than just the equations: it includes an explanation of how the equations are to be interpreted.

There does seem to be a consensus among physicists that these demands are reasonable. Still, it can be tempting to extract physical facts from the mathematics whenever they are appealing, to refuse to do so whenever they would be embarrassing, and not to enquire too closely as to whether any discernible principle underlies these choices. In particular, when a new approach, such as the consistent histories formalism, is discovered, it is sometimes suggested that the only interesting problem is extending the mathematics, and that interpretation is just a matter of common sense. Yet this view cannot easily be maintained, once one recalls the intensity of the debates which led eventually to the Copenhagen interpretation. It seems even more naive when one considers the history of various attempts at a many-worlds interpretation.[4] And it is certainly not the view of the authors of the consistent histories formalism, who have each set out their interpretational ideas at some length.

We shall try, then, to be careful to specify exactly how one translates the consistent histories formalism into physics. Fortunately, as we describe later, there are a number of ways to do this, so that it is possible to use the formalism to obtain physical results. To understand the nature of these results, we need some description of the various consistent sets and some characterization of the events they each describe. In Section 2, we set out the formalism, discuss the various consistency criteria which have been proposed, and examine equivalence relations and maximality criteria for consistent sets.

Any realistic discussion of the general properties of consistent sets in cosmological models is hopelessly impractical. In fact, exploratory calculations on finite-dimensional Hilbert spaces show that, even given a simple initial density matrix, the spaces of consistent sets of histories become complicated very quickly; it is a formidable task to calculate the full class of consistent sets of histories in dimensions greater than about three. In

[4] A collection of early papers on the subject can be found in DeWitt and Graham's book.[11] A detailed discussion of many-worlds interpretations is beyond our present scope; several critiques can be found in the literature.[12-14] Briefly, our view is that although the ideas of Everett *et al.* have motivated interesting work, including some of the papers we shall discuss, no well-defined scientific theory has yet been described in the many-worlds literature except for Bell's intentionally pathological Everett–de Broglie–Bohm hybrid,[15] and that most of the ideas hinted at in earlier many-worlds papers can more naturally be understood in the language of consistent histories.

Section 3 we set out the general algorithm for classifying consistent sets and give some fragmentary results describing the properties of the space of consistent sets. These results, which are obtained by exact calculation of the very simplest examples, give a limited and qualitative test of the conclusions one can draw from naive counting arguments. Fortunately, we need these results and counting arguments only to support propositions which already seem very plausible. We also set out some elementary but useful lemmas and illustrative examples which seem to characterize the basic mathematical facts that lead to problems—or, at least, unfamiliar features—in the various interpretations of the formalism.

In Section 4 we briefly examine two of Gell-Mann and Hartle's proposals. The first is the use of approximate consistency. It is easy to set out the conditions for a set of histories to have probabilities satisfying the usual sum-rules. Gell-Mann and Hartle[5, 16] argue that it is natural to consider, on an equal footing with these consistent sets of histories, those sets for which the probability sum rules are slightly violated. They point out that, if the violation is sufficiently small, no experiment could detect the discrepancy, and that in any case one can remove the sum rule violation by an ad hoc, but equally undetectable, renormalization of the probabilities. Still, this seems a rather casual disruption of the mathematical structure of a fundamental theory. It also seems unnecessary. The intuition one can glean from calculations of consistent histories on finite-dimensional Hilbert spaces is admittedly limited, but it does suggest that the solution space of the exact consistency equations is quite large enough for all theoretical purposes so far suggested. In particular, naive but plausible counting arguments suggest that, in the neighborhood of generic approximately consistent sets of histories, an exactly consistent set can be found. This, we shall suggest, means that there is no need to consider approximately consistent sets in any fundamental formulation of the theory. Given that there are many consistent, incompatible sets of histories, it is natural to ask whether there might be any good way of singling one out. The second proposal of Gell-Mann and Hartle that we comment on is that there is some simple and natural measure of the quasiclassicality of consistent sets, which is sharply peaked (ref. 5, pp. 448–449). While this goes beyond the framework of consistent histories, it is an attractive idea. Finite-dimensional history calculations give no clue as to its plausibility, but illustrate in a little more detail the questions involved.

In Section 5, we set out various possible interpretations of the formalism and analyze their status as scientific theories. Our main preoccupation here is the extent to which any interpretation allows us to make deductions about the past or present and predictions of the future. We explain that there are several different senses in which a consistent histories theory can be said to account for the past, and try to determine which particular sense is intended by particular authors. We then attempt to understand the extent to which any consistent histories interpretation can account for our perception of a single, persisting, nearly classical world. Since one can show that, unless the particular set of histories which describes our experiences up to the present has remarkable and previously unsuggested properties, present-day quasiclassicality in a consistent set does not imply persisting quasiclassicality in that set, and in fact quasiclassicality does not persist in generic consistent future extensions; this is clearly a nontrivial problem. We also attempt to understand the extent to which consistent histories interpretations allow us to make ordinary physical inferences. Can we show, for instance, from terrestrial experiments, that the moon is following an essentially classical orbit around the earth, or more generally that the world around us is approximately described by classical physics?

Here we take the interpretations case by case. We first consider the logical interpretation of the formalism which has been set out by Omnès. Omnès uses the formalism to analyze the logical relations among various collections of statements, and reaches an interpretation of the formalism which is logically consistent and which, in a certain sense, extends the Copenhagen interpretation. There are interesting ideas in this approach. However, we conclude that on a key point—the identification of true statements which, given a particular set of data, can be deduced unambiguously—Omnès' conclusions are incorrect. We argue that in fact there are generally no useful true statements describing the future, and only a very limited class of true statements about the past. Omnès accepts this criticism. Work on a revised criterion for "truth" is in progress: as we shall explain, we are sympathetic to this program, though we believe it will necessarily go beyond the consistent histories formalism. Meanwhile, as presently formulated, Omnès' interpretation does not allow us to make any unconditional tests of the formalism, and in particular gives no explanation for the apparent persistence of quasiclassicality.

We next mention an interpretation set out by Griffiths, which relies on a notion of quantum logic rather different from that of Omnès. As we explain, Griffiths' logical structure, in itself, allows no useful notion of prediction or retrodiction. We then set out a many-histories interpretation, inspired by Griffiths' ideas, but requiring only classical logic. This natural-seeming interpretation is quite different in spirit from the previous two; it allows different arguments to be made and suggests different directions for possible extensions of the formalism. However, its practical implications are essentially the same: the past is unreconstructable, and the persistence of quasiclassicality is inexplicable.

We come next to Gell-Mann and Hartle's work on the interpretation of the consistent histories formalism in the light of quantum cosmology. Gell-Mann and Hartle argue that creatures such as ourselves, whose evolution is best described quasiclassically, will naturally perceive a persisting quasiclassical world (ref. 5, p. 454). This is an important claim which deserves careful scrutiny: if it holds, then the consistent histories approach, together with a theory of the boundary conditions, constitutes a genuinely predictive scientific theory; if it fails, then it seems very unlikely that any useful and unconditional predictions of our future observations can be made using the existing formalism. We attempt to analyze Gell-Mann and Hartle's arguments and to identify the premises on which they rely. We conclude that the argument that quasiclassicality will appear to persist relies crucially on assumptions about an as-yet-unknown theory of experience. We note also that, although they often write in ordinary language about the reality of past events and the possibility of communication, the only interpretation we can find in which Gell-Mann and Hartle's prediction of our experience of a quasiclassical world can be hoped to be made entails solipsism.

In the final subsection on interpretations, we describe a minimalist interpretation in which our continuing perception of a quasiclassical world is very clearly an unexplainable mystery. We argue that the Omnès, Griffiths, and many-histories interpretations can all be characterized as variants of this "unknown set" interpretation.

The apparent persistence of quasiclassicality is central to our discussions. In Section 6 we look at another aspect of the question. We show that generic cosmological theories which contain a consistent set describing quasiclassical behavior up till some time do not contain a consistent set in which quasiclassicality persists. More precisely, we show that, if the initial density matrix is pure and some consistent collection of present and past facts is given, and if this collection of facts has probability p when the final density matrix is trivial, then, given any nontrivial statement about the future, generic final density matrices for which the given facts are also consistent with probability p render the future statement inconsistent. In other words, it is impossible to ascribe consistency to *any* nontrivial statement about the future without invoking some general theory of the final density matrix. In Section 7, we examine the case for imposing any consistency criterion. We point out that there are logically coherent interpretations involving inconsistent sets, and indeed that it is not completely impossible that the world as we see it is described by an inconsistent set. However, the success of experimental science to date is a compelling argument against this possibility.

In the concluding section, we summarize the earlier discussions and our view of the current status of the consistent histories formalism and its interpretations. We concur with its authors that it is a natural approach to quantum theory. However, we conclude that, whichever interpretation one adopts, the formalism at present provides only a very weakly predictive scientific theory. We suggest that the formalism ought to be augmented, and briefly discuss how this could be done.

We close this introduction with a few historical remarks. The consistent histories formalism has evoked a good deal of interest over the last decade, and has provoked many thoughtful criticisms. Griffiths' original article[1] included a discussion of some novel and counterintuitive features of consistent sets of histories, and he, Omnès, and Gell-Mann and Hartle have since devoted much effort to elaborating and explaining the consequences of such features. In a thoughtful critique,[17] whose conclusions we cite later, D'Espagnat examined the early papers of Griffiths and Omnès, and noted some divergence between a language of definite physical events and formal discussions in which these reassuringly realistic pictures cannot quite be justified. Griffiths[2] and Omnès[4] have developed their ideas further in response to D'Espagnat. Among other critics, Zurek, in an illuminating discussion[18] of decoherence-based approaches to the problems of quantum mechanics, has emphasized that consistency does not constrain histories enough to eliminate even very nonclassical evolutions of the state vector, and argued for an interpretation based on the division of the universe into subsystems. Dürr, Goldstein, and Zanghí, (DGZ) in their careful and detailed analysis[19] of Bohmian mechanics, noted the analogy between the trajectories of Bohmian mechanics and Gell-Mann and Hartle's speculative program aimed at understanding how quasiclassical domains might be characterized and investigating their properties and the possibility that an essentially unique domain might exist.[5] Albrecht[20] discussed the relationship between descriptions of a combined system and apparatus in terms of consistent histories and of the Schmidt decomposition, with the aid of a series of computational experiments. Paz and Zurek[21] further explored this relationship theoretically, and showed that, when particular dynamical approximations hold, a large class of branch-dependent consistent histories can be identified.

It has, in short, been recognized for some time that in the consistent histories approach to quantum cosmology there are likely to be a large

[5] To quote DGZ: "they [Gell-Mann and Hartle] propose a program to extract from the quantum formalism a 'quasiclassical domain of familiar experience' which, if we understand them correctly, defines for them the basic ontology of quantum mechanics." In fact, as we explain later, DGZ do not understand Gell-Mann and Hartle correctly on this point, although we agree this would be a clearer and more natural interpretation than any Gell-Mann and Hartle offer. With this caveat, we recommend DGZ's remarks (Appendix of ref. 19, point 22) to the reader.

number of consistent sets, and that many of these will not be standard quasiclassical descriptions of the world. Whether this causes scientific problems (and if so which), and whether these problems can be dealt with by simple postulates or interpretational arguments, are questions which depend very much on the details. We here try, *inter alia*, to characterize much more precisely the general properties of consistent sets, to analyze carefully the scientific problems which these properties cause, and to engage in detail with the interpretational arguments which Omnès, Gell-Mann and Hartle, and Griffiths offer in defense of the formalism.

There are, of course, interesting formulations of quantum theory outside the consistent histories framework—most notably, in our view, Bohmian mechanics[22] and Samols' recent stochastic model of a relativistic quantum field theory.[23] There is no doubt that these approaches solve some of the problems afflicting current interpretations of the consistent histories formalism. In particular they explain the persistence of quasiclassicality by introducing auxiliary variables—often misleadingly called "hidden variables"—which, very roughly speaking, do the job that a good set selection criterion would do within the consistent histories framework. On the other hand, these approaches raise questions (of relativistic invariance in the case of Bohmian mechanics, and of generalizability and naturalness) of their own. These questions, though very interesting, go beyond our scope, and we shall generally avoid comparisons of the consistent histories framework to other approaches: the interested reader is encouraged to consult refs. 19, 23, and 24.

2. CONSISTENT HISTORIES

2.1. Formalism

It is simplest to describe the consistent histories formalism as it applies to nonrelativistic quantum mechanics, in the Heisenberg picture, using the language of projection operators and density matrices. As in standard quantum mechanics, we assume that a separable Hilbert space $\mathscr{H}$ and Hamiltonian H are given, that Hermitian operators correspond to observables, that the commutation relations among the Hamiltonian and physically interesting observables (such as position, momentum, and spin) have been specified, and that the operators corresponding to the same observables at different times are related by

$$P(t) = \exp(iHt/h)\, P(0)\, \exp(-iHt/h) \tag{2.1}$$

We are interested in a system (in principle, the universe) whose initial density matrix ρ_i is given. As usual, we require that $\mathrm{Tr}(\rho_i) = 1$ and that ρ_i is Hermitian and positive semidefinite .

The consistent histories formalism also admits an interesting time-symmetric generalization of quantum mechanics,[25] in which both initial and final boundary conditions are specified via ρ_i and a *final* density matrix ρ_f which is Hermitian positive semidefinite and normalised[6] so that $\mathrm{Tr}(\rho_i \rho_f) = 1$. This is particularly interesting in the context of quantum cosmology, removing as it does the usual time asymmetry in the boundary conditions (see, e.g., ref. 16, Section IV.2). We shall often include a final density matrix, but, where the value of ρ_f is not explicitly discussed, it can be set to I without affecting the arguments. We shall use the convention that, where nontrivial final density matrices are excluded, the initial density matrix is called ρ rather than ρ_i. The initial and final density matrices give boundary conditions for the system at times t_i and t_f, with $t_i < t_f$; these times can generally be set to 0 and ∞. The basic physical events we are interested in correspond to sets of orthogonal Hermitian projections $P^{(i)}$, with

$$\sum_i P^{(i)} = 1 \qquad \text{and} \qquad P^{(i)} P^{(j)} = \delta_{ij} P^{(i)} \tag{2.2}$$

We shall want to consider the projections as attached to some particular time t; where we wish to emphasize this, we shall write $P^{(i)}$ as $P^{(i)}(t)$. It may be useful to visualize these as projections onto disjoint intervals of the spectrum of some observable, for example, the position of a single particle. For a single set of this type, with trivial ρ_f, standard quantum mechanics supplies an interpretation: specifying the set corresponds to specifying the possible results of a particular experimental measurement at time t, and—if the experiment is carried out—the probability of obtaining result i is

$$p(i) = \mathrm{Tr}(P^{(i)}(t)\, \rho_i P^{(i)}(t)) \tag{2.3}$$

In fact, so long as a single set of projections has somehow been fixed, it is perfectly consistent to speak about the state of the system without invoking the notion of measurement. If we wish, we can simply postulate that the projections define an extra variable which represents the "real" physical description of the system, and say that at time t the system *is* in the range of precisely one of the projections, and the probability is $p(i)$ that

[6] With this normalization, since $Tr(\rho_f) \neq 1$ in general, ρ_f is not, strictly speaking, a density matrix. We shall nonetheless refer to it as such, since both the normalization and the terminology are convenient.

this projection is $P^{(i)}(t)$—although, of course, the formalism does not imply this interpretation.

One needs to take more care in assigning probabilities to *histories* —that is, to sequences of events in time. We now allow nontrivial ρ_f, and suppose we have sets of projections $\sigma_j = \{P_j^{(i)};\ i = 1, 2,..., n_j\}$ with $n_j > 1$ and j running from 1 to n, at times t_j with $t_i < t_1 < ... < t_n < t_f$, which satisfy the conditions (2.2). Then the histories given by choosing one projection from each σ_j in all possible ways are an exhaustive and exclusive set of alternatives, $\mathscr{S}$. It seems natural to interpret a history $\{P_1^{(i_1)}(t_1),..., P_n^{(i_n)}(t_n)\}$, where $P_n^{(i_j)}(t_j) \in \sigma_j$, as corresponding to the proposition that the system was in the range of P^{i_1} at time t_1, in the range of $P^{(i_2)}$ at time t_2, and so on. The natural rule for the probability of this history is[7]

$$p(i_1 \cdots i_n) = \mathrm{Tr}(P_n^{(i_n)} \cdots P_1^{(i_1)} \rho_i P_1^{(i_1)} \cdots P_n^{(i_n)} \rho_f) \tag{2.4}$$

However, one would also like the standard probability sum rules for exclusive events to be reflected in the mathematics in the natural way. We shall discuss this point further in Section 6, but let us give the standard argument here. It is that the projection operator $(P_r^{(a)} + P_r^{(b)})(t_r)$ (where a and b are distinct labels in the range 1 to n_r) should correspond to the statement that the system was in the union of the ranges of $P_r^{(a)}$ and $P_r^{(b)}$ at time t_r, and the probability of a history including this event should be calculable either directly, using this projection operator, or as the sum of the probabilities of the two finer-grained histories in which the system was, respectively, in the range of $P_r^{(a)}$ and in the range of $P_r^{(b)}$. That is, for consistency we require that

$$\begin{aligned}
&\mathrm{Tr}(P_n^{(i_n)} \cdots (P_r^{(a)} + P_r^{(b)}) \cdots P_1^{(i_1)} \rho_i P_1^{(i_1)} \cdots (P_r^{(a)} + P_r^{(b)}) \cdots P_n^{(i_n)} \rho_f) \\
&\quad = \mathrm{Tr}(P_n^{(i_n)} \cdots P_r^{(a)} \cdots P_1^{(i_1)} \rho_i P_1^{(i_1)} \cdots P_r^{(a)} \cdots P_n^{(i_n)} \rho_f) \\
&\qquad + \mathrm{Tr}(P_n^{(i_n)} \cdots P_r^{(b)} \cdots P_1^{(i_1)} \rho_i P_1^{(i_1)} \cdots P_r^{(b)} \cdots P_n^{(i_n)} \rho_f)
\end{aligned} \tag{2.5}$$

and similar conditions involving sums of more than two projections. These conditions do not hold for general sets $\mathscr{S}$. As Griffiths[1] pointed out, Eq. (2.4) will be consistent with these probability sum rules if, and only if, the projection operators satisfy

$$\begin{aligned}
&\mathrm{Re}(\mathrm{Tr}(P_n^{(i_n)} \cdots P_r^{(i_r)} \cdots P_1^{(i_1)} \rho_i P_1^{(i_1)} \cdots P_r^{(i'_r)} \cdots P_n^{(i_n)} \rho_f)) \\
&\quad = \delta_{i_r i'_r} p(i_1 \cdots i_n)
\end{aligned} \tag{2.6}$$

[7] This will be recognized as the formula for the probability, calculated in the Copenhagen interpretation, that the history is *measured* in a sequence of ideal measurements.

for all r and all choices of $i_1, \ldots, i_n$ and i'_r. We shall follow Gell-Mann and Hartle in imposing the stronger condition that

$$\mathrm{Tr}(P_n^{(i_n)} \cdots P_1^{(i_1)} \rho_i P_1^{(j_1)} \cdots P_n^{(j_n)} \rho_f) = \delta_{i_1 j_1} \cdots \delta_{i_n j_n} p(i_1 \cdots i_n) \tag{2.7}$$

As Gell-Mann and Hartle stress, it is natural to require (2.7) if one is modeling the physical decoherence mechanisms of quasiclassical variables; it also seems to be mathematically the most natural, and most convenient, consistency condition. When the conditions (2.7) hold, we say that the set of histories $\mathscr{S}$ is consistent.[8] We use (2.7) rather than the Griffiths conditions (2.6) because our general preoccupation will be with the number of consistent sets and the problems this causes in prediction and retrodiction. We look favorably on any natural criterion which reduces the number of sets, in the hope of reducing these problems, or at least—since we believe that adopting (2.6) would not significantly change our discussion—in assuring ourselves that we have done what we can.

There are other versions of the theory, for example, the path integral approach (refs. 16 and 26; ref. 5, pp. 432–434), in which a set of coarse grained histories is a partition of the set of fine grained paths, and the generalization to "history-dependent projections" (ref. 3; ref. 5, pp. 450–451; ref. 9, p. 3352), in which the set of projections at a certain time may depend on the previous choices of projections. Other interesting generalizations of the formalism are also being examined.[27] It is clear that the nonrelativistic formalism which we examine here cannot be fundamental, and it will be necessary eventually to investigate whether the other proposals give so radically different a picture of quantum cosmology that they evade the difficulties we discuss later. However, this seems unlikely at first sight, and since it anyway seems sensible to look first at the simplest nonrelativistic form of the theory, we shall restrict our attention in this work to the projection operator formalism described above.

2.2. Characterizing Histories

We turn to the question of when we are to regard two sets of histories as different. There clearly are equivalence relations which can naturally be imposed on the consistent sets.[16] It would be useful to characterize such

[8] The term *medium decoherent*, coined by Gell-Mann and Hartle, is used in the literature to describe sets satisfying (2.7). *Weakly decoherent* sets satisfy only the real part of (2.7). This condition is motivated by the path integral version of the nonrelativistic theory, in which not only individual projections, but entire histories are required to obey probability sum rules. Note that Griffiths' original consistency condition (2.6) is weaker still. For more comments see ref. 9. We prefer the condition (2.7) as explained in the text, and Griffiths' terminology because of its implication of virtue.

relations completely, but this has not yet been done, and may not even be a well-defined problem. Here we just make a few simple observations.

Let us start with the complete characterization of a set of histories that we have so far:

$$\mathcal{S} = (\rho_i, t_i, \rho_f, t_f, \{\sigma_j\}, \{t_j\}) \tag{2.8}$$

with $j = 1, 2, \ldots, n$ and $t_i < t_1 < \ldots < t_n < t_f$ and the σ_j complete sets of orthogonal projections $\{P_j^{(i)}: i = 1, 2, \ldots, n_j\}$ as above. We are interested in possible equivalence relations on the set of all such sets of histories. In the following discussion, we denote possible relations by $\sim$, and take their reflexivity, symmetry, and transitivity for granted. Thus, for example, if we impose the relation $A \sim B$, we also require $B \sim A$.

The first question to decide is whether any equivalence relations should preserve the physical interpretation of the consistent sets we actually use for calculations, or whether it is only the algebraic structure that need be preserved. The first seems more natural if one accepts the argument (which we shall make later) that some selection criterion will have to be found to explain why we use particular sets, since one would hope that a good selection criterion would rely on the Hamiltonian and canonical variables, would assign times to the relevant projections, and so forth. If one intends to use the formalism while treating all sets democratically, the second alternative is probably more sensible.

We start with the first, wherein equivalence relations respect the full Hamiltonian structure and thus should preserve the time coordinates in (2.8). Equations (2.7) are invariant under arbitrary unitary transformations, but if we take the Hamiltonian H to be a fixed operator then we consider consistent sets to be physically equivalent only if related by a Hamiltonian-preserving unitary map, that is

$$(\rho_i, \rho_f, \{\sigma_j\}) \sim (U\rho_i U^{-1}, U\rho_f U^{-1}, \{U\sigma_j U^{-1}\}) \tag{2.9}$$

if and only if U is a unitary operator such that $H = UHU^{-1}$. Here $U\sigma_j U^{-1}$ is shorthand for the set of projections $\{UP_j^{(i)} U^{-1}: i = 1, \ldots, n_j\}$. We can alternatively, following Gell-Mann and Hartle,[28] consider equivalences between complete theories, which include the Hamiltonian; in this case we have

$$(H, \rho_i, \rho_f, \{\sigma_j\}) \sim (UHU^{-1}, U\rho_i U^{-1}, U\rho_f U^{-1}, \{U\sigma_j U^{-1}\}) \tag{2.10}$$

On the other hand, even if one anticipates a set selection criterion, the arguments in favor of restricting consideration to Hamiltonian-preserving equivalences are compelling but not quite conclusive. Many of the consistent sets in a theory describe a rich mathematical structure, even if a

Hamiltonian is not given: it may seem implausible that this structure alone could be sufficient to characterize the quasiclassical world we see, but it is hard to see how one could argue the point with complete certainty.

We shall take a pragmatic approach. We chose the Gell-Mann–Hartle consistency conditions rather than the Griffiths conditions because they were mathematically more convenient and led to fewer consistent sets; likewise, in the calculations we set out later we allow equivalence relations which do not preserve the Hamiltonian structure because they lead to simpler calculations and to fewer equivalence classes. If we restricted ourselves to relations preserving the structure, our later arguments and illustrative calculations describing the multiplicity of equivalence classes would only be strengthened.

So, we first assume that the only relevance of the times $\{t_j\}$ and $\{t_i, t_f\}$ is their ordering. Thus

$$(\rho_i, t_i, \rho_f, t_f, \{\sigma_j\}, \{t_j\}) \sim (\rho_i, t'_i, \rho_f, t'_f, \{\sigma_j\}, \{t'_j\}) \tag{2.11}$$

if $t'_i < t'_1 < \ldots < t'_n < t'_f$, and we henceforth drop the time labels.

Second, impose the relation

$$(\rho_i, \rho_f, \{\sigma_j\}) \sim (U\rho_i U^{-1}, U\rho_f U^{-1}, \{U\sigma_j U^{-1}\}) \tag{2.12}$$

where U is any unitary operator on the Hilbert space. [In fact, we can take $U \in SU(\mathscr{H})$, since both sides of (2.12) are obviously equal for $U = \exp(i\theta) I$.]

With these relations, although the members of an equivalence class are essentially indistinguishable as sets of projection operators, some may be much more easily related to simple physical observables than others. Given a consistent set, one would have to search through its equivalence class to see whether (for example) it could be represented in terms of projections describing coarse grained particle trajectories.

A careful and detailed discussion is given in Gell-Mann and Hartle's recent elegant treatment of equivalence relations within the formalism,[28] which is intended to supersede their earlier discussions (ref. 5; ref. 16, pp. 25–26). Gell-Mann and Hartle consider equivalence relations in the absence of any hypothetical selection mechanism; their new proposals seem very natural, and greatly refine those we have described above. Apart from this paragraph and occasional footnotes, we have not amended the discussion of this section in the light of the new treatment, so as to conform with our general approach in this and the next section, in which we simplify arguments and calculations by neglecting the dynamical aspects of the formalism where possible.

There are further relations which one might consider. For example, it might seem sensible to assign no significance to the time-ordering of commuting sets of projection operators if no other projections intervene. Thus, if σ_k and σ_{k+1} are pairwise commuting sets of projections, and if we define σ'_k to be the set of nonzero projections among the operators

$$\{P_k^{(i_k)}P_{k+1}^{(i_{k+1})}: i_k = 1,..., n_k, i_{k+1} = 1,..., n_{k+1}\} \tag{2.13}$$

one could impose the following relation:

$$(\rho_i, \rho_f, \{\sigma_j\}) \sim (\rho_i, \rho_f, \{\sigma'_{j'}\}) \tag{2.14}$$

where $n = n' + 1$ and

$$\begin{aligned} \sigma_j &= \sigma'_j, \qquad j = 1,..., k-1 \\ \sigma_j &= \sigma'_{j-1}, \qquad j = k+2,..., n' \end{aligned} \tag{2.15}$$

However, since it is less obvious that one wants all the relations implied from this by symmetry and transitivity, we shall not impose this last relation.

We have no compelling principle to determine exactly which equivalences should be imposed, and quite possibly several others may eventually be adopted: the relation of trivial extension, discussed in the next section, is another strong candidate. However, we shall adopt only the particularly simple relations (2.11) and (2.12). If new equivalences are suggested, it may eventually be necessary to investigate whether they substantially affect our arguments below; however, it seems unlikely that they will. Even if all sets of histories with the same collection of nonzero probabilities were made equivalent—and this is surely an upper bound on what might conceivably be sensible—we do not believe that it would make any qualitative difference to our arguments in practical applications.

2.3. Other Conditions

It is useful to have a condition of maximality when characterizing consistent sets, since otherwise one has to consider all subsets of a consistent set on an equal footing. No doubt it is not essential, but we adopt it in line with our general strategy of accepting plausible rules which reduce the number of sets. The condition we shall define here differs from that given by Gell-Mann and Hartle[5] and is tailored to our purpose of characterizing the sets.

We first need some preliminary definitions. We say the time-ordered set

$$\mathscr{S}' = (\rho_i, \rho_f, \{\sigma_1,..., \sigma_k, \tau, \sigma_{k+1},..., \sigma_n\}) \tag{2.16}$$

is a *consistent extension* of a consistent set of histories $\mathscr{S} = (\rho_i, \rho_f, \{\sigma_1, ..., \sigma_n\})$ by the set of projections $\tau = \{Q^i: i = 1,..., m\}$ if τ satisfies (2.2) and $\mathscr{S}'$ is itself consistent.[9] We say the consistent extension $\mathscr{S}'$ is *trivial* if, for each history $\{P_1^{(i_1)},..., P_k^{(i_k)}, P_{k+1}^{(i_{k+1})},..., P_n^{(i_n)}\}$ in $\mathscr{S}$, at most one of the extended histories $\{P_1^{(i_1)},..., P_k^{(i_k)}, Q^i, P_{k+1}^{(i_{k+1})},..., P_n^{(i_n)}\}$ has nonzero probability. We call $\mathscr{S}' = (\rho_i, \rho_f, \{\sigma_1,..., \sigma_{k-1}, \sigma'_k, \sigma_{k+1},..., \sigma_n\})$ a *consistent refinement* of the consistent set $\mathscr{S}$ if $\mathscr{S}'$ is consistent and the projective decomposition

$$\sigma'_k = \{P_k'^{(1)},..., P_k'^{(n'_k)}\}$$

is a refinement of $\sigma_k = \{P_k^{(1)},..., P_k^{(n_k)}\}$, by which we mean that $n'_k > n_k$ and each $P_k^{(i)}$ can be written as the sum of one or more of the $P_k'^{(j)}$; we define *trivial* consistent refinement in the same way as trivial consistent extension. We use the term (*trivial*) *consistent fine graining* to mean either a (trivial) consistent refinement or a (trivial) consistent extension. We extend these definitions by taking (trivial) consistent refinement, (trivial) consistent extension, and (trivial) consistent fine graining to be transitive relations.[10] We say a consistent set $\mathscr{S}$ is *maximally extended* if it has no nontrivial consistent extension. Finally, we say a consistent set of histories $\mathscr{S}$ is *fully fine grained* if it is maximally extended, has no consistent refinement, and is not itself a trivial extension of any consistent set. We follow Gell-Mann and Hartle in defining a *maximally refined* consistent set to be one with no (trivial or nontrivial) consistent fine grainings.[11]

At first blush, it seems natural to define a further equivalence relation by setting $\mathscr{S} \sim \mathscr{S}'$ if $\mathscr{S}'$ is a trivial consistent extension or trivial consistent refinement of $\mathscr{S}$. We shall not do so because, while in some interpretations there is a case for considering $\mathscr{S}$ and $\mathscr{S}'$ as representatives of the same fundamental object, it is not clear to us that all the sets related (via transitivity) by a chain of such relations should also be taken as equivalent. The equivalence classes of fully fine grained consistent sets of histories—which for brevity we shall call the *fundamental consistent* sets—could reasonably be taken as the fundamental objects in any given cosmological theory, although one needs to flesh them out by trivial extensions when describing quasiclassical domains or the behavior of IGUSes. We turn now to the problem of characterizing the fundamental consistent sets.

[9] Though for simplicity of notation we use sets of finite length, this and later definitions should be understood as extending to infinite-length sets in the obvious way.

[10] Gell-Mann and Hartle use the term fine graining to cover both the notions of extension and refinement defined in the text, without any assumption of consistency.[5]

[11] Gell-Mann and Hartle also use the term *maximal* for maximally refined sets,[5] and define the related notion of a *full* consistent set.[7,9]

3. PROPERTIES OF CONSISTENT SETS

Since physics is described by (at least) the fundamental consistent sets of histories, we need to know how many of these sets there will be, and what properties they are likely to have, in typical cosmological theories. As no demonstrably good quantum cosmological theory exists, and as it is clear that the relevant calculations would be extremely complicated in any cosmological theory, this seems a hopeless task at present. However, we can find a few hints about the general features of consistent sets by looking at finite-dimensional Hilbert spaces. Of course, it is possible that these calculations are misleading: conceivably, there is some good cosmological theory in which the variety of consistent sets has quite different properties from those suggested by finite-dimensional intuition. However, as far as we are aware, no one has suggested that the true cosmological theory should have this surprising feature. It seems reasonable to assume that finite-dimensional intuition is correct, until a good counterargument is produced.

In this section, we generally take the Hilbert space $\mathscr{H}$ to be of finite dimension n, although we shall allow $\mathscr{H}$ to be infinite dimensional and separable if the same proof covers both cases. We generally omit the final density matrices ρ_f, which are taken to be 1 unless otherwise stated.

3.1. Classification

The basic objects in the formalism are the projective decompositions of the identity $\sigma_j = \{P_j^{(i)} : i = 1, 2, \ldots, n_j\}$, where the $P_j^{(i)}$ are orthogonal Hermitian projections, so that

$$\sum_i P_j^{(i)} = 1 \qquad \text{and} \qquad P_j^{(i)} P_j^{(i')} = \delta_{ii'} P_j^{(i)} \tag{3.1}$$

These decompositions are parametrized by: (i) the set of ranks

$$\{r_j^{(1)}, r_j^{(2)}, \ldots, r_j^{(n_j)}\} \tag{3.2}$$

of the projection operators, where $n = \sum_{i=1}^{n_j} r_j^{(i)}$ and we take $r_j^{(1)} \geqslant r_j^{(2)} \geqslant \ldots \geqslant r_j^{(n_j)}$, and (ii) (up to discrete symmetries) the generalized Grassmannian manifold

$$G(n; r_j^{(1)}, r_j^{(2)}, \ldots, r_j^{(n_j)}) = \frac{U(n)}{(U(r_j^{(1)}) \times U(r_j^{(2)}) \times \cdots \times U(r_j^{(n_j)}))} \tag{3.3}$$

More precisely, if the first k_1, the next k_2,..., and the last k_l of the $r_j^{(1)}$, ordered as above, are equal—where the $k_j \geqslant 1$, and so $\sum_{j=1}^{l} k_j = n_j$—then the parametrization manifold is actually

$$G(n; r_j^{(1)}, r_j^{(2)},..., r_j^{(n_j)})/(S_{k_1} \times \cdots \times S_{k_l}) \tag{3.4}$$

where S_k is the group of permutations of k elements.

It is easy to use this parametrization in explicit calculations: we can define projections $\{P^{(1)},..., P^{(n_j)}\}$ of ranks $\{r^{(1)}, r^{(2)},..., r^{(n_j)}\}$ by choosing an orthonormal basis of vectors $\{x_1,..., x_n\}$, so that

$$P^{(1)} = \sum_{i=1}^{r_1} x_i(x_i)^\dagger, \qquad P^{(2)} = \sum_{i=r_1+1}^{r_1+r_2} x_i(x_i)^\dagger \tag{3.5}$$

and so on. The redundancies in this parametrization correspond to the actions of $U(r^{(1)}) \times U(r^{(2)}) \times \cdots \times U(r^{(n_j)})$ and $S_{k_1} \cdots \times S_{k_l}$. They can be eliminated, and the equivalence relations (2.12) and (2.14) can be imposed, at any convenient stage of the calculation.

Thus, in principle, we can simply fix the form of the initial density matrix, fix the ranks of the projection operators in the type of consistent set we wish to classify, and then impose the consistency conditions (2.7). These define (real) algebraic curves in the generalized Grassmannians (3.3); their intersection is an algebraic variety whose generating polynomials can be obtained by the usual reduction methods. For example, one finds the following in two dimensions:

Example 1. If

$$\rho = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}, \qquad \text{with} \quad p \neq 0, 1/2, \text{or } 1$$

and we set

$$P(z) = \frac{1}{1+zz^*} \begin{pmatrix} 1 & z \\ z^* & zz^* \end{pmatrix} \tag{3.6}$$

then the only consistent sets of length two are of the form

$$P_1^{(1)} = P(0), \qquad P_1^{(2)} = P(\infty), \qquad P_2^{(1)} = P(z), \qquad P_2^{(2)} = P(-1/z^*) \tag{3.7}$$

where z is any nonzero complex number. These sets are equivalent [by (2.12)] to sets with z on the positive real axis, and so the fundamental consistent sets are of the form (3.7) with z positive real.

Example 2. If

$$\rho = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

then the only length-two consistent sets are of the form

$$\begin{aligned} P_1^{(1)} &= P(z_1), \qquad P_1^{(2)} = P(-1/z_1^*) \\ P_2^{(1)} &= P(z_2), \qquad P_2^{(2)} = P(-1/z_2^*) \end{aligned} \tag{3.8}$$

where z_1 and z_2 are any complex numbers. Imposing the equivalence relation (2.12), we again find that the fundamental consistent sets take the form (3.7), with z positive real.

More interestingly, in three dimensions we have the following.

Example 3. If

$$\rho = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and we set

$$P(z_1, z_2) = \frac{1}{1 + z_1 z_1^* + z_2 z_2^*} \begin{pmatrix} 1 & z_1 & z_2 \\ z_1^* & z_1^* z_1 & z_1^* z_2 \\ z_2^* & z_2^* z_1 & z_2^* z_2 \end{pmatrix} \tag{3.9}$$

then the equivalence classes of consistent sets of length two, involving binary projective decompositions, are represented by

$$\begin{aligned} P_1^{(1)} &= P(0, x_{12}), \qquad P_1^{(2)} = I - P_1^{(1)} \\ P_2^{(1)} &= P(x_{21}, x_{22} + i y_{22}), \qquad P_2^{(2)} = I - P_2^{(1)} \end{aligned} \tag{3.10}$$

where $x_{12}, x_{21}, x_{22}, y_{22}$ are real numbers such that

$$\begin{aligned} &\text{(i)} \qquad x_{12} y_{22} = 0 \\ &\text{(ii)} \qquad \text{either} \quad 1 + x_{12} x_{22} = 0 \quad \text{or} \quad x_{12}(x_{12} - x_{22}) = 0 \end{aligned} \tag{3.11}$$

That is, the solution space is a two-dimensional real algebraic variety in the parameter space.[12] This simple example illustrates the general

[12] We note for use in Example 4 that (i) and (ii) are necessary and sufficient conditions for the sets to be consistent, given that the parameters are real.

phenomenon. Already here, though, a calculation of the equivalence classes of fundamental consistent sets would be rather complicated, and for general sets in higher dimensions the algebraic equations are, practically speaking, intractable. Fortunately, we are interested in the qualitative features of the solution spaces in general, rather than the precise details of particular cases. What one expects, naively, is a solution space of dimension given by the number of parameters minus the number of independent consistency equations. Now $G(n; r_j^{(1)}, r_j^{(2)},..., r_j^{(n_j)})$ has real dimension $n^2 - \sum_{i=1}^{n_j} (r_j^{(i)})^2$, while a set $\sigma_1,..., \sigma_k$ of projective decompositions of length $n_1,..., n_k$ gives rise to no more than $n_1 \cdots n_{k-1} n_k(n_1 \cdots n_{k-1} - 1)$ real equations. The equivalence relation (2.12) accounts for not more than $(n^2 - 1)$ real parameters, depending on the form of ρ. This suggests a solution space of dimension no smaller than

$$\prod_{j=1}^{k} \left(n^2 - \sum_{i=1}^{n_j} (r_j^{(i)})^2\right) - n_k\left(\left(\prod_{j=1}^{k-1} n_j\right) - 1\right) \prod_{j=1}^{k-1} n_j - (n^2 - 1) \quad (3.12)$$

For example, if we restrict ourselves to considering projective decompositions of a $2N$-dimensional space into $N-$dimensional subspaces and consider sets of k such decompositions, the estimate (3.12) gives

$$(2N^2)^k - 2^k(2^{k-1} - 1) - 4N^2 \sim 2^k N^{2k} \quad (3.13)$$

parameters. This suggests that there are solutions for arbitrarily large k.

Whether these naive calculations are generally a good guide is open to doubt, but it is certainly possible to find inequivalent sets of solutions for arbitrarily large k. For example, consider a space of dimension $N \geqslant 4$ with an orthonormal basis $\{e_1,..., e_N\}$, let $\rho = |e_1\rangle\langle e_1|$, and let

$$\begin{aligned} P_i^{(1)} &= |e_1\rangle\langle e_1| + Q_1, & P_i^{(2)} &= 1 - P_i^{(1)} & &\text{for } i \text{ odd} \\ P_i^{(1)} &= |e_2\rangle\langle e_2| + Q_2, & P_i^{(2)} &= 1 - P_i^{(1)} & &\text{for } i \text{ even} \end{aligned} \quad (3.14)$$

where Q_1 and Q_2 are noncommuting projections on the subspace spanned by $e_3,..., e_N$. Then, writing $\sigma_i = \{P_i^{(1)}, P_i^{(2)}\}$, we have that the set $S_r = (\rho, \{\sigma_1, \sigma_2,..., \sigma_r\})$ is consistent for arbitrarily large r, and is not equivalent to any shorter consistent set. The same is true, even more trivially, of $\mathscr{S}'_r = (\rho, \{\sigma_1, \sigma_1,..., \sigma_1\})$, the σ_1 being repeated r times.

3.2. Lemmas and Examples

As the last examples illustrate, the full space of equivalence classes of consistent sets is not very interesting. The sets $\mathscr{S}_r$ and $\mathscr{S}'_r$ for $r \geqslant 2$ are trivial extensions of $\mathscr{S}_1$ and $\mathscr{S}'_1$, respectively. We are really interested in fully fine grained sets, and these have qualitatively different properties:

Lemma 1. Let $\mathscr{S}=(\rho, \{\sigma_1,..., \sigma_k\})$ be a consistent set that is not a trivial extension of any consistent set, defined on a space $\mathscr{H}$ of dimension n, with initial density matrix ρ of rank r. Then the length k of $\mathscr{S}$ obeys $k \leqslant rn$. [In particular, (i) this holds for any fully fine grained consistent set, and (ii) if ρ is pure, then $k \leqslant n$.]

Proof. Let $\rho=\sum_{i=1}^{r} p_i |\psi_i\rangle\langle\psi_i|$, where the p_i are positive, $\sum_{i=1}^{r} p_i=1$, and the $|\psi_i\rangle$ are the first r elements of an orthonormal basis $|\psi_1\rangle,..., |\psi_n\rangle$ of $\mathscr{H}$; let $\sigma_j=\{P_j^{(i)}: 1 \leqslant i \leqslant n_j\}$. We write $\rho^{1/2}=\sum_{i=1}^{r} p_i^{1/2}|\psi_i\rangle\langle\psi_i|$ and define the operators A_{ij} (for i, j running from 1 to n) on $\mathscr{H}$ by $A_{ij}|\psi_k\rangle=\delta_{jk}|\psi_i\rangle$.

The matrix ρ defines a Hermitian form $\langle A, B\rangle=\mathrm{Tr}(A^{\dagger}\rho B)$ on the operators on $\mathscr{H}$. The form's null space N is spanned by the operators A_{ij} for $r+1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant n$, and there is a natural isomorphism

$$O'=\mathrm{span}\{A_{ij}: 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant n\} \equiv \mathrm{Op}(\mathscr{H})/N \tag{3.15}$$

given by $A_{ij} \leftrightarrow A_{ij}+N$. We have a natural positive-definite Hermitian form, which we denote by $(\cdot, \cdot)$, on O', given by $(A, B)=\mathrm{Tr}(A'B)$. Since $\mathscr{S}$ is consistent,

$$\mathrm{tr}(P_k^{i_k} \cdots P_1^{i_1} \rho P_1^{i'_1} \cdots P_k^{i'_k})=p(i_1,..., i_k)\, \delta_{i_1 i'_1} \cdots \delta_{i_k i'_k} \tag{3.16}$$

so that the operators $\rho^{1/2}P_1^{i_1}...P_k^{i_k}$ are orthogonal with respect to $(\cdot, \cdot)$. Thus there can be no more than rn histories with nonzero probability.

Now by hypothesis each of the σ_j for $j=1, 2,..., k$ defines a nontrivial extension of the consistent subset $\hat{\mathscr{S}}_j=(\rho, \{\sigma_1,..., \hat{\sigma}_j,..., \sigma_k\})$ obtained by its deletion. *A fortiori*, each of the σ_j for $j \geqslant 2$ defines a nontrivial extension of the subset $\mathscr{S}_j=(\rho, \{\sigma_1,..., \sigma_{j-1}\})$, so that we can obtain $\mathscr{S}$ as a series of such nontrivial extensions, starting from the length-one set $(\rho, \{\sigma_1\})$. This means that for each j from 1 to $k-1$, there is some set of indices $i'_1,..., i'_j$ such that there are two projections $P_{j+1}^{i'_{j+1}}$ and $P_{j+1}^{i''_{j+1}}$ in σ_{j+1} with

$$p(i'_1,..., i'_j, i'_{j+1}) \neq 0 \neq p(i'_1,..., i'_j, i''_{j+1}) \tag{3.17}$$

Since each nontrivial extension thus increases the total number of histories with nonzero probability, the lemma follows.

Note also that, when ρ is pure, it follows that a set is maximally extended if and only if it has n histories of nonzero probability. Another simple result in the pure ρ case, which we shall use in Section 5, is as follows.

Lemma 2. Let $\mathcal{S}$ be a set of consistent histories that is not maximally extended, with a pure initial state ρ, and let $\mathcal{H}$ be either finite dimensional or separable. Then there exists a continuous family of nontrivial extensions for each history in $\mathcal{S}$ with nonzero probability.

Proof. Let $\rho = |\psi\rangle\langle\psi|$ and $\mathcal{S} = (\rho, \{\sigma_1,..., \sigma_l\})$. For simplicity of notation, take $\mathcal{H}$ to be finite dimensional, and let n be its dimension: the proof can trivially be rewritten for separable $\mathcal{H}$. Consider the subset of histories with nonzero probability, which we can label by products of projections S_j for $j = 1, 2,..., r$, where each S_j is an ordered product, taking one projection from each of the σ_i. We define the states $|h_j\rangle = S_j |\psi\rangle / \|S_j |\psi\rangle\|$. The consistency condition means that the $|h_j\rangle$ are orthogonal. Since $\mathcal{S}$ is not maximally extended, $r < n$: and we can extend the set of $|h_j\rangle$ to an orthonormal basis by adding states $\{|e_k\rangle : k = r+1,..., n\}$. Choosing states $|h_j\rangle$ and $|e_k\rangle$ and complex numbers z_1, z_2 with $z_1^* z_1 + z_2^* z_2 = 1$, we can define two orthogonal states

$$|\phi_+\rangle = z_1 |h_j\rangle + z_2^* |e_k\rangle, \qquad |\phi_-\rangle = z_2 |h_j\rangle - z_1^* |e_k\rangle \tag{3.18}$$

Define projections

$$P_\pm = |\phi_\pm\rangle\langle\phi_\pm|, \qquad P_3 = 1 - P_+ - P_- \tag{3.19}$$

which satisfy (2.2), and let $\tau = \{P_+, P_-, P_3\}$. Then $\mathcal{S}' = (\rho, \{\sigma_1,..., \sigma_l, \tau\})$ is a nontrivial consistent extension of $\mathcal{S}$. The histories in $\mathcal{S}'$ with non-zero probabilities correspond to the states $P_\pm |h_j\rangle$ and the $|h_i\rangle$ for $i \neq j$. Since there is a consistent extension for every choice of z_1 and z_2, the lemma holds.

Note also that these extensions are generically incompatible. The basic point here—that, with a pure density matrix, one can produce many consistent sets using simple vector space algebra—was first remarked on by Gell-Mann and Hartle (ref. 5, p. 441). One immediate consequence is worth noting:

Lemma 3. If the Hilbert space is infinite dimensional and the initial state is pure, then there exist consistent sets $\mathcal{S}$ which are of infinite length and are nontrivial extensions of all their subsets.

Proof. We can explicitly build such sets by repeatedly applying the construction of the last proof in the obvious way.

The above results rely on the assumption that the final density matrix is fixed to be I. Taking ρ_f as a variable changes the picture:

Lemma 4. Let $\mathcal{S} = (|\psi\rangle\langle\psi|, \{\sigma_1,..., \sigma_l\})$ be as in Lemma 2, with $\mathcal{H}$ finite dimensional, and let $\mathcal{S}' = (|\psi\rangle\langle\psi|, \{\sigma_1,..., \sigma_l, \sigma_{l+1}\})$ be a nontrivial consistent extension. Then there exists some final density matrix ρ_f such that:

(i) $\mathrm{Tr}(|\psi\rangle\langle\psi|\rho_f)=1$ and $\mathscr{S}_f=(|\psi\rangle\langle\psi|, \rho_f, \{\sigma_1,..., \sigma_l\})$ is consistent.

(ii) The histories of $\mathscr{S}_f$ have the same probabilities as those of $\mathscr{S}$.

(iii) The extension $\mathscr{S}'_f=(|\psi\rangle\langle\psi|, \rho_f\{\sigma_1,..., \sigma_l, \sigma_{l+1}\})$ is inconsistent.

Moreover, a generic ρ_f satisfying (i) and (ii) also satisfies (iii).

Proof. It is enough to prove the result in the case where $\sigma_{l+1}=\{P, 1-P\}$. Define r, n, the nonzero history states $|h_i\rangle$, and the orthogonal complement basis states $|e_i\rangle$ as in the proof of Lemma 2. Since $\mathscr{S}'$ is a nontrivial extension of $\mathscr{S}$, we can assume without loss of generality that $P|h_1\rangle$ does not lie in the subspace spanned by the $|h_j\rangle$. By the consistency of $\mathscr{S}'$, we have

$$P|h_1\rangle = a_1|h_1\rangle + \sum_{i=r+1}^{n} a_i|e_i\rangle \tag{3.20}$$

with $a_1\neq 0$ and not all the $a_i=0$, $i=r+1,..., n$. Now a density matrix $\rho_f=(\rho_f)_{ij}$ satisfies conditions (i) and (ii) provided that $(\rho_f)_{ij}=\delta_{ij}$ for $1\leqslant i, j\leqslant r$. If the set $\mathscr{S}'_f$ is consistent, then in particular

$$\mathrm{Tr}(\rho_f P|h_1\rangle\langle h_1|(1-P))=0 \tag{3.21}$$

which implies that

$$\begin{aligned} a_1(1-a_1^*) + \sum_{i=r+1}^{n} ((1-a_1^*)\,a_i(\rho_f)_{1i} - a_1 a_i^*(\rho_f)_{i1}) \\ + \sum_{i,j=r+1}^{n} a_i a_j^*(\rho_f)_{ij} = 0 \end{aligned} \tag{3.22}$$

This is not true for generic Hermitian positive-semidefinite ρ_f unless all the a_i (for $i=r+1,..., n$) are zero, which contradicts our earlier assumption.

Returning to the case where ρ_f is trivial, we can sharpen the non-uniqueness statement of Lemma 2:

Lemma 5. Let $\mathscr{S}=(|\psi\rangle\langle\psi|, \{\sigma_1,..., \sigma_l\})$ be as in Lemma 2, with $\mathscr{H}$ finite dimensional or separable. Then there is no projective decomposition σ_{l+1} such that: (i) $\mathscr{S}'=(\rho, \{\sigma_1,..., \sigma_l, \sigma_{l+1}\})$ is a consistent extension of $\mathscr{S}$, and (ii) any consistent extension $\mathscr{S}''=(\rho, \{\sigma_1,..., \sigma_l, \tau\})$ of $\mathscr{S}$ has a consistent extension $(\rho, \{\sigma_1,..., \sigma_l, \tau, \sigma_{l+1}\})$.

Proof. Suppose otherwise. We first show that σ_{l+1} must be a trivial extension of $\mathscr{S}$.

Suppose σ_{l+1} is a binary projective decomposition, $\sigma_{l+1} = \{P, 1-P\}$. As in Lemma 2, we can define the orthonormal set of history states $\{|h_j\rangle : j = 1, 2,..., r\}$ belonging to $\mathcal{S}$, and extend it to an orthonormal basis with the states $\{|e_k\rangle : k = r+1,..., n\}$, where if $\mathcal{H}$ is separable, we set $n = \infty$. We know that $r < n$, since $\mathcal{S}$ is not maximally extended; we shall assume for the moment that $r > 1$. We pick states $|h_j\rangle$ and $|e_k\rangle$ and complex numbers z_1, z_2 as above, and define $|\phi_\pm\rangle$ and the projections $P_\pm, P_3$ by Eq. (3.18) and (3.19). Since σ_{l+1} itself defines a consistent extension, we have that

$$\langle h_a| P |h_b\rangle = 0 \qquad \text{for} \quad 1 \leqslant a < b \leqslant r \tag{3.23}$$

Condition (ii) implies that

$$\langle h_a| P_\pm P P_3 |h_b\rangle = 0 \qquad \text{for} \quad 1 \leqslant a, b \leqslant r \tag{3.24}$$

But this is true for any choice of $|h_j\rangle$ and $|e_k\rangle$, which means that, by considering all combinations such that $b \neq j$ and $a = j$, we obtain

$$\langle e_k| P |h_b\rangle = 0 \qquad \text{for} \quad 1 \leqslant b \leqslant r \quad \text{and} \quad r+1 \leqslant k \leqslant n \tag{3.25}$$

Equation (3.25) means that

$$\text{span}(\{P |h_a\rangle\}) \subseteq \text{span}(\{|h_a\rangle\}) \tag{3.26}$$

and the same is true for $(1-P)$. Hence $\{P, 1-P\}$ defines a trivial extension of $\mathcal{S}$. If $r = 1$, we can use

$$\langle h_1| P_+ P P_- |h_1\rangle = 0 \tag{3.27}$$

to show that

$$\langle e_k| P |h_1\rangle = 0 \qquad \text{for} \quad 2 \leqslant k \leqslant n \tag{3.28}$$

and again $\{P, 1-P\}$ must be a trivial extension. Finally, if σ_{l+1} is a general projective decomposition $\{Q_1, Q_2,..., Q_q\}$, then we can apply the above arguments to the decompositions $\{Q_p, 1-Q_p\}$ (for $p = 1, 2,..., q$), which have the same consistency properties as $\{P, 1-P\}$. Hence (3.26) holds for each of the Q_p, and once again σ_{l+1} trivially extends $\mathcal{S}$.

So, in all cases, we have that σ_{l+1} is a trivial extension, and without loss of generality can be taken to be binary. We now show that this leads to a contradiction. We have that $\sigma_{l+1} = \{P, 1-P\}$ is a trivial extension of $\mathcal{S}$ and can be made to any consistent extension of $\mathcal{S}$. Choose the states $\{|e_{r+1}\rangle,..., |e_n\rangle\}$ to be eigenstates of P and $(1-P)$. Choose $a \in \{1, 2,..., r\}$ and $k \in \{r+1,..., n\}$ such that $P|h_a\rangle = |h_a\rangle$ and $(1-P)|e_k\rangle = |e_k\rangle$; we may do this without loss of generality, interchanging P and $(1-P)$ if necessary. Construct $|e_+\rangle = (1/\sqrt{2})(|h_a\rangle + |e_k\rangle)$ and let $Q = |e_+\rangle\langle e_+|$. Then $\{Q, 1-Q\}$ consistently extends $\mathscr{S}$, but $(|\psi\rangle\langle\psi|, \{\sigma_1,..., \sigma_l, \{Q, 1-Q\}, \{P, 1-P\}\})$ is inconsistent. This completes the proof.

We can also use this argument to find a useful property of trivial extensions:

Lemma 6. Let $\mathscr{S}' = (\rho, \{s_1,..., s_l, t\})$ be a trivial consistent extension of the consistent set $\mathscr{S} = (\rho, \{s_1,..., s_l\})$ in a finite-dimensional space $\mathscr{H}$. Let $s_j = \{P_j^{(i)}: i = 1, 2,..., n_j\}$, and let $t = \{Q^i: i = 1, 2,..., n\}$. Then for any $i_1,..., i_l$ and i in the given ranges, we have

$$Q^i P_l^{i_l} \cdots P_1^{i_1} \rho^{1/2} = \begin{cases} P_l^{i_l} \cdots P_1^{i_1} \rho^{1/2} & \text{if } \ p(i_1,..., i_l, i) \neq 0 \\ 0 & \text{if } \ p(i_1,..., i_l, i) = 0 \end{cases} \tag{3.29}$$

Proof. The case when $p(i_1,..., i_l, i) = 0$ is immediate. But by triviality, for each $\{i_1,..., i_1\}$ there is only one i with $p(i_1,..., i_l, i) \neq 0$, and since $Q^i = 1 - \sum_{j \neq i} Q^j$ the result follows.

This implies that, if a projective decomposition occurs more than once in a consistent set, there is at least one class of consistent extensions which *can* be made to all consistent extensions:

Lemma 7. Let

$$\mathscr{S} = (\rho, \{s_1,..., s_j, t, t_1,..., t_l, t, s_{j+1},..., s_k\}) \equiv (\rho, \{S_1, t, T, t, S_2\})$$

be a consistent set in which the projective decomposition t is repeated. Let $\mathscr{S}' = (\rho, \{S_1, t, T_1, t, T_2, t, S_2\})$ be an extension of $\mathscr{S}$ by a further repetition of t at some point between the first two, so that $\{T\} = \{T_1, T_2\}$. Then $\mathscr{S}'$ is also consistent.

Proof. Let α, β, γ denote histories from S_1, T, S_2 respectively, and a, b projections from t. Since

$$\sum_\beta p(\alpha, a, \beta, b, \gamma) = p(\alpha, a, b, \gamma) = \delta_{ab}\, p(\alpha, a, \gamma)$$

we have that $p(\alpha, a, \beta, b, \gamma) = \delta_{ab} p(\alpha, a, \beta, \gamma)$, so that the consistency of $\mathscr{S}$ implies that it is a trivial extension of $(\rho, \{S_1, t, T, S_2\})$. Likewise the set $(\rho, \{\bar{S}_1, t, \bar{T}, t, \bar{S}_2\})$ is a trivial consistent extension of $(\rho, \{\bar{S}_1, \bar{T}, t, \bar{S}_2\})$, for any subsets $\bar{S}_1, \bar{S}_2, \bar{T}$ of S_1, S_2 and T. In particular, $(\rho, \{S_1, t, T_1, t\})$ is a trivial consistent extension of $(\rho, \{S_1, t, T\})$ for any subset T_1 of T. Let

P^1_α and Q^1_β denote general histories from S_1 and T_1, respectively, and let P_a denote a general projection from t. Lemma 6 implies that

$$P_a Q^1_\beta P_b P^1_\alpha \rho^{1/2} = \delta_{ab} Q^1_\beta P_b P^1_\alpha \rho^{1/2} \tag{3.30}$$

Hence the consistency conditions for $\mathscr{S}'$ follow from those for $\mathscr{S}$, as required. Note that, by the same argument, any consistent fine graining of $\mathscr{S}$ can be consistently fine grained by a further repetition of t between the first two.

We conclude with a short list of counterexamples. By Lemma 7, a consistent set in which a projective decomposition t is repeated can always be extended consistently by adding an arbitrary number of repetitions of t between the originals. However, Lemma 5 shows that adding t after all existing occurrences of t is not possible in all consistent extensions. We cannot even, in general, add t to the past of all the original t's:

Example 4. As in Example 3, we take a three-dimensional Hilbert space, with pure initial state ρ, and consider the one-dimensional projections $P(z_1, z_2)$ defined by (3.9). Defining histories by the series of projective decompositions $\sigma_i = \{P^{(1)}_i, P^{(2)}_i\}$ for $i = 1, 2, 3$, where

$$\begin{aligned} P^{(1)}_1 &= P(0, x_{12}), \qquad & P^{(2)}_1 &= I - P^{(1)}_1 \\ P^{(1)}_2 &= P(0, 0), \qquad & P^{(2)}_2 &= I - P^{(1)}_2 \\ P^{(1)}_3 &= P(0, x_{12}), \qquad & P^{(2)}_3 &= I - P^{(1)}_3 \end{aligned} \tag{3.31}$$

Again we see from (3.11) that, while the set $(\rho, \{\sigma_2, \sigma_3\})$ is consistent, its extension by past repetition, $(\rho, \{\sigma_1, \sigma_2, \sigma_3\})$, is not consistent if $x_{12} \neq 0$.

Another interesting question is the extent to which the usual notion of correlated subsystems carries over into the consistent histories formalism. The answer is that, while one can find a consistent set which describes the correlation, one can also find consistent sets which are incompatible with the correlation, in the sense that they have no consistent extension in which the correlation can be displayed:

Example 5. Consider the Hilbert space $\mathscr{H} = \mathscr{H}_1 \otimes \mathscr{H}_2$, where $\mathscr{H}_1$ and $\mathscr{H}_2$ are two-dimensional spaces with orthonormal bases $\{v_1, v_2\}$ and $\{w_1, w_2\}$, respectively. Take

$$\rho = p\,|v_1\rangle\,|w_1\rangle\langle v_1|\,\langle w_1| + (1-p)\,|v_2\rangle\,|w_2\rangle\langle v_2|\,\langle w_2| \tag{3.32}$$

define the projections $P_i = |v_i\rangle\langle v_i|$, $Q_i = |w_i\rangle\langle w_i|$, $i = 1, 2$, and $Q_3 = \frac{1}{2}(|w_1\rangle + |w_2\rangle)(\langle w_1| + \langle w_2|)$. Let

$$\begin{aligned} \sigma_1 &= \{P_1 \otimes Q_1, I \otimes I - P_1 \otimes Q_1\} \\ \sigma_2 &= \{P_1 \otimes I, I \otimes I - P_1 \otimes I\} \\ \sigma_3 &= \{P_1 \otimes Q_3, I \otimes I - P_1 \otimes Q_3\} \end{aligned} \tag{3.33}$$

Then the set

$$\mathcal{S}_1 = (\rho, \{\sigma_1, \sigma_1, ..., \sigma_1\}) \tag{3.34}$$

is consistent and describes the correlation of the subsystems corresponding to $\mathcal{H}_1$ and $\mathcal{H}_2$, in the sense that the reduced density matrices corresponding to the histories of nonzero probability are

$$\begin{aligned} p^{-1} R_1 \cdots R_1 \rho R_1 \cdots R_1 &= |v_1\rangle\ |w_1\rangle\langle v_1|\ \langle w_1| \\ (1-p)^{-1}\ R_2 \cdots R_2 \rho R_2 \cdots R_2 &= |v_2\rangle\ |w_2\rangle\langle v_2|\ \langle w_2| \end{aligned} \tag{3.35}$$

where $R_1 = P_1 \otimes Q_1$ and $R_2 = (I \otimes I - P_1 \otimes Q_1)$. However, the set

$$\mathcal{S}_2 = (\rho, \{\sigma_3, \sigma_2, \sigma_2, ..., \sigma_2\}) \tag{3.36}$$

is also consistent. $\mathcal{S}_2$ does not describe the correlations, and cannot be consistently extended to a set which does: any set of the form

$$\mathcal{S}_3 = (\rho, \{\sigma_3, \sigma_2, ..., \sigma_2, \sigma_1, \sigma_2, ..., \sigma_2\}) \tag{3.37}$$

is inconsistent.

This simple result suggests, in particular, that one cannot generally deduce the quasiclassical behavior of one subsystem from a quasiclassical description of a disjoint subsystem, since if the descriptions are for a sufficiently short time, then the effect of the interaction Hamiltonian can be neglected. Put picturesquely, it is possible, given suitable initial conditions, and restricting oneself to a sufficiently short time interval, to find consistent sets in which the earth is described quasiclassically, and the moon is not, and in which no consistent fine graining allows one to recover a quasiclassical description of the moon. We shall suggest later that a similar problem arises for long time intervals.

Suppose we accept the probability rules (2.4) for consistent histories. We can still ask whether there might perhaps be some other, presently unknown rule which assigns probabilities to general inconsistent histories in such a way that the probability sum rules hold. As Goldstein and Page[29] have stressed, standard no-local-hidden-variables theorems[30] show that this cannot be done. For example:

Lemma 8. There exist inconsistent sets of projective decompositions with the property that there is no probability distribution on their histories from which one can derive the standard probabilities (2.4) for all histories belonging to consistent subsets.

Proof. With the Hilbert space and density matrix of Example 2, and using the notation defined there, consider the inconsistent set $\mathscr{S} = (\rho\{\sigma_1, \sigma_2, \sigma_3\})$, where $\sigma_i = \{P_i^{(1)}, P_i^{(2)}\}$, with $P_i^{(1)} = P(i)$ and $P_i^{(2)} = P(-1/i)$ for $i = 1, 2, 3$. Let a, b, c—each taking the values 1 or 2—label the projections in the various decompositions. Each length-two subset of $\mathscr{S}$ is consistent, so that any assignment of probability weights $p(a, b, c)$ to the histories $P_1^{(a)} P_2^{(b)} P_3^{(c)}$ of $\mathscr{S}$ would have to satisfy

$$p(1, 1, 1) + p(1, 1, 2) = \mathrm{Tr}(P_2^{(1)} P_1^{(1)} \rho P_1^{(1)}) \tag{3.38}$$

and 11 similar equations. From the general solution, which involves one free parameter, we find in particular that $p(1, 2, 1) + p(2, 1, 2) = -1/25$, so that at least one of the hypothetical probability weights must be negative.

As we shall explain in the next two sections, the various results we have listed are useful in clarifying the basic features of the consistent histories formalism and its interpretations. They also illustrate that very little is known of the properties of consistent sets even in finite-dimensional spaces, and there are many simple unanswered questions. How far, for instance, can the bound of Lemma 1 be improved? What is the length of a typical consistent set? Or the expected length of a consistent set constructed by choosing the first projective decomposition, and each successive extension, randomly (using some natural measure on the generalized Grassmannian spaces) from among the consistent alternatives? Can Lemmas 2–5 be extended to arbitrary initial density matrices? Does a consistent set $\mathscr{S} = (\rho, \{\sigma_1, ..., \sigma_l\})$, which is not maximally extended, always have an extension in the future, $\mathscr{S}' = (\rho, \{\sigma_1, ..., \sigma_l, \sigma_{l+1}, ..., \sigma_L\})$, which is maximally extended, or is it possible to find examples where nontrivial extension requires projective decompositions to be inserted within the existing sequence?

We expect that in practical applications, where the space of possible extensions tends to be of much larger dimension than the number of consistency equations, the naive counting arguments are qualitatively correct. That is, we expect that under these conditions the solution space of nontrivial consistent extensions generically contains manifolds of large dimension. Likewise, in practical applications the sets under consideration tend to be very far from fully fine grained, and we expect the conclusion of Lemma 2 to hold generically, whether or not the initial density matrix is pure.

In this section we have applied the condition that the objects of interest in the theory are the fundamental consistent sets. From now on we shall relax the equivalence relations and the condition of full fine graining. The description of IGUSes and quasiclassical domains which we shall need are most straightforwardly made assuming both that the projection operators are assigned times (contrary to our adopted equivalence relation) and that we can make trivial consistent fine grainings. The latter are needed since one of the most important sorts of prediction we want to make is of classical deterministic behavior which is most familiarly described in terms of repeated trivial extensions.

4. QUASICLASSICALITY AND APPROXIMATE CONSISTENCY

It is not obvious from our discussion so far that the consistent histories formalism is able to give any useful description of the world we see, or the laboratory experiments we do. The case that it can relies on a study of what Gell-Mann and Hartle have called the *quasiclassical* projection operators describing our own classical domain: operators such as the integral of the mass density, or the density of chemical species, over small regions, the approximate position of an apparatus pointer, or the approximate current through a photomultiplier. As many authors have stressed,[31–34, 9] interactions with the environment are an impressively efficient mechanism for quantum decoherence, even when that environment consists only of the cosmic microwave background radiation. Hence quasiclassical projection operators obey the consistency equations (2.7) to an extremely good approximation. However, since no natural decoherence mechanism is perfectly efficient, none of the quasiclassical projection operators listed above is ever likely precisely to obey the consistency equations; nor is it easy (even in simple models) to find quasiclassical projections which do. In the context of the consistent histories formalism, this point was first raised by Griffiths,[1] and has been discussed in more detail by Gell-Mann and Hartle.[5, 16] Gell-Mann and Hartle suggest that, in applying the formalism, approximately consistent sets should be accepted for the purpose of making predictions, so long as the failure of consistency (and hence of the probability sum rules) is so small as to be experimentally undetectable.

This proposal has excited lively controversy. Clearly, the use of approximately consistent sets does raise new problems. Most obviously, the problem of set selection is worsened, since the class of approximately consistent sets is larger than the class of consistent sets. More fundamentally, many physicists would prefer a theory in which probabilities are precisely defined and precisely obey the sum rules. The consistent histories formalism does, after all, have a natural mathematical structure: why weaken it with ad hoc prescriptions?[13] Moreover, the formalism has at least one natural interpretation in which it is impossible to ascribe a fundamental role to the approximately consistent sets.[14] Finally, it is often thought a sign of weakness if one needs apparently contingent facts to make sense of a fundamental theory: the overwhelming efficiency of environmental decoherence and the impossibility of performing an infinite sequence of identical experiments are both contingent on the Hamiltonian and boundary conditions, whereas the consistent histories formalism itself applies to any Hamiltonian and boundary conditions.

There are counterarguments for each of these points, and Gell-Mann and Hartle certainly do not regard any of these problems as serious weaknesses. Their defense rests on a particular view of the scientific enterprise and the role of probabilistic theories: expositions can be found in refs. 5 and 16. We do not, however, want to enter this debate: rather, we suggest that it is most likely irrelevant. A typical analysis of decohering quasiclassical projection operators applies a relatively short series of relatively coarse grained projections to a system whose Hilbert space is large or infinite dimensional. For example, an analysis of a photon two-slit experiment might use a few fairly coarse grained projective decompositions of the local densities of chemicals in a photographic plate at times either side of the arrival of the photon at the plate, while the relevant Hilbert space would describe a mesoscopic collection of particles. Unless an implausible number of partial derivatives vanish, this suggests that if the analyzed set $\mathscr{S}$ satisfies Equations (2.7) approximately, then there is a large-dimensional manifold of exactly consistent sets including points near $\mathscr{S}$, since the generalized Grassmannian spaces characterizing sets of projective decompositions with ranks those of $\mathscr{S}$ will be of dimension much larger than the number of consistency equations.

More precisely, for sets of the same length whose projection operators have fixed ranks, one can define a metric by, say, taking

$$d(\mathscr{S}, \mathscr{S}') = \sum_{j=1}^{l} \sum_{i=1}^{n_j} d(P_i^{(j)}, P_i'^{(j)}) \tag{4.1}$$

[13] Would similar violations of Lorentz invariance be acceptable?

[14] This is the many-histories interpretation described in Section 5.

where $d(P, P')$ is the operator norm of $(P - P')$. Then if $\mathscr{S}$ is consistent to within a small parameter ε, one expects to find an exactly consistent set $\mathscr{S}'$ (of the same length and with the same collection of ranks) within a distance $A\varepsilon$ of $\mathscr{S}$, where A depends on partial derivatives of the decoherence functionals at $\mathscr{S}$ and can in principle be estimated.

It is crucial for this argument that the analyzed set $\mathscr{S}$ is far from being fully fine grained. Given a pure initial density matrix in a space of dimension n, one can construct examples in which an approximately consistent set with no approximately consistent nontrivial fine graining contains more than n nontrivial histories, and so cannot possibly be approximated by an exactly consistent set.[15] Thus, roughly, the conjecture is that for any finite-dimensional Hilbert space $\mathscr{H}$ there is some positive bound $\varepsilon(M)$, defined for finite integer M, such that a generic set in the space of sets of histories, having M histories and approximately consistent to order $\varepsilon < \varepsilon(M)$, contains an exactly consistent set in its neighborhood; that $\varepsilon(M)$ is large enough that the approximately consistent sets of physical interest are well approximated by exactly consistent sets; and that this last fact holds true for histories defined on infinite-dimensional separable Hilbert spaces.

More precise mathematical statements (which neighborhood? what bounds can be found for A?) should probably be formulated. (There is some progress in this direction.[35]) It would also be very interesting to see some detailed investigations in models. Still, for the moment we see no strong case for arguing that approximately consistent sets need be given any fundamental role, and we consider only exactly consistent sets hereafter.

It is worth mentioning here another of Gell-Mann and Hartle's suggestions: the possibility that a single quasiclassical domain, or a small number of such domains, might emerge from quantum cosmology. The point here is that at present the notion of a quasiclassical domain is only an intuitive one:

> Roughly speaking, a quasiclassical domain should be a set of alternative decohering histories, maximally refined consistent with decoherence, with its individual histories exhibiting as much as possible patterns of classical correlation in time. Such histories cannot be *exactly* correlated in time according to classical laws because sometimes their classical evolution is disturbed by quantum events. There are no classical domains, only quasiclassical ones. (Ref. 5, p. 445.)

Gell-Mann and Hartle here consider maximally refined—and not fully fine grained—sets since a quasiclassical domain is supposed to involve

[15] We thank Matthew Donald, to whom this point is due.

many trivial extensions, so that the redundancy needed to characterize classical variables is built into its description.

It would be preferable to have a precise notion of quasiclassicality—a calculable quantity, defined by some sensible, but as yet unknown, measure. The hope is that the quasiclassical domain or domains would be given by one or more maximally refined consistent sets with a much higher degree of quasiclassicality than all the others:

> We wish to make the question of the existence of one or more quasiclassical domains into a *calculable* question in quantum cosmology and for this we need criteria to measure how close a set of histories comes to constituting a "classical domain". (Ref. 5, p. 445.) It would be a striking and deeply important fact of the universe if, among its maximal sets of decohering histories, there were one roughly equivalent group with much higher classicities than all the others. That would then be *the* quasiclassical domain, completely independent of any subjective criterion, and realized within quantum mechanics by utilizing only the initial condition of the universe and the hamiltonian of the elementary particles. (Ref. 5, p. 454.)

This last proposal can be fleshed out a little, under the assumption—which finite-dimensional calculations support—that the fully fine grained (and hence the maximally refined) consistent sets are parametrized by algebraic curves. One would expect any measure of quasiclassicality to be continuous, so that the most that could be hoped for is that there is a consistent set $\mathscr{S}$ on which the measure attains a global maximum, with the consistent sets in its neighborhood arbitrarily close to this maximum quasiclassicality, and which is much higher than any other local maxima. Gell-Mann and Hartle's suggestion is then that all those sets in the neighborhood of $\mathscr{S}$ whose quasiclassicality is near that of the maximum should describe very similar domains, so that to make predictions significantly different from those of $\mathscr{S}$ one would have to use sets of much lower quasiclassicality. The notions of "similar" and "near" need to be made precise here. Very roughly speaking, the hope seems to be that the relevant matrix of second derivatives is positive definite and that its eigenvalues have some lower bound. Finite-dimensional calculations give no clue as to the plausibility of this proposal, but their complexity does suggest that, even if the relevant quantities can be defined, it may be rather hard to test.

5. INTERPRETATIONS OF THE FORMALISM

Having set up the formalism and shown that there will be very many consistent sets of histories in even the simplest models, we must now consider how to make contact between the theory and our observations. In the consistent histories formalism, there is no distinction between the consistent sets in a given model of the universe. True, most of the sets will describe the history of the universe in terms of very complicated and nonlocal projection operators and it may happen that one or more of the sets contains histories which are described by projections that are much more familiar to us, but the formalism ascribes the same validity to them all. The problem appears only to be worsened when the formalism is applied to relativistic quantum field theory. How, then, are we to make sense of the formalism? Or, to put things more practically, if we are given a quantum cosmological theory, how (and to what extent) can we check whether it is right?

5.1. Cosmological Tests

There are two essentially distinct tests of a quantum cosmological theory: does it describe the past? and can it predict the future? Neither question is as simple as it seems. We first consider the problem of accounting for the past.

It seems obvious that any good theory must be consistent with what Omnès calls the *actual facts*—the events we regard as having actually been observed (in short, the data). That is, the projection operators corresponding to our actual experiences must comprise a history belonging to at least one consistent set. Let us call this the *actual history*. Further, the conditional probabilities assigned by the formalism must be borne out by what we have seen. For example, our two-slit experiments should have produced roughly the right interference patterns.

The difficulty here is that the notion of actual history is ambiguous.[16] This raises unavoidable questions about the proper scope of a cosmological theory. One could, solipsistically but quite consistently, include only one's own experiences in the actual history, or one could try to include those reported by others. One could try to describe only the present state of one's mind, or one could assume that one's memories are essentially accurate and try to construct a history corresponding to the past events which one recollects. More generally, one can take what appear to be historical records—Darwin's notebooks, the fossil record, the microwave background—merely as present facts to be described by present-day projection operators, or as prima facie evidence for past facts which should be included in the actual history. These are live issues in quantum cosmology. Of course, no one seriously supposes that we shall be faced with rival cosmological theories, one of which claims that Darwin's notebooks were written in the ordinary way, and the other that our present day records of Darwin's observations originate from a fortuitous quantum fluctuation in

[16] We thank Andy Albrecht for discussions of this point.

the neighborhood of the Cocos Islands on 1 April 1836. But we cannot interpret the formalism sensibly, or consider how it might best be developed, without taking a definite view on these points. Indeed, the entire formalism is unnecessary if one is content only to account for the "marvellous moment" and to treat all memories and records simply as statements about the present. [17] This "solipsism of the moment," which is argued against by Hartle (ref. 16, Section V. 1.2), is the most extreme antihistorical position. If one takes this line, everything can be described by a single projective decomposition now. Theories of the initial density matrix or the Hamiltonian are then simply data compression algorithms: one does physics (or rather, since processes extended in time have no meaning, has the illusion of doing physics) in an attempt to describe more succinctly one's momentary experience.

Our own view is that quantum cosmological theories ought to allow ordinary quasiclassical historical reasoning during the era in which, according to conventional cosmology, the universe was quasiclassical. Hartle makes the point well.

> The simple explanation for the observed correlation that similar dinosaur bones are located in similar strata is that dinosaurs *did* roam the earth many millions of years ago and their bones are persistent records of this epoch. It is by calculating probabilities between the past and different records today that correlations are predicted between these records. It is by such calculations that the probability for error in present records is estimated. Such calculations cannot be carried out solely on one spacelike surface. (Ref. 16, p. 77)

Of course, memory can deceive and records can mislead, and for these reasons alone there may never be complete agreement on what the actual facts are. Moreover, as we shall discuss presently, there are real theoretical dilemmas in selecting the actual facts. There are also differences of opinion over how, or perhaps even whether, criteria of beauty and simplicity should be applied to select cosmological theories.

Let us put these difficulties aside for the moment, assuming that the actual facts are given and restricting attention to theories which are agreed to be suitably simple and beautiful. We believe that a successful cosmological theory ought to describe the vast bulk of our own and others' experiences, and the past events reconstructable from records available to us, as events in a history from a single consistent set. Although we hope that most would concur with this, we are not sure there is a consensus. The formalism certainly allows a wide range of possible views. At one extreme, we

[17] Bell seems to have been the first to discuss this proposal in the context of many-worlds interpretations.[15] As he mentions, a well-known variant on the "marvellous moment" philosophy accepts records of recent history, but forbids historical reasoning beyond a certain point in the past.[36]

have pure solipsism of the moment: advocates of this view will accept any theory that describes their present experience. Next we have recent-history solipsism: this accepts any theory which describes actual facts over some period in time up to and including the present. Some recent-history solipsists might simply want to describe their own experiences, and see no need to include events occurring before their birth; others might choose some other interval. A third position is that any theory must allow us to describe historical events at arbitrarily early times, but not necessarily in the same consistent set. It is hard to produce convincing examples, but, suspending disbelief for the sake of illustration, one might imagine a quantum cosmological theory in which it was possible to describe early time matter density anisotropies leading to galaxy formation but *not* possible to describe all the relevant anisotropies in the same consistent set. Finally, one can demand that the theory describe, within one history of a single consistent set, essentially all the actual facts implied by quasiclassical historical reasoning.

We subscribe to this last position. The views of the developers of the consistent histories approach will be examined in later subsections. Meanwhile we should note that the extracts we have cited from Hartle's papers have to be interpreted with great care. Read naively, as statements in ordinary language, they set out a still stronger position. They demand that a cosmological theory should give a clear account of the past which unambiguously states that, with probability very close to one, key historical events actually took place. As it happens, this is indeed our view, but it is not Hartle's: as we shall explain, the cited extracts are *not* meant in their ordinary sense.

Now, if we adopt any position other than solipsism of the moment, we must explain what criteria we would use to identify actual facts. It is a complicated question. The assumption of quasiclassical behavior some way into the past seems to be part of the answer, but only part. It is closer to the truth to say that we would first build and test incomplete, high-level cosmological theories—describing the evolution of terrestrial life, the formation of the solar system, or the seeding of anisotropies from which galaxies develop—and then use successful theories of this type as actual facts to be explained by a fundamental quantum cosmology.[18] This

[18] Although we think it unlikely that there could be serious conflict between the two levels of theory, one could perhaps imagine some tension at the boundaries of quasiclassicality. Thus, while it seems inconceivable that there could be any quantum cosmological theory, consistent with the present data, in which no consistent set contains a quasiclassical description of terrestrial evolution, it does seem conceivable (if perhaps only as a result of our present ignorance) that the desire to maintain some elegant cosmological theory of structure formation might conflict with the desire to adopt the simplest possible theory of the boundary conditions. It is hard to see how any general algorithm could be found to decide which theory should be rejected.

procedure is very interesting. It would be good to see a serious examination of exactly what is involved and how it can be justified since, if quantum cosmology ever produces testable theories, we shall need to understand precisely which tests we are applying and why.

Discussions of actual facts, then, must be understood in the context of some high-level cosmological theory or theories. We now take this for granted, and refer to *the* actual facts, and we shall assume these are to be included in one history of a single consistent set.

The need to account for the actual facts in the consistent histories formalism gives a strong constraint on quantum cosmologies. It may well be impossible to satisfy this constraint in some, or all, versions of the formalism. For example, even assuming that the basic ideas of the consistent histories formalism and of quantum cosmology are correct, one can well imagine that the actual facts cannot be described by a history within a consistent set in the projection operator formalism that we consider in this paper, and that, as Gell-Mann and Hartle suggest (ref. 5, pp. 450–451; ref. 9, p. 3352) we really need to consider histories belonging to a branch-dependent consistent set—that is, one in which the projective decompositions after any time t are allowed to depend on the history selected up to that time.[19] In other words, the various proposals in the consistent histories program already constitute a collection of theories of nature, and these theories are in principle falsifiable. However, since all of these alternative proposals only exacerbate the problems we discuss in this section, we shall continue to restrict our attention to consistent sets of histories defined by branch-independent projection operators. Note also that, as the proposals considered in Section 7 illustrate, it remains possible that the fundamental principles of quantum cosmology (the description of the universe by Hamiltonian evolution from quantum initial and final conditions) are correct, but that no consistent histories formalism is.

We turn now to the future. Here we must be careful with our language, since we shall see later that the formalism has different interpretations, in which the discussion has to be phrased differently. The point, though, is that we can make predictions about the future only once a particular consistent set, incorporating the actual history is specified. In particular, if we want to predict—as we surely do—that we shall continue to experience a quasiclassical world, we need a consistent set on which this is described as following, with a high conditional probability, from the actual history. How good—that is, how strongly predictive—the cosmological

[19] Why, for instance, should a microscopic difference in local densities of chemical species in a laboratory experiment necessarily decohere in histories in which the solar system was never formed?

theory is very much depends on whether we can find an argument which genuinely depends only on the consistent histories formalism, the density matrix, and the particular interpretation of the formalism which we are using, and which selects out the relevant consistent set and makes clear why it is the one we must use. This, we suggest, is the crucial point on which, in the end, interpretations of the formalism stand or fall: we continue the discussion in the subsections below, which deal with the various possible interpretations.

We have still to explain what we mean by an interpretation of the formalism, and why it requires one. By an interpretation we simply mean a clear set of rules which explain how, in a general cosmological model, we are supposed to translate mathematical statements about the formalism into statements about physics. In this sense, of course, any mathematical theory of physics needs an interpretation. We suspect that the reason why physicists are often loath to discuss problems of interpretation is that, for many interesting theories, it seems that nothing is to be gained by discussion. There seems to be general agreement on how to make sense of classical mechanics, or general relativity, so that there appears to be little need to set out precisely what it is that we agree on. No such agreement yet exists for the consistent histories formalism. There seems to be real confusion about what the formalism means or is intended by its authors to mean. Many arguments in the literature rely on some notion of interpretation, and it is far from clear that all these arguments can actually simultaneously be made in any single interpretation.

In the remaining subsections we shall try, as far as possible, to find interpretations in which the key arguments in the literature can be justified. We shall see that there are at least four distinct interpretations of the formalism. Each interpretation is a serious attempt to make sense of the formalism; unfortunately, each has problems.

5.2. Omnès

Omnès uses the formalism as a way of analyzing the logical consistency of various collections of statements, reaching an interpretation of the formalism which is logically impregnable. As Omnès points out, this interpretation allows an analysis of all the propositions which the Copenhagen interpretation would produce, and a great many more. However, as we shall explain, this involves a loss of predictive power. Omnès' approach is deliberately cautious but uncomplacent in spirit. Omnès attempts to describe the minimum of logical structure necessary to make sense of the formalism but, as he has stressed,[37] his approach is not intended to exclude other interpretations, nor is it meant to discourage proposals to extend the formalism. Omnès' own primary concern is with quantum mechanics rather than quantum cosmology. We believe, however, that he has given a sensible procedure for evaluating quantum cosmological theories. We shall try to sketch the method below: for a less superficial treatment we recommend Omnès' thoughtful and detailed expositions.[3]

Suppose we have some cosmological theory which passes the first test for viability: that the actual facts are described by a history from at least one consistent set. We will certainly want to use the theory to make predictions. We may also want to make statements about the past. It is useful to divide these into two types: statements about events before any of the actual facts, and statements about events occurring at times in between actual facts. Past statements cannot be used to test the theory, but they are of interest nonetheless, and we shall use the terms *retrodictions* and *paradictions* to refer to the two cases.[20] In principle, Omnès does indeed allow us to make predictions, retrodictions, and paradictions. However, the only new statements which we can strictly deduce are what he terms *true propositions*—those which hold with probability one in any fundamental consistent set fine graining the actual history. We slightly extend Omnès' discussion by calling these *definite* true propositions and further define a *probabilistic* true proposition to be a proposition which holds with some probability (not equal to zero or one) in every consistent fine graining. There might possibly be surprising and nontrivial true propositions in some theories. However, in most theories the true propositions seem to be very limited. Let us consider the possibilities case by case.

The situation is clearest if the initial density matrix is pure and if we consider future true propositions. Let $\mathscr{S}$ be a consistent set including a history describing the actual facts. Suppose P is a true proposition, definite or probabilistic. Then $\{P, 1-P\}$ may be consistently added to any consistent fine graining of $\mathscr{S}$. Lemma 5 shows that if this is the case, then $\mathscr{S}$ must be maximally extended. In other words, either the future is entirely

[20] The term retrodiction has generally been used to apply to both, but it seems useful to make the distinction even in the nonrelativistic formalism. In the relativistic case, some such distinctions seem essential. If the actual facts describe projections localized within a particular region R of spacetime, we clearly want to refer respectively to derived statements describing propositions in regions which are in the future and past light cones of all points in R as predictions and retrodictions. We refer to all other derived statements—describing projections which are not localized either entirely in the future or entirely in the past of R—as paradictions. It is probably then useful to refine the class of paradictions further. For example, defining the future boundary $\partial_f R$ of R to be the set of points in the closure of R whose future light cones do not intersect R, one can define *weak predictions* to be statements referring to events lying in the union of the future light cones of the points in $\partial_f R$, and and similarly *weak retrodictions*.

predictable, or there are no true future propositions.[21] This may seem counterintuitive. One might like to believe that future predictions divide up into classical and quantum events, with the former inherently predictable and the latter intrinsically probabilistic; and it is tempting to think that this means the classical events correspond to definite true propositions and the quantum events to probabilistic true propositions. Within a single consistent set this remains an attractive picture: trivial consistent fine grainings give deterministic predictions and nontrivial consistent fine grainings give truly stochastic ones. But, when all the sets are considered on the same footing, as they are in the Omnès interpretation, this is not the case (at least for pure initial states). Either all future events are described by definite true propositions, or none are either definite or probabilistic true propositions. It is logically possible that the unpredictability we ascribe to quantum experiments is not genuine, and that in principle—if only we knew the full set of actual facts and the correct quantum cosmological theory—we could predict the outcome of all experiments with certainty. However, unless one is willing to make this leap, unpropelled either by evidence or argument, one is forced to accept that there are no future true propositions.

We have no analog of Lemmas 2 and 5 in the cases where the initial density matrix is impure or where there is a nontrivial final density matrix. Presumably, future true propositions (either definite or probabilistic) are not generally to be expected in these cases either, unless (perhaps) the actual history's set is in some sense close to fully fine grained. In any case, given the current state of our understanding, there seems no prospect of any future true proposition being identified.

Nor are past true propositions easy to come by. Lemma 7 tells us that if the actual history contains a repetition, so that the same projection P describes events at times t_1 and t_2, then further repetitions of P at times t with $t_1 < t < t_2$ *can* consistently be included in all consistent fine grainings. But these repetitive paradictions are the only general type of past true propositions which we have been able to identify. In particular, Example 4 shows that trivial past extensions, like trivial future extensions, do not generally describe true propositions. We know of no examples of probabilistic past true propositions, either retrodictive or paradictive.

Omnès was certainly well aware that there are relatively few true propositions (beyond those asserting the actual facts). However, the two

[21] Note that this precise result assumes the consistency conditions (2.7). It would be interesting to see what difference other (weaker or stronger) sets of conditions make. The problems with Omnès' interpretation however are not solved (nor claimed to be) by his choice of consistency conditions.

examples of true propositions identified in his paper do not deserve that title. Omnès uses the term *sensible logic* to describe any consistent set which incorporates the actual facts. (Here *logic* simply means consistent set.) He argues that (ref. 4, p. 366) "Measurement theory ... can be used to prove that the result of an experiment is always true. Another example comes from determinism: a past classical property that can be reconstructed logically in a deterministic way from present records can be said to be true, even when the sensible logics one is using involve only the present facts" and adds the comment:

> It seems that the two examples just given are the only ones. Thus one recovers essentially Heisenberg's point of view as far as quantum events are concerned, except for a deeper understanding of the meaning of truth among the phenomena themselves. The second example also answers an old question: It shows that a standing object at which nobody is looking is still nevertheless at the same place, and this can be taken to be true despite the fact that classical physics relies upon quantum mechanics.

This is very much the sort of result one would hope for. Unfortunately, neither example is correct, as Omnès now accepts.[22] The first, in fact, needs no new discussion of ours. Omnès' definition of an idealized measurement is an interaction between a measured system Q and a macroscopic object M, taking place between an initial time t_m and a final time t'_m, with the usual correlative properties, so that when the initial state of Q is an eigenstate of some discrete observable A with eigenvalue a_n, the final state of M is an eigenstate of a pointer observable B with eigenvalue b_n. The *result* of the experiment is "a quantum property, namely $A = a_n$, stating something about the measured system (for instance, a value for a spin component). It is a property holding at the time t_m when the interaction begins" (ref. 4, p. 362). We can, for example, take Q to be a spin-1/2 particle prepared in the state $\sigma_x = 1/2$ at time $t_i < t_m$ and then left undisturbed until the onset of the measurement interaction, and A to be the observable σ_z, and suppose that the final state of M points to the result $\sigma_z = 1/2$. The actual history here is the final state of M and the prepared state of Q. Now we can consistently extend the actual history to include projections onto σ_z at times t such that $t_i \leqslant t \leqslant t'_m$. However, as Griffiths points out (ref. 1, Section 4.3) we can also consistently extend the actual history to include projections onto σ_x at time t_m, and we cannot consistently add a projection onto σ_z at time t_m to this extension. [Strictly speaking, this is trivially true, since the formalism does not allow noncommuting projections to be applied at the same time. A criticism more sympathetic to the spirit of the discussion is that for "$\sigma_z = 1/2$ at time t_m" to be regarded as true, it should at least be possible to include the projection onto σ_z consistently at times

[22] We are very grateful to Roland Omnès for many helpful discussions on this point.

$(t_m \pm \varepsilon)$ for arbitrarily small ε, and this cannot be done for the negative choice of sign.]

Omnès' second example is particularly interesting. In the first place, one needs to establish some correspondence between classical determinism and the triviality of the corresponding consistent sets. Essentially, what is needed is to show that the Heisenberg picture quasiprojection describing a classical system in a cell of phase space at time t_1 is the same operator as the quasiprojection describing the system in a deterministically evolved cell at some other (earlier or later) time t_2. Omnès gives a careful discussion of the conditions under which this can be shown (ref. 4, Sections D–F). Let us assume that the correspondence has been established, and further that we can replace quasiprojections by projections. These are the most favorable possible assumptions, and they reduce the statement to the claim that if a projection operator belongs to the actual history, its past repetitions describe true propositions. We have shown above that is not generally so; only when the actual facts contain repetitions are there generally any true propositions, and in this case we can only make further, repetitive paradictions. In Omnès' language, one might—with careful analysis—be able to argue that if a tree was first discovered standing in a forest yesterday, and is still there today, and these discoveries are included in the actual facts, then it is true that it continued to stand while unobserved last night. But it is not generally true that it did so last year, or, for that matter, will do so tomorrow.

Omnès' position on the role of historical reasoning in identifying actual facts is unclear to us. What is clear is that "truth" is not the criterion which we, as adherents to ordinary historical reasoning, would use when building up the actual facts. For, once we have discovered the tree, we may well wish to include statements about its past classical history among the actual facts. We may, for instances want to deduce from its growth rings that it has been standing for the past 30 years, just as we want to deduce the past existence of live dinosaurs from the fossil record. These deductions are made using classical determinism (or more accurately quasiclassical near-determinism) rather than any criterion within the formalism.

In summary, an interpretation which relies on the use of true propositions is unlikely to allow any testable predictions. While Omnès did not explicitly discuss prediction, we believe he would agree this is a shortcoming in an interpretation. Our own view is that, because of the overwhelming variety of consistent sets and the strength of the condition that a proposition be includable in every consistent extension of the actual history, it will not be possible to resolve this particular problem without going beyond the formalism of consistent histories. As we shall discuss later, it seems to us that the simplest and most natural way to do this is to supplement the formalism by a set selection criterion.

However, remaining within the formalism of consistent histories, one can still make testable predictions using Omnès' concept of *reliable propositions*—propositions which have probability one in at least one, but not all, of the consistent sets which fine grain the set of the actual history. Again we term such propositions *definite* reliable propositions and define *probabilistic* reliable propositions to be propositions which hold with some probability not equal to one in at least one consistent fine graining. The statements physicists are accustomed to predicting are reliable propositions. For example, the predictions of classical determinism are examples of definite reliable propositions and the predictions of the results of quantum experiments are probabilistic reliable propositions. Unfortunately, as Lemma 2 makes clear, the same is true of many more statements which we do not want to predict: if there are probabilistic reliable propositions, then there are continuous families of consistent sets containing distinct reliable propositions. If we want to predict that a Stern–Gerlach experiment will end up with a pointer indicating the result $\sigma_z = 1/2$ with probability $p \neq 0, 1$, we have to accept that this prediction belongs to a continuous family of predictions, each of which is given equal status in Omnès' interpretation. It is hard to say precisely what all these predictions will be in a realistic laboratory setting, but they would have to involve statements about macroscopic superpositions of (at least) the pointer.

However, by considering the possible consistent fine grainings of the actual history, we can make perfectly good *conditional* predictions about the future. Thus, *if*, for the duration of an experiment, a bubble chamber is described within the formalism by projection operators delimiting the densities of chemical species within small volumes, and if that description includes the appearance of a particle track, then in the absence of external fields the track will be approximately linear and the probability distribution for the track angle will be that given by standard quantum calculations.

Note, though, that it is almost certainly impossible to use the formalism to make conditional predictions of the usual type, which involve quasiclassical inferences. For instance—and here we go beyond the discussion after Example 5—we almost certainly cannot deduce that if the earth is in the expected orbit throughout next Thursday then so will the moon be. To make the deduction, one would have to show that, from the actual facts describing the earth's trajectory quasiclassically, the moon's orbit (as described by appropriate projection operators) followed as a series of true propositions. This seems theoretically implausible—we have no proof, but doubt that any nontrivial paradictions are true propositions—and, practically speaking, an impossible calculation. To predict the moon's trajectory within the formalism we need the independent assumption that the moon will be described by the right sort of projection operator. If the projection operators are those onto densities of chemical species within small volumes of the moon's orbit, then they will describe the moon's expected orbit, whose predictability follows from the quasiclassicality of those operators. However, there are almost certainly other descriptions, consistent with the quasiclassicality of the earth but inconsistent with that of the moon.

Still, the conditional predictions which can be made are extremely useful. We would no doubt be impressed if a cosmological theory were produced which were consistent with the actual facts and which, in Omnès' interpretation, made many correct conditional predictions, even though it made no unconditional predictions. This would nonetheless be a very weakly predictive sort of theory. It would leave us puzzled as to why the world appears quasiclassical and unable to predict that this quasiclassicality will persist. It remains a perfectly consistent and testable scientific theory; it is logically possible that nature allows no unconditional predictions and that, within its domain of validity, an Omnès-interpreted cosmological theory is the best we can hope for. But persistent quasiclassicality is such a basic feature of our experience, and failure to predict it leaves such a significant gap in our understanding of the world, that we feel that some argument, or some new interpretation, or some new formalism, that would amend the situation should be sought.

The only argument relevant to this point in the literature is that of Gell-Mann and Hartle on the properties of IGUSes. We shall come to it later. Note though, that it cannot be made in the language of actual facts and true propositions.

5.3. Griffiths

In this subsection we very briefly sketch the views on interpretation that Griffiths has set out in a recent article.[2] We here discuss Griffiths' proposals only in so far as they relate to the problems of prediction with which we are concerned; the reader is encouraged to consult the original paper.[23]

To deal with the problem that different consistent sets allow different descriptions of nature, Griffiths introduces a set of rules for reasoning in quantum theory which he calls *Logic*. Griffiths' Logic contains, in a precise way, all of Omnès' sensible logics. However, Logic is a much weaker propositional calculus than ordinary classical logic. Most strikingly, the

[23] We shall not discuss retrodiction since, as far as we are aware, Griffiths has set out no definite position on the issue of historical reasoning and the nature of the actual facts.

rules of Logic state that if proposition P is inferred with probability one in a sensible logic and proposition Q is inferred with probability one in a second, incompatible, sensible logic, then "P" and "Q" are separately predicted but their conjunction "P and Q" is declared meaningless. Griffiths' case for Logic is, essentially, that it is a natural way to interpret the consistent histories formalism; it may be a conceptual weakening, but so (in some sense) were other advances in physics, such as the abandonment of absolute time in special relativity.

It seems to us that this analogy does not hold. Theoretical revolutions typically involve radical conceptual revision rather than simple weakening and, when they succeed, do so because of their greater predictive power. But on questions of prediction, Logic is of no help. For, once again using Lemma 2, we see that if there are nontrivial predictions to be made at all and if the initial state is such that the quasiclassical variables decohere, then Logic predicts both that "the universe will continue to be quasiclassical" and "the universe will not continue to be quasiclassical". The fact that this does not Logically imply "the universe will continue to be quasiclassical and the universe will not continue to be quasiclassical" saves the scheme from self-contradiction, but will not, in itself, be of comfort to many.[24] Griffiths' view is that one needs not only the formalism (of which Logic is a part), but also a theory of human experience in order to make unconditional predictions.[38] However, if one appeals to a theory of experience in order to select a consistent set in which we can make unconditional predictions, it is not clear that one has any scientific need for a logical scheme relating the statements from distinct consistent sets.

In summary, the predictive content of Griffiths' interpretation is meant to follow from an implicit theory of experience. We shall consider what this entails when we come to examine Gell-Mann and Hartle's ideas.

5.4. Many Histories

We now set out another interpretation of the formalism, which requires that we take seriously the notion of many coexisting but noninterfering histories. This is inspired by Griffiths' interpretations but requires no change in the rules of logic: different histories peacefully coexist.[25] In

[24] Some may regard these statements as alternative rather than contradictory descriptions. Note though that a Logical analysis of Lemma 8 or the standard no-local-hidden-variables results does produce direct contradictions.

[25] We are grateful to Bob Griffiths for mentioning this interpretation in discussions.

addition to the changes needed to ensure this, we also introduce a slight technical modification by insisting on sets which are exactly (rather than approximately) consistent.

To set up this interpretation, we require that from each of the fundamental consistent sets $\mathscr{S}$ precisely one history $H(\mathscr{S})$ is chosen, the probability of any particular history being chosen being precisely its probability $p(H)$, defined in the usual way. The interpretation then states that all of the chosen histories, and no others, are realized. The true description of nature, in this interpretation, is the list of all the chosen histories $\{H(\mathscr{S})\colon \mathscr{S}$ a fundamental consistent set$\}$, and each history constitutes a complete description of one of an infinite collection of (for want of a better term) "parallel worlds".

Perhaps it is worth stressing a few details since, despite its simplicity, this interpretation seems to cause confusion. It is important that just one history from each set is realized. If one postulates that all the histories of each set $\mathscr{S}$ are realized (as is sometimes suggested[(39)]) then no role has been assigned to the probabilities, and there seems no obvious way of introducing further assumptions which would allow probabilistic statements to be deduced. In other words, there is no precise analogy here with Everett's suggestion[(40)] that all possible results of quantum measurements are always (in some sense) equally realized: this is a many-histories interpretation, but the many histories belong to different consistent sets.[26]

It is also crucial for this interpretation that exactly consistent sets are used. The statement that a history is realized is supposed here to be an absolute statement of fact, quite independent of any anthropocentric considerations about the accuracy to which we can test the history's predictions.[27]

Some may feel uneasy at the use of a probability distribution to select histories, or find it hard to make sense of the idea that the chosen histories "are realized". To take the second point first, we mean simply that the chosen histories are supposed to correspond to nature; in just the same way, general relativity admits a large class of four-dimensional manifolds, equipped with the appropriate tensors, as solutions, of which one is supposed to correspond to the universe we inhabit. As for the use of probability, we see nothing particularly distinctive about its use here: the relevant

[26] It should nonetheless be true that all possible results of most quantum measurements are realized, but for a different reason: most measurements can be described in infinitely many different fundamental consistent sets if Lemma 2 is any guide, or if our conjectures about the generic dimensions of the solution spaces for the consistency equations are correct.

[27] Attempts have been made to consider approximate histories.[(39)]

debate is really not over the consistent histories formalism at all, but over whether any fundamental probabilistic theory of nature is meaningful or acceptable.[28]

Still, the idea of describing reality by a collection of sets of histories is unfamiliar and perhaps uncomfortable. We shall not argue that the picture is attractive, or particularly plausible: we do claim, though, that the proposal makes sense. In any sensible cosmological theory, the interpretation will produce a highly structured set of mathematical objects exhibiting complex patterns and it seems to us that, although its character is rather different from familiar theories such as general relativity, it therefore should be considered a candidate mathematical theory of physics. We now need to examine whether our experiences, and the world they describe, could be reflected by some subset of those mathematical objects, and if so, whether that subset is identifiable *a priori* or only *post hoc.*

Here we are not optimistic. The interpretation seems a very natural one, using as it does only the basic notions of consistent set, history, and probability. It also gives an unambiguous description of a definite reality, which some like. In practical terms, though, it is indistinguishable from Omnès'. Our own experiences are supposed to be described by one of the realized histories from a given cosmological theory. To test this hypothesis, we first have to be able to find some consistent history in which we can describe our past and present experiences, and any historical events which we decide we should deduce from records. We have, in other words, again to find Omnès' actual facts within the theory. We then turn to predicting the future, and—for the reasons outlined above—discover that we can make conditional predictions, but cannot make unconditional ones. Among the histories which are realized, there will be many which describe the actual facts. In most of these, quasiclassicality will not persist in the future. We are again left puzzled as to why quasiclassicality does persist in the consistent history we actually experience.[29]

[28] We are happy with such theories, but this debate is beyond our scope.

[29] This interpretation does perhaps give truly hardened anthropicists a little more room for maneuver. It might possibly be argued that some theory of experience will be found which predicts that we have no awareness in nonquasiclassical histories. It could then be maintained that an uncountable number of close (though imperfect) facsimiles of ourselves do indeed perish in histories where quasiclassicality dissolves, while we survivors are inevitably left bemused by our fortune. We do not ourselves take this type of argument seriously, but shall not pursue the point—which involves questions about the nature of identity and the validity of anthropic reasoning that go beyond our scope. We note, though, that any criticism of Gell-Mann and Hartle's position—which we shall discuss in the next subsection—applies equally to this argument, so that the extravagant multiplicity of facsimiles and the invocation of anthropic principles seem to be drawbacks which bring no compensation.

5.5. IGUSes and Quasiclassical Domains

5.5.1. Preliminaries. We now come to Gell-Mann and Hartle's seminal work on the consistent histories formulation of quantum cosmology. As in the earlier two subsections, our aim is to find a formal interpretation that accords with Gell-Mann and Hartle's views and in which we can draw their stated conclusions. Here we must admit that we are in difficulties. Gell-Mann and Hartle's writings on the consistent histories formalism sparkle with ideas, proposals, suggestive observations, and rhetorical arguments. The totality of views expressed does not, it seems to us, constitute a coherent and internally consistent interpretation. In part, this is a consequence of an evolving view: for example, some of their most controversial interpretational arguments are contained in an as-yet-unpublished paper, the current revised version of which[(41)] differs from the preprint version which is presently available on the bulletin boards.[(28)] Here our criticisms refer in detail to the earlier version.[30]

In part, though, we believe that Gell-Mann and Hartle's problems arise from an attempt to maintain two irreconcilable lines of argument. As we have repeatedly stressed, a key question for an interpretation of the consistent histories formalism is whether it does or does not *explain* our perception of a quasiclassical world. It appears that in developing their program, Gell-Mann and Hartle had the problem of the apparent persistence of quasiclassicality very much in mind. They introduce the notions of an information gathering and utilizing system, or IGUS (ref. 5, p. 454), and of a quasiclassical domain (ref. 5, p. 445), and they argue that certain types of IGUS—such as ourselves—"evolve to exploit" the regularities in some particular quasiclassical domain: "The one reason such systems as IGUSes exist, functioning in such a fashion, is to be sought in their evolution in the universe" (ref. 5, p. 454).

It is, they argue, as a consequence of this that we perceive quasiclassicality as a persistent property of the world around us, although, they claim, nothing in the formalism ascribes a special role to persistently quasiclassical consistent histories:

> If there are many essentially inequivalent quasiclassical domains, then we could adopt a subjective point of view ... and say the IGUS 'chooses' its coarse graining of histories and therefore 'chooses' a particular quasiclassical domain ... It would be better, however, to say that the IGUS evolves to exploit a particular quasiclassical domain or set of domains. Then IGUSes, including human beings, occupy no special place and play no preferred role in the laws of physics. (Ref. 5, p. 454.)

---

[30] It would be confusing to do otherwise, given that Gell-Mann and Hartle's subsequent, and still tentative, modifications are motivated at least in part by the criticisms made here. We will, however, comment briefly on the later version in footnotes.

This, Gell-Mann and Hartle maintain, is an argument which explains, within the consistent histories formalism, the mystery of why we continue to experience a quasiclassical world. It is an argument which, we suggest, to make sense requires a particular, not at all straightforward, type of interpretation of the formalism. The same, we suggest, is true of Gell-Mann and Hartle's discussions (ref. 28, p. 33; see also ref. 42, pp. 164–165) of communication between IGUSes in incompatible quasiclassical domains or (as rephrased) of inference of each other's features.[41] We attempt to set out the sort of interpretation that is needed, and the difficulties that it involves, in Sections 5.5.2 and 5.5.3. Gell-Mann and Hartle do not wish to advocate such an interpretation. While we sympathize with this view—since the interpretations described in Sections 5.5.2 and 5.5.3 seem to us ad hoc, inelegant, and convoluted—we do not see how their arguments can otherwise be supported.

On the other hand, Gell-Mann and Hartle also maintain that one can make *no* predictions within consistent histories that are not conditional on the set in which they are made.[41,43] We see this, in itself, as the core of a natural and internally consistent interpretation of the formalism—which, for completeness, we shall set out in Section 5.6. It leads, though, to the conclusion that the persistence of quasiclassicality is an unexplainable mystery, and, we suggest, contradicts the intent of Gell-Mann and Hartle's extensive discussion of IGUSes and their evolution. It also entails a novel, and to us unacceptably ahistorical, view of the past. Hartle's statement that "dinosaurs *did* roam the earth many millions of years ago" (ref. 16, pp. 77) is not to be read as a statement about a definite, observer-independent physical event; it is meant as a shorthand for the claim that there is a particular consistent set which includes the actual facts presented to us and in which quasiclassical dinosaur density projection operators followed largely deterministic equations of roaming through the Cretaceous era. Other consistent fine grainings of the actual facts contain different operators, incompatible with such a picture. All these pictures are just theoretical contstructs—a vast library of consistent historical fables, of which one is free to choose any or none.[31,32]

We have found it impossible to reconcile these two positions, and feel that some clarification of Gell-Mann and Hartle's views would be useful. There seem, however, to be divergent views on what a clearly defined interpretation involves. Hartle, for example, gives a revealing discussion of what

[31] "Fable" is not, of course, Gell-Mann and Hartle's preferred term.

[32] Whitaker,[44] citing these comments of ours, puts it well: in this view the formalism might better be described as one of "consistent stories" rather than "consistent histories".

he calls the "buzzwords" of quantum mechanics in the appendix to his Jerusalem lecture notes,[16] and suggests:

> Only a casual inspection of the literature reveals that many interpreters of quantum mechanics who agree completely on the algorithms for quantum mechanical prediction, disagree, often passionately, on the words with which they describe those algorithms. This is the 'words problem' of quantum mechanics. The agreement on the algorithms for prediction suggests that such disagreements may have as much to do with people as they do with physics. This does not mean that such issues are unimportant because such diverging attitudes may motivate different directions for further research. However, it is important to distinguish such motivation from properties of the theory as it now exists.

Many physicists would certainly agree, and it is easy to understand why—interpretational arguments have often led nowhere, and real progress has often been made by putting them aside and concentrating on development of the formalism. Still, the consistent histories formalism is a new description of quantum theory, and its use in quantum cosmology does raise new questions, in the light of which it seems sensible to reexamine the problem of interpretation.

We persist in worrying about interpretations of consistent histories, not only in the hope of motivating future developments, but also because we do not believe that there is full understanding of, or agreement on, the algorithms for prediction. For example, the interpretations we described in the preceding subsections clearly do not predict, unconditionally, that we shall observe a persistently quasiclassical domain. Gell-Mann and Hartle argue that this can be explained, and we shall exhibit the implicit assumptions that we believe this argument involves. Similarly, various interpretations differ on whether, and when, communication between IGUSes is possible, and even on what is meant by the word. There seems to be no consensus on how to resolve all these questions, and we do not believe this merely reflects lack of thought or unfamiliarity with the literature. It seems to us, rather, that there are physical questions which simply cannot unambiguously be answered by referring to the existing literature, that the various possible answers rely on arguments which have not always been set out in detail, and that it is necessary to try to set out the possibilities.

The approach we have taken in this subsection is to look for a formal interpretational scheme that can accomplish the aims set out in Gell-Mann and Hartle's IGUS discussions, those of predicting our experience and explaining why it is of a quasiclassical world. This divides into two parts. We first examine the serious problems inherent in defining an IGUS. Then, under the assumption that these problems can somehow be solved or finessed, we set out some interpretations which are essentially restrictions of the many-histories interpretation to some subclass of the consistent sets. We discuss Gell-Mann and Hartle's ideas within these interpretations, and describe the new and rather awkward consequences which follow. We stress, though, that these interpretations are entirely unauthorized, and we expect that Gell-Mann and Hartle will reject them.

5.5.2. Problems of IGUS Definition. We turn first to IGUSes. The idea of an IGUS—a type of creature which is coupled by some form of sensory organs to its environment, able to model the local environment by some form of logical processing, and able to act on the results of its computations—is simple enough. To characterize an IGUS within a particular consistent set it must, of course, be identified with a coarse graining of that set, just as any subsystem of the universe must. The operators involved in the coarse graining must possess a certain amount of temporal stability—they must be more or less the same operators at neighboring times. The functioning of the sensory organs of an IGUS, its logical processing, and its resultant behavior can all then be described by correlations of the coarse grained operators.

IGUSes, then, are part of the formalism; had we wished, we could have characterized supernovas in the same sort of way. There is no significant notion of free will which can be attached to the formalism. Just as in general relativity, IGUSes do what the equations say they will do, though the predictions here are probabilistic. Thus, if we study possible experiments which might be performed by an IGUS, we must bear in mind that it is the boundary conditions and probability rules—not any extraphysical free choice of an experimenter —which determine the experiment that actually takes place. What any particular IGUS perceives can be bounded, though not completely delineated, by the standard assumption of psychophysical parallelism.[45] That is, the perceptions, sensations, thoughts—in sum, the consciousness—of an IGUS must be paralleled in a natural way by some description of the IGUS in the formalism. Again, there is nothing unusual here; the same assumption is necessary to make sense of any physical theory.

Now we come to an awkward question. To set out any argument about the experiences of quasiclassical IGUSes such as ourselves, and to calculate any probabilities relating those experiences, we need to apply psychophysical parallelism. But to what? Exactly what in the formalism is supposed to correspond to a single IGUS? One possible answer is suggested by the many-histories interpretation of the previous subsection: each fundamental consistent set contains a realized history, and if an IGUS can be identified in the realized history, we can apply psychophysical parallelism. This is a perfectly clear, and quite natural, prescription, but it is certainly not the prescription which Gell-Mann and Hartle have in mind. It tells us that if a consistent set contains a quasiclassical description of an IGUS up to time t, and if—once again, following Lemma 2—that set has a continuous family of extensions, then we must apply psychophysical parallelism to each copy of the IGUS in each of the extended sets. We conclude that an IGUS just before time t, being unable to tell which set to use, has no reason to expect quasiclassicality to persist beyond time t, for there is clearly no reason to suppose that psychophysical parallelism applied to a nonquasiclassical description leads to quasiclassical experiences.

In fact, Gell-Mann and Hartle are quite clear that they do not want to speak of many copies of the same IGUS, associated with the various extensions; in their interpretation of the formalism, there is only the one IGUS.[(43)] They do not want IGUSes to correspond to a pair of objects: a coarse grained history and some particular consistent set in which it can be found. Instead, they make the correspondence directly with the coarse grained operators.[33] However, once these coarse grained operators are identified, the remaining details of the consistent set are not included in the definition of the IGUS. There are many consistent sets which contain all of the coarse grained operators. Any of these can be used, but this is not supposed to imply a multiplicity of IGUSes corresponding to the sets. It is further supposed that a quasiclassical IGUS will continue to have the potential for quasiclassical experiences so long as there is some appropriate consistent set which contains its quasiclassical coarse grained operators.

This, though, raises some new technical problems. First, in defining an IGUS, one runs into the familiar problem of where to make the cut between organism and environment. Should humans be described only by the quasiclassical physics inside their skins, or should a little of the environment be included so as to allow a quasiclassical description of the information-gathering process? Or should the relevant operators describe only the brain, or some subset of brain activity? No principle is known that would allow us to identify any natural cut, and it is hard to see how one could be found. Of course, these problems would be relatively unimportant if an IGUS were simply a rough characterization of an interesting class of objects within an already well-defined interpretation of the consistent histories formalism.[34] They are of crucial importance here because, we

[33] More accurately, an IGUS should generally be described by many coarse grained histories of nonzero probability—histories of zero probability may be ignored—within a consistent set, since the course of its life need not generally be classically determined. We might, for example, arrange our travel plans to depend on the results of quantum experiments. A precise interpretation, capable of supporting Gell-Mann and Hartle's arguments, would have to allow for this. We shall ignore this complication, although it seems rather serious.

[34] This, in fact, is how Gell-Mann and Hartle's discussions of IGUSes are often understood.

claim, Gell-Mann and Hartle's arguments require an interpretation which uses the IGUS as a fundamental concept in its very definition. We shall give our attempts at such an interpretation in the next subsubsection.

A related point, to which we shall return, is that while it is a plausible assumption that the experiences of an IGUS are described in the same set as that of some coarse grained operators which describe its information gathering and utilizing behavior, it is still an assumption: in the end, it will need to be justified by attaching a theory of experience to the formalism.

And then there is no reason to expect to find a particular coarse graining which is uniquely well-suited to describing the IGUS. Which one should we use? Why? This raises the question of how we can tell whether two IGUSes are the same. One could imagine an IGUS such as the reader being described in terms of two slightly different sets of operators, not coarse or fine grainings of each other, not equivalent by our proposed equivalences. For example, one could use the hydrodynamic variables defined by integrals of energy densities over small spacetime regions marked out by some lattice which has the minimal grid size to give decoherence, and also the same variables but defined with the lattice translated by half a lattice spacing in some direction. We would want to identify the two IGUSes in these two consistent sets and give general conditions under which two descriptions are to be identified as the same IGUS.

We can sharpen this last problem. Since we want humans to be examples of IGUSes, we must allow the operators to vary slightly over time in such a way that after a long time they can be significantly altered. This in itself poses no problem in describing human IGUSes, but it leaves open the awkward hypothetical possibility that there exist two consistent sets, one containing an IGUS whose lifespan is described by some series of operators and another in which an IGUS is described identically to the first for half its life and which then slowly diverges from the first, until the operators which describe it are entirely different. We have in mind here a picture in which the operators from the two sets are inconsistent and yet *both* sets of operators describe the IGUS gathering and exploiting information from an environment: nothing in the formalism or in the definition of an IGUS implies that this is impossible. If such an IGUS is calculating before the split, how is it supposed to predict its future? Which set of operators should it use? Even a probabilistic answer to the last question would be satisfactory, but the formalism gives none. Still worse possibilities can be imagined, involving many branchings and recombinings of the IGUS operators.

All of this suggests that, if we are really to use the formalism to speak about the experiences of general IGUSes, then, even if we grant the assumption that the experiences of the IGUS are closely linked to its quasiclassical description, we need a much more precise notion of how to define an IGUS than anything which has so far been suggested. We need, in fact, a set of rules which tell us, *a priori*, which sets of coarse grained operators describe IGUSes and which IGUSes are distinct. We can either, optimistically, hope that these rules can somehow be formulated so as to lead to no branching ambiguities of the type we have just mentioned, or accept the ambiguities as possible features of the formalism; we assume for the moment that the ambiguities can be removed. There seems no prospect that any sensible, precise definition of an IGUS will be found; still, let us assume there is one.

5.5.3. Interpretational Problems. It is easy, given this assumption, to provide a formal, IGUS-centric interpretation. Given a cosmological theory, we can identify all the consistent sets of coarse grained operators which correspond to (distinct) IGUSes in the theory. Precisely one history is chosen from each of these sets, the choice being governed by the standard probabilities, and the experiences of the IGUS correspond to the history chosen from its set. We are forced, in this interpretation, to give up the hope of describing what happened in the very early universe, or anything else which IGUSes do not experience. In return we hope that we shall be able, for example, to predict, *unconditionally*, that human IGUSes will continue to experience a quasiclassical world. However, it will be apparent that this interpretation is solipsist. For there is, in this interpretation, no correlation between the experiences of these splendidly isolated IGUSes; each may well believe itself in communication with others, but the others may be experiencing a quite different history, or nothing at all. Note also that by this device we have not found a fundamental answer to the question of why quasiclassicality seems to persist; we have simply shifted the difficulty so that it becomes part of the IGUS definition problem.

Clearly we are not forced to use the definition of an IGUS in an interpretation of this sort; one can try to eliminate the solipsism by using something bigger. The obvious choice is that of a quasiclassical domain.[35] The same problems arise: quasiclassical domains might branch; they are certainly imprecisely defined; we need a set of rules which precisely characterizes the coarse grained operators corresponding to a particular domain. Given such a set of rules, we can again follow the same path: we use the probability weights to choose precisely one history from each of the sets of

[35] Gell-Mann and Hartle's program aimed at characterizing the notion of a quasiclassical domain has sometimes, understandably, been read[(19)] as an attempt to define an interpretation of precisely this form. This, though, is not what Gell-Mann and Hartle intend: they do not, for example, regard the question of whether a good characterization exists as a test of whether a sensible interpretation of the formalism can be defined.

coarse grained operators corresponding to quasiclassical domains, and then suppose that each of these histories is realized, and in particular that each describes the experiences of all the IGUSes within the given domain. This allows IGUSes in a quasiclassical domain to have genuine communication, in which they can, for example, discuss features of the domain which form part of their common experience. Again, we have not explained our perception of persistent quasiclassicality by this interpretation; we have simply assumed it. Nonetheless, this seems more promising than the previous interpretation: it avoids complete solipsism, and it is easier to believe that quasiclassical domains can be characterized elegantly than that IGUSes can.

However, Gell-Mann and Hartle's recent discussion of IGUS communication cannot be carried out in either of these interpretations. They envisage that:

> If two essentially distinct quasiclassical domains exist, they may overlap, in the sense of having some features in common so that sets of histories possess a common coarse graining. It is then possible that IGUSes in different but overlapping domains could make use of some of the common features and thus communicate by observing and manipulating alternatives of this common coarse graining. (Ref. 28, p. 16)

We have to be careful about the meaning of communication in an interpretation of the consistent histories formalism. It is important to distinguish between genuine communication and the IGUSes' beliefs about the matter. For instance, either of the interpretations we have described allows a situation in which we can describe two IGUSes—perhaps both working on the interpretation of the consistent histories formalism—each of which believes itself to be corresponding with the other, but in which there is no agreement about the contents of the correspondence. No one, though, would describe this as communication. We would like to translate the statement that two IGUSes are communicating to mean that the formalism allows us simultaneously to describe both their experiences, including some correlated pieces of information, and gives a probability distribution for the experiences which respects the correlations. It seems to us that anything weaker can hardly be referred to as communication.

Now, if the paragraph quoted is taken in anything like the ordinary sense—in which we live in and experience our domain, and aliens live in and experience theirs; the formalism is taken to mean that neither we nor they exist in multiple copies; and we are understood to be transmitting and receiving between the domains, and agreeing on the information transferred—then it is incorrect.[36] Probabilities of such a process may not be

[36] One can easily convince oneself of this by trying to describe the communication process in a simple model.

assigned in either of the interpretations given above; neither, as we discuss below, can it be done in any extended interpretation without some extra axiom.

Is there another interpretation in which we can understand the paragraph? We now investigate some possibilities. Since the first interpretation above allows no communication at all, and the second allows communication only within a quasiclassical domain, they cannot support Gell-Mann and Hartle's conclusion. So, why not adopt the obvious solution, and make a similar interpretation using something other than quasiclassical domains? The problem is—what? If a pair of IGUSes I_1 and I_2 live in distinct quasiclassical domains D_1 and D_2, and if there is some consistent set $\mathscr{S}_1$ containing a description of both IGUSes (at least for some period of time), one can always declare by fiat that one history from $\mathscr{S}_1$ is realized, in which case communication between I_1 and I_2 is restored. But how, abstractly, is $\mathscr{S}_1$ to be characterized? And what are we to do if there is another IGUS I_3 in a third domain D_3, and sets $\mathscr{S}_2$ (containing a description of the IGUSes I_1 and I_3) and $\mathscr{S}_3$ (containing I_2 and I_3), with the property that $\mathscr{S}_1$, $\mathscr{S}_2$ and $\mathscr{S}_3$ are pairwise inconsistent? If I_1 and I_2 are in communication, then so must each pair of IGUSes be. Yet the formalism gives no joint probability distribution for the experiences of the three and, even if we wished to go beyond the formalism, Lemma 8 tells us that we cannot generally find *any* joint probability distribution consistent with those describing the pairwise communications: in other words, the probabilities specified by the decoherence functional distribution on the sets $\mathscr{S}_i$ are incompatible.

The question is clear: either I_1 and I_2 are in communication, or they are not. In particular, we ought to be able to derive a probability for I_1–I_2 communication, and similarly for the other pairs. We cannot get around this by suggesting that I_1 has a choice of communication with I_2 or I_3, but not both, since the formalism allows no notion of free choice: however things seem to I_1, it must be possible to calculate the probabilities of the decisions I_1 appears to make. Nor can we exclude the possibility of communication between any of the pairs without either excluding the possibility of communication between all distinct quasiclassical domains or else finding some rule which explains under which circumstances IGUSes in distinct domains can communicate and which gives a joint probability distribution for the experiences of all IGUSes in communication. The first of these would contradict Gell-Mann and Hartle's discussion; the second would go beyond the formalism.

Here we have used only the assumption that, if a pair of IGUSes are in communication, there must be a joint probability distribution for their experiences, which we believe must be true under any standard definition of communication. We have *not* explicitly assumed that pairwise communication implies that information can freely be shared among the three IGUSes; we note, though, that the literature again gives no clear and generally agreed rule which either forbids this or imposes any other well-defined constraint on the information which an IGUS can pass on.

We must conclude, then, that Gell-Mann and Hartle's recent discussion of interdomain communication is wrong.[37]

Since this is a strong claim, let us repeat the main argument. We proposed two interpretations in which the fundamental notions are, respectively, IGUSes and quasiclassical domains. In the first, there is no communication at all; in the second there is communication only within quasiclassical domains. Consider then the naturally extended interpretation in which two IGUSes are in communication when there is a consistent set that includes them both. This interpretation leads to a contradiction when we have three IGUSes, pairwise in communication, for which there exists no joint probability distribution on all three which respects the joint distributions for the pairs. The interpretation demands that a given IGUS concludes that it is communicating with the other two, but this contradicts the lack of an overall joint probability distribution.

What of Gell-Mann and Hartle's discussions in which IGUSes "evolve to exploit" certain variables? This is the argument IGUS-centric interpretations are designed to support. Given a clear definition of IGUSes, one could indeed hope to study their general properties, observe that IGUSes do tend, as time progresses, to become described by operators which correlate better and better with more and more quasiclassical variables and which perform more and more sophisticated calculations simulating quasiclassical equations of motion. And, given the assumption that IGUSes are *the* fundamental objects in the theory, one could then use these results (together, as always, with the assumption of psychophysical parallelism) as an *explanation* of their (and in particular our) perception of a persistently quasiclassical world.

[37] We should mention here that the current draft of the paper[41] differs from the preprinted version quoted from in the text in that it contains no explicit reference to "communication" between IGUSes. We believe, however, that our criticisms are just as relevant to the current draft. The claims remain virtually unaltered:[41] "It is ... possible that IGUSes evolving in different but overlapping realms could make use of some of the common features ... to infer features of each other. The problem of inferring the existence of other IGUSes using distinct but overlapping realms is not so very different from that involved in ordinary searches for extraterrestrial life." That it *is* rather different, we believe we have demonstrated in the text. The IGUSes are now inferring the existence of each other rather that communicating, but our arguments depended only on the translation that there exist a joint probability distribution for the IGUSes and so apply equally well.

In an interpretation in which quasiclassical domains, rather than IGUSes, are fundamental, IGUS evolution can be described and investigated. The only variables that IGUSes can *possibly* be described as exploiting, however, are quasiclassical, so that one can hardly speak of IGUSes as *evolving* so as to exploit these variables rather than others. Since it is in this sense that Gell-Mann and Hartle's statements about evolution are to be understood, this interpretation is inadequate to accommodate them. And, of course, in this interpretation, the persistence of quasiclassicality is imposed by fiat rather than explained.

In the remainder of the subsection we shall reiterate the properties of the IGUS-centric interpretation which is the closest we have been able to come to the spirit of Gell-Mann and Hartle's ideas of explaining our experience (though, as we have seen, it fails to allow some of their more detailed assertions).

We can—it is hoped—predict what we shall experience by identifying the relevant set of future projection operators, and then test these predictions. We have mentioned the problem of branching ambiguities. Perhaps there is no general way of finding these experiential projection operators for arbitrary IGUSes, for human IGUSes in arbitrary cosmological models, or even for some future humans in our own cosmology. But in defense it might be argued that none is needed: any IGUSes afflicted by branching ambiguities will (or may) experience something after the branching; even if they have perfect knowledge of the boundary conditions they cannot make any probabilistic statement about what (or whether) they will experience. This would be scientifically unfortunate for them, but need it concern us? Should the hypothetical problems of other IGUSes, or even the real problems of humans in some future circumstances, be a deadly objection to a scientific theory? If, at present, we can predict our own future experiences, should we not be content?

Possibly, but this still begs the question: *can* we identify the operators that allow us to predict our own future experiences? Nobody seriously claims that the operators can be precisely identified at the moment. But many claim that we can say roughly what form the operators must take: that they are almost surely quasiclassical descriptions of the brain in operation. It is important here to make the distinction between the practical question (what is the state of things?) and the theoretical question (do we understand why?). On the practical question, we agree: we have no reason to suppose that anything other than a quasiclassical description of the brain is needed. But we see the theoretical question as crucial. At a fundamental level, we do not understand why the operators relevant for describing our experience must be a subset of the quasiclassical projection operators which are most convenient for describing the brain's logical functioning. It is an assumption—a necessary assumption to make the formalism describe our experience; a plausible assumption as part of a general theory of experience; but still, no more than an assumption. This forces the conclusion that we cannot answer the question as to why we perceive a persistently quasiclassical world within the formalism. Instead, this question becomes part of the general problem of finding a theory of experience. And to do the job we require—to explain, finally, why we perceive a quasiclassical world—the sought-for theory of experience will have to be framed not in quasiclassical terms (which would merely assume the answer), but in the language of the consistent histories formalism.

At the risk of repetition, let us stress this crucial point. To do the job, a theory of experience would have to take as input some characterization of an IGUS and return as output a list of the projection operators describing its experiences. There is no *a priori* link between the behavior of the IGUS, as described in some history of some consistent set, and its experience. Indeed, it is crucial that the link is generally absent, since we can be described (albeit perhaps inconveniently) in many consistent sets which clearly are not suitable for describing our experiences. Hence the theory would have, for each IGUS, to construct such a link with a particular set and explain, from its axioms, why this set is the correct one.

Nothing in the consistent histories account of our evolution implies the form of such a theory, and no coherent theory of experience, either quasiclassical or quantum theoretic, has yet been framed. There is no consensus on the form such a theory might take and, indeed, some doubt that a scientific theory of experience can ever be found. Nor is there any evidence that any such theory should be naturally framed in the language of the consistent histories formalism.[38] For consistent historians to appeal to an unknown theory of experience in order to make unconditional predictions is then, at best, to postpone the question of whether such predictions can in fact be made.

Another concern is the solipsism which the IGUS-centric interpretation implies. We can describe our own experiences within the formalism. We can also describe other IGUSes', but not simultaneously.

We cannot, moreover, see how to avoid solipsism in extended interpretations of this type: to avoid it one needs to ascribe reality to at least one set that extends our experience. If one ascribes reality to all such sets,

[38] Moreover, if a theory really can be found to identify the scientifically relevant set or sets of histories, perhaps the consistency criteria need not separately be imposed. For one can certainly imagine the consistency of the scientifically relevant sets following directly from the hypothetical theory. In this case there seems no need to discuss general consistent sets at all.

one then runs into the difficulties encountered in our discussion of the many-histories interpretation: when these histories contain quasiclassical descriptions of IGUSes, these descriptions generically cease to be quasiclassical very abruptly. So one requires a selection principle as to which sets to ascribe reality to—quasiclassical domains are a possibility we have mentioned, but this, again, has its own problems.

Neither of the features to which we have drawn attention is unprecedented. It is disappointing that the problem of the apparent persistence of quasiclassicality is left as a mystery to be solved by some future theory of experience, but other approaches to quantum mechanics use the same strategy, which goes back at least to Wigner.[46] It is awkward that experience thereby becomes entangled with the quantum formalism at a fundamental level, but of course this could conceivably turn out to be unavoidable—who can tell for sure? Likewise, it is not unknown for interpreters of quantum mechanics to find themselves driven to solipsism, and some physicists find this a respectable scientific position.[39]

Gell-Mann and Hartle characterize their program as essentially an elucidation of quantum theory. Indeed they often refer to the consistent histories approach as "quantum mechanics" or "quantum theory." We would quarrel with this nomenclature, since the consistent histories is only one among many, currently existing and potential, descendants of the quantum mechanics of Bohr and Dirac. We do, however, agree with them that it is a particularly important development. One can trace a line of thinking in which, gradually, through the work of many people, the provisional or irrelevant notions cluttering certain earlier interpretations have been stripped away. No notion of measurement by classical apparatus is necessary; no collapse of the wave function need be considered—this much was suggested by Everett. Everett's attempt at a many-worlds interpretation, though, was left incoherent by the lack of any criterion by which physical histories could be extracted from the formalism. Griffiths, Omnès, Gell-Mann, and Hartle have supplied such criteria. We are left—and we think this is most clearly seen in the quantum cosmological picture of Gell-Mann and Hartle—with a clear characterization of the fundamental points on which there is disagreement about the interpretation of standard quantum theory. As Penrose has stressed in other contexts,[47] there is a fundamental divergence between those who seek to add a theory of experience to quantum theory and those who would prefer a theory of reality. Perhaps the debate really has finally reached impasse, and can only be resolved by new science on one side or the other. Certainly, each is left

[39] We do not, but this debate goes beyond our scope.

with a hard task: one side needs to construct the theory of experience which will shore up their position; the other to find a realistic theory which replaces or extends quantum theory.

One can imagine a variety of radically different outcomes to these programs. But, if the consistent histories formalism is fundamentally the right setting, then each program could be completed by a selection criterion which picks out particular consistent sets[40]: on the one side, a class of sets would be selected to describe the experiences of IGUSes, or at least humans; on the other, a single set would be used to describe reality. It is common for those who appeal to a theory of experience to claim that, while alternatives are not excluded, theirs is the natural null hypothesis. In the light of the consistent histories formalism, and in particular of these last observations, we suggest that this position is very clearly untenable.

5.6. The Unknown Set Interpretation

Most of the interpretational ideas in the literature add complicated new ideas ("truth", Logic, and the IGUS, to name but three) without, it seems to us, resolving any of the problems in making sense of the consistent histories formalism. In this subsection we suggest a simpler interpretation, which we think achieves all that any other interpretation has achieved without adding conceptual frills or suggesting a resolution of unresolved problems. It is simply this: the world is described by precisely one history from one consistent set. Given the set, the history is chosen randomly according to the decoherence functional probabilities. We do not know which is the correct set, or how it should be characterized, or why it has the properties that it appears to have. If we are willing to assume an agreed list of historical events (in the terminology we have adopted from Omnès, actual facts), we can pin down some of the past sets of projections in the realized history. These, however, as we have seen, will not generally allow us to make useful unconditional predictions.

What can we make of the apparent persistence of quasiclassicality in this interpretation? It is either an illusion or a mystery—we have the impression that the realized history has been quasiclassical so far, and that is all there is to be said. As a practical matter, we should no doubt assume that the realized history will be quasiclassical in the future. Conditioned on this assumption, we can make probabilistic predictions with the same scientific content as those of the Copenhagen interpretation. Then, if we are to assume persistent future quasiclassicality, we may as well (since it hardly

[40] Such a criterion need not be deterministic: in principle, a probabilistic criterion defined by a measure on the space of consistent sets could do the job—if a sensible measure could be found.

deepens the mystery further) assume past quasiclassicality, too. However, this interpretation does not pretend any explanation of these facts.

This interpretation, it seems to us, is in practical terms equivalent to Griffiths' Logical interpretation, to the reasoning suggested by one of Gell-Mann and Hartle's lines of argument (though not their discussions of IGUSes), and to Omnès' approach (unless and until any interesting new definition of future true facts can be found). It has the advantage of clarity: indeed, it seems to us the clearest formulation of quantum mechanics, and the best understanding of that theory currently attainable, that involves neither a set selection hypothesis nor auxiliary variables. Either of these could in principle solve the problem it leaves open—the mystery of persistent quasiclassicality.

We have come full circle. As we have said already in our discussion of Omnès interpretation, our own view is that persistent quasiclassicality is such a basic feature of our experience, and failure to predict it leaves such a significant gap in our understanding of the world, that we feel that some argument, or some new interpretation, or some new formalism, that would amend the situation should be sought. Others may find the present interpretation perfectly acceptable as a fundamental theory. We hope, at least, that it will be recognized that it *does* leave a mystery which *could* possibly have an explanation.

6. PERSISTENCE AND PREDICTION

Much of our discussion on formal interpretations has involved the question of whether our persisting experience of a quasiclassical world can be predicted within a given interpretation. It seems to us *the* crucial question. If it were not for their failure on this score, both the Omnès and the many-histories interpretations would be quite acceptable; if Gell-Mann and Hartle's arguments, which lead to the conclusion that our experience *can* be predicted, did not rely crucially on premises about human experience, they would be complete. It also seems to us a perfectly legitimate scientific question, and we would no more happily accept a fundamental theory which cannot supply an answer than we would accept one which cannot explain celestial mechanics. Still, it is not a traditional question, and this may lead some to conclude that it is somehow beyond the scope of science, and not something about which down-to-earth equation-solvers need be concerned.

However, a little reflection shows that this is false. There are obvious quantum cosmological effects—the tunneling through to a region of true vacuum, propagating at the speed of light, for example—which would destroy (or at least discontinuously disrupt) persistent quasiclassicality. If a theory predicts that a vacuum bubble should engulf the solar system at some time after noon, Poisson-distributed with mean 1 hr, then, should we survive till midnight, we will certainly reject the theory.

Moreover, once nontrivial final density matrices are allowed, we need not rely on manufactured cosmological catastrophes to cause the breakdown of quasiclassicality. Quite the converse: the inevitable, consistent-set-independent breakdown of quasiclassicality becomes a generic feature. Suppose that we are given any model in which some consistent set $\mathscr{S} = (\rho_i \{\sigma_1, \sigma_2, \ldots\}, \{t_1, t_2, \ldots\})$ describes a persisting quasiclassical domain. (We have included the projection times, since they are necessary to establish quasiclassicality.) Choose any time t such that there is some r with $t_r < t < t_{r+1}$ and such that the truncated set $\mathscr{S}_t = (\rho_i, \{\sigma_1, \sigma_2, \ldots, \sigma_r\}, \{t_1, t_2, \ldots, t_r\})$ is nontrivially extended by σ_{r+1}. Then Lemma 4 implies that we can choose a final density matrix ρ_f such that

$$\mathscr{S}_t^f = (\rho_i, \rho_f, \{\sigma_1, \sigma_2, \ldots, \sigma_r\}, \{t_1, t_2, \ldots, t_r\})$$

is consistent and that the corresponding histories in $\mathscr{S}_t^f$ and $\mathscr{S}_t$ have the same probabilities, but such that

$$\mathscr{S}_f = (\rho_i, \rho_f, \{\sigma_1, \sigma_2, \ldots, \sigma_{r+1}\}, \{t_1, t_2, \ldots, t_r, t_{r+1}\})$$

is inconsistent.

In other words, if we have a theory describing our own quasiclassical domain up to the present time, and if the theory tells us that the results of some quantum experiment we are about to perform are genuinely unpredictable, then we can find another theory which reproduces the description up to the present time, but in which the standard quasiclassical description of the experimental results cannot be made. This, at least, is true if the original theory had pure initial density matrix and trivial final density matrix, since under these conditions Lemma 4 holds in both finite-dimensional and infinite-dimensional Hilbert spaces.[41]

Clearly, then, we cannot avoid discussions about the breakdown of quasiclassicality and its implications for the experience of IGUSes such as ourselves: if we want to understand general quantum cosmologies, then the formalism forces these questions on us. It seems, in fact, that among theories accurately describing the present quasiclassical world, nearly all will predict the eventual breakdown of quasiclassicality. In particular, any interpretation that discusses the experiences of IGUSes has to assume something about the experience of an IGUS whose lifespan is not describable quasiclassically beyond a certain point in time. This is quite

[41] No doubt similar results hold far more generally.

distinct from any conventional picture involving the quasiclassical destruction of the IGUS. No doubt the simplest assumption is that any experience requires a persistent quasiclassical description. But again, while this is a useful assumption if one wants to avoid awkward questions about the predictability of future experience, it is still an assumption. If one makes it, one must suppose that it will somehow follow from a general theory of experience, and this is pure guesswork at the moment. Even if we assume that when there are relevant quasiclassical variables they invariably correspond to quasiclassical experience, we can deduce nothing about the situation when there are no relevant quasiclassical variables.

It is also interesting to note that Lemma 4 implies that we cannot make any useful predictions, either unconditional or conditional, without a theory of the final density matrix. Neither consistency with the actual data nor the assignment of a particular probability to those data can be used as selection criteria for ρ_f: at any given point in time, a large family of final density matrices will do equally well. This does not appear to be a grave problem: there are apparently sensible choices—such as taking ρ_f to be I—which can be made on grounds of simplicity.

7. WHY BE CONSISTENT?

Since we consider in this paper several different versions and interpretations of the consistent histories formalism, we here discuss why any consistency criterion should be retained. We first mention recent interesting work by Goldstein and Page,[29] who use an earlier result of Page[48] to show that the Gell-Mann and Hartle consistency conditions can be considerably weakened while preserving the probability sum rules. Goldstein and Page point out that if $\mathscr{S}=(\rho_i, \rho_f, \{\sigma_j\})$, is a consistent set according to (2.7), then

$$\begin{aligned}\operatorname{Tr}(\rho_f P_n^{(i_n)} \cdots P_1^{(i_1)} \rho_i P_1^{(i_1)} \cdots P_n^{(i_n)}) &= \sum_{i'_1 \cdots i'_n} \operatorname{Tr}(\rho_f P_n^{(i'_n)} \cdots P_1^{(i'_1)} \rho_i P_1^{(i_1)} \cdots P_n^{(i_n)}) \\ &= \operatorname{Tr}(\rho_f \rho_i P_1^{(i_1)} \cdots P_n^{(i_n)}) \qquad (7.1)\end{aligned}$$

giving a simpler expression than (2.4) for probabilities:

$$p(i_1 \cdots i_n) = \operatorname{Tr}(\rho_f \rho_i P_1^{(i_1)} \cdots P_n^{(i_n)}) \qquad (7.2)$$

They go on to observe that, since (7.2) is linear, it automatically satisfies the sum rules (2.5) so long as the projections $P_r^{(i_r)}$ belong to projective decompositions σ_r, whether or not they form a consistent set. For general sets of σ_r, the expressions for $p(i_1 \cdots i_n)$ include negative quantities, and so cannot be given a probability interpretation. However, one can simply restrict to the sets for which all $p(i_1 \cdots i_n)$ are positive. Goldstein and Page refer to these as sets of linearly positive histories. As they point out, these sets have all the relevant mathematical properties of consistent sets, and give a broad extension of the consistent histories formalism.

Goldstein and Page further point out that if one uses their conditions there is no obvious mathematical reason to require ρ_i (or indeed ρ_f) to be positive semidefinite: whether they are or not, one can simply restrict attention to the sets whose histories have positive probabilities.[42] We include linear positivity among the class of possible consistency conditions when we refer below to consistent sets.

Now let us consider a more general possibility. We have seen that if a set $\mathscr{S}$ satisfies any of the various consistency conditions, then its histories obey the standard probability sum rules. This is certainly a very pleasing property, and gives an interesting selection criterion for sets of projective decompositions. But are there compelling reasons for restricting attention to such sets? After all, even for a general set $\mathscr{S}$ of which either the initial or final density matrix is unity, the weights (2.4) still have the fundamental properties of a probability distribution:

$$p(i_1,\dots,i_n) \geqslant 0 \qquad \text{and} \qquad \sum_{i_1,\dots,i_n} p(i_1,\dots,i_n) = 1 \tag{7.3}$$

Even if both ρ_i and ρ_f are nontrivial, one can in general simply renormalize the individual weights, which are all nonnegative, so that their sum is one: this fails only if all the weights are zero. Why not simply accept these weights as probabilities for the histories, and calculate those for coarse grained histories simply by summing: why not, in other words, accept the right-hand side of (2.5) as the correct coarse grained probability calculation and ignore the left-hand side entirely? Is it possible to make physical sense of such probabilities?

This last question was raised in Griffiths' original article:

> Another direction in which one might hope to extend the consistent histories approach is to find some physical interpretation of the weights ...when the events in question do *not* form a consistent history. One such interpretation is already implicit in Theorem 7 of section 5: aside from normalisation, the

[42] This loss of structure seems a little worrying, and suggests that the Goldstein–Page conditions, taken alone, are rather too loose: as Goldstein and Page themselves stress, further selection principles are needed. We prefer to consider the Gell-Mann–Hartle conditions for the pragmatic reason that they produce fewer consistent sets. However, as Goldstein[49] points out, even here things are complicated: it is not so clear that the Gell-Mann–Hartle conditions necessarily produce fewer fully fine grained or maximally refined sets.

> weights for histories belonging to a particular (inconsistent) family are the probabilities that the corresponding *consistent* histories *would* occur in a combined system which includes idealized measuring instruments which detect the different events in the original system at the appropriate times. However, this interpretation is neither simple nor a source of much intuition, given all the peculiarities associated with quantum measurements. Can one do better? (Ref. 1, Section 6.2.)

One can: it is perfectly possible to modify each of the interpretations outlined in Section 6 to allow for inconsistent sets. This causes no logical contradiction: it is trivial to extend the many-histories interpretation to inconsistent sets, and with a little more care the other interpretations also extend.

An immediate objection is that this weakens the formalism enormously, since now even the actual facts barely test a cosmological theory: any collection of actual facts can be incorporated in many different inconsistent sets, and we cannot even make use of probabilistic tests, since the probability assigned to the actual facts depends on the inconsistent set in which they are embedded. This would be the end of the story, were it not for the point which we have stressed—any satisfactory interpretation of the consistent histories formalism as a fundamental physical theory must eventually identify one particular consistent set in which to calculate.[43] This raises the question as to why we should insist that the choice be restricted to consistent sets. Could a theory of experience perhaps rely on a correspondence of mental states with operators in an inconsistent set? Or could those who would plump for a set selection criterion instead postulate that reality is described by some particular inconsistent set?

It has to be admitted that these are possibilities. There is no logical contradiction involved in using an inconsistent set in this way. The reason why we tend to reject these proposals is that inconsistent histories disagree with experimental data, unless they are fine tuned to be undetectably inconsistent. Thus the success of conventional quantum mechanics, which tells us that experimentally observed conditional probabilities relating various events can be calculated using only an initial density matrix, the Hamiltonian, and a description of those events in the language of projection operators, is strong evidence against generic inconsistent histories. We know, in practice, that we are able to ignore the parts of the universe outside the spacetime region of our experiments and to coarse grain the description of the experiment itself, and the consistent histories formalism guarantees that this coarse graining works. If nature were really described

[43] As we have seen, there are several different lines of argument which could be used to justify different possible choices.

by a history from a generic inconsistent set, with the probability interpretation above, the successes of reductionist science and the reproducibility of simple experiments would be inexplicably lucky accidents. For example, to calculate the rate of a chemical reaction in a generic inconsistent set, one would have to calculate the probability of every history in that set. Depending on the projections in the set, this could involve complicated calculations of the nuclear spin trajectories of the atoms involved, or of the local matter densities in distant nebulas, or of many grotesquely uninterpretable variables. There would be no reason to expect that the calculation for any pair of atoms could be performed according to Copenhagen (or any other simple) rules of thumb and, even if such a calculation could be performed, there would be no useful connection between the results for different pairs. In contrast, if the set were consistent, one could calculate the rate by considering only the coarse grained operators corresponding to various chemical densities inside the test tube and then applying the usual Copenhagen rules.

An inconsistent histories formalism is not absolutely excluded. It is still possible that the universe is described by an inconsistent set, the probabilities of whose histories, when summed to give the probabilities of events such as the measurement of the position of an electron on a screen in a double slit experiment, give the observed answers. If a theorist produces an otherwise elegant and successful quantum cosmological theory which implies that nature is described by a particular inconsistent set, we should want a very good explanation for the apparent consistency (in both senses) of our experimental data, but we should not reject the theory out of hand. However, assuming that one uses the decoherence functional to define probabilities for coarse grained histories, it seems to us that presently the case for some consistency criterion in quantum mechanics is compelling.

This, it should be stressed, is not a criticism of other history-like formulations of quantum mechanics[22, 23, 50] or of related theories.[51] There are sensible probability measures apart from the one defined by the decoherence functional and sensible sample spaces apart from spaces of coarse grained histories, and once one abandons one or both the comments above no longer apply.

8. CONCLUSIONS

The virtues of the consistent histories approach are worth reasserting. We have a natural mathematical criterion which is empirically supported and which identifies the physical content of quantum theory to be propositions about particular collections of events encoded in the consistent sets.

This gives a framework in which attempts to set out a quantum theory of the universe, and the problems inherent in this idea, can be sensibly discussed, and in which intrinsically quantum cosmological questions (such as the role of early anisotropies in seeding galaxy formation) can, at least in principle, be posed. There is a natural and attractive path integral version of the formalism, which is manifestly Lorentz invariant. The formalism gives a new way of thinking about the problems of quantum gravity and of studying measures of information in quantum theory. And, as Gell-Mann and Hartle have stressed, the abstract characterization of a quasiclassical domain within the formalism is a new and interesting research program. Many attempts have been made to find natural mathematical structures and interpretational postulates that allow one to use quantum theory to make statements about physics that go beyond the Copenhagen bounds. In our view, the consistent histories formalism is one of the most significant developments. Whether it will eventually form part of a theory with greater explanatory power than the Copenhagen interpretation remains in doubt —the possibility cannot be excluded, but there is as yet no positive evidence. Meanwhile, the consistent histories approach has the undoubted virtue of illustrating, more sharply than its predecessors, the problems inherent in quantum theory.

Advocates of the formalism thus have a good case. Perhaps we have even made one or two marginal additions. The suggestion that approximately consistent histories might play a key role has worried many who, like us, prefer the equations of their fundamental theories to hold exactly: it is reassuring to note that any interesting physical process can almost certainly be characterized by exactly consistent sets. The derivation of consistency criteria from the probability sum rules is an absolutely key insight, and is obviously mathematically very natural, but leaves some wondering whether the criteria are justified by any scientific principle. The answer is clear, and has been repeatedly pointed out by the authors of the formalism by illustrating the failure of inconsistent histories to reproduce the observed data in two-slit and other quantum experiments. But since obstinate critics can always say that such illustrations only exclude very particular inconsistent histories, a general discussion of inconsistency and reductionism may be helpful.

Nonetheless, we feel that the worrying features of the consistent histories approach deserve more attention. First, there are important points of principle to be resolved: at present, there is no clear agreement on how the actual facts should be selected, nor on which criterion (if any) should be used to deduce properties beyond the actual facts. Second, as we have seen, there are several possible attitudes which consistent historians can adopt to the past, and it would be good to see this question addressed more clearly. If one accepts the formalism as it stands, and introduces no historical actual facts, one seems to be led toward solipsism of the present, since it is unlikely that any past event can be described in every consistent set, and each consistent set gives an equally valid description of the past. Similarly, if Omnès' criterion for true propositions is adopted—and no other inferential rule has been suggested in the literature to date—the formalism almost certainly allows no ordinary quasiclassical deductions to be made unambiguously. It seems clear, practically speaking, that we cannot deduce, from the tides, from our perception of moonlight, or from any quasiclassical event on earth that the moon is in a quasiclassical orbit. And, as we have argued, although it is hard to find a rigorous demonstration, this does also seem likely to be a problem in principle: a quasiclassical description of events on earth should be consistent with an infinite number of pictures, in nearly all of which the moon does not behave quasiclassically. Since the most natural collection of actual facts appears to be our own experiences, and since the same inferential problem then arises for statements about our fellow creatures, the formalism also appears to lead to personal solipsism.

The prediction of the future is an even more pressing difficulty. Why do we continue to experience a quasiclassical world? The only answer we have found in the literature is that supplied by Gell-Mann and Hartle. We predict that we will experience a quasiclassical world because our experience will be described by certain decohering variables, which, for now, we simply assume to be quasiclassical ones coupled to a quasiclassical environment, and which a to-be-found theory of experience will identify as the fundamentally correct ones among all the possibilities offered by the formalism. We must assume that such a theory of experience can be found, since we *do* experience a quasiclassical world. We put the argument starkly, but we believe the translation is accurate. It is a coherent position, and anyone who holds it will feel happy developing the formalism along the current lines. No one, though, will mistake it for an ultimately satisfactory answer to the question. If it is correct, there seems to be no hope of our understanding the answer in the foreseeable future. We are clearly a very long way from a theory of experience, and there is no guarantee that such a theory, of the type needed, will or can ever be found.

There is an alternative which cuts through all these problems. It is to accept, once and for all, that quantum theory is not sufficient to describe the world, and that it should be augmented by a further axiom which takes the form of a selection principle. The consistent histories formalism has taught us that there are infinitely many incompatible descriptions of the world within quantum mechanics. Perhaps some simple criterion can be found to pick out one of these descriptions, by selecting one particular consistent set. Such a criterion should explain persistent quasiclassicality, not as a consequence of our own biased perceptions, but as a deducible fact; it would remove all solipsist tendencies from the theory; it would restore definiteness to statements about the past. To be remotely persuasive, of course, any such criterion would have to be simple and elegant: there is no point in merely setting out a characterization of the world as we see it.

Consistent historians will certainly agree that this is an interesting program. Omnès has recently set out a speculative proposal along precisely these lines.[52] As others have noted,[19, 43] Gell-Mann and Hartle's program aimed at characterizing the notion of a quasiclassical domain could equally be viewed as a search for a set selection criterion, though this is not their motivation. Moreover, Gell-Mann and Hartle's explorations of stronger decoherence criteria (such as M-decoherence) and of alternative set selection rules (such as replacing projections by sums over fine grained position space histories) point out ways in which the set selection problem can at least be diminished.

Nonetheless, the search for a selection principle which will pick out a unique consistent set is a line of development that tends to be placed firmly in the category of unorthodox proposals occupied by alternative dynamical theories such as those of Ghirardi *et al.*[51] or Gisin[53] and Percival[54] and by auxiliary variables theories.[22, 23] Interpretations of the consistent histories formalism which do not explicitly rely on set selection, on the other hand, have been regarded as setting out *the* correct development of quantum theory and defining the natural null hypothesis given the present experimental data. If there is one single point which we would wish to emerge from this paper, it is that this view is indefensible. The apparent persistence of quasiclassicality is a central problem for the consistent historians. Either it is fundamentally bound up with the problem of a theory of experience or the two problems are separate. In the former case, there is presently no hope of a solution. In the latter case, there are many interesting ideas which can be explored. In the former case, we are led to solipsisms; in the latter case, we can hope to recover historical and quasiclassical inferences in an entirely straightforward way. There is no methodological or scientific reason to prefer the former position: it is tenable, but it certainly occupies no distinguished high ground. Although alternative dynamical theories are very interesting, it remains a sensible null hypothesis to suppose that the dynamical principles of quantum theory are correct. However, having made that supposition, and under the assumption that the consistent histories formalism defines the correct framework for making sense of quantum theory, it is no less wild to hypothesize that the formalism should be augmented, at a fundamental level, by a theory of experience than that it needs a set selection criterion. If quantum dynamics is fundamentally correct, and the consistent histories formalism provides the correct physical interpretation, then it seems that one hypothesis or the other must be right: at any rate, no others have yet been suggested. Both hypotheses go beyond quantum theory as it is currently understood; neither is strongly supported by current science.

One sometimes encounters the following objection. Suppose that a good set selection hypothesis were found. Would we not still ultimately need a theory of experience? If so, what would have we gained? The answer, of course, is that a theory of experience would still be sought—consciousness is a deep and fascinating problem in its own right. What would be gained, though, by the selection of a quasiclassical history—*the* quasiclassical history of the world—and the assumption of psychophysical parallelism is the knowledge that any theory of experience would necessarily be formulated in terms of quasiclassical variables. The task of forming such a theory would be simplified enormously over that of forming a theory in the unaugmented consistent histories formalism (which is not to say that it would be easy even were the variables known, it just speaks to the extraordinary difficulties of forming the theory otherwise). Then, making the weak assumption that whatever the theory of experience is, it must be the case that it would describe our experience (largely) mirroring real physical events—i.e., we *see* a table (usually) because there *is* a table there—we would not need the details of the theory of experience to be able to predict the appearance to us of a quasiclassical world. This last assumption amounts to no more than the standard and well-supported hypothesis that the quasiclassical variables in our brains are correlated, by mechanisms familiar to neuroscientists, with those in the external world.

Another objection occasionally raised is that it remains conceivable that the operators relevant to our experience are *not* quasiclassical, perhaps because they describe truly microscopic events at the subneuronal level.[55] There is no neuroscientific evidence for this hypothesis, and even if it were correct, the dichotomy would remain: either a theory of experience or a theory of reality would be needed, though a purely quasiclassical set selection rule would of course be excluded.

We would like to conclude by drawing attention to d'Espagnat's admirable critique,[17] which covers the earlier papers of Griffiths and Omnès. D'Espagnat examines the question of whether the consistent histories formalism allows a realistically interpretable local formulation of quantum mechanics, and finds that it cannot, despite the impression which might be gained by superficial readings. To quote his conclusion:

> It must be granted that several of the interpretative comments the quoted authors make of their theories stand quite at odds with the main conclusions reached here. Indeed, while these authors do not actually *say* their theories are realistically interpretable, they somehow give at various places the impression that they *mean* just precisely that. Such a somewhat disquieting state of affairs seems to indicate that we physicists still have efforts to make before we succeed in imparting to the *words* we use (and especially to the nonoperationally defined ones) a strictness of meaning comparable with the strictness of our mathematical manipulations. This will presumably only be achieved when we have convinced ourselves that it is impossible to freely switch between an ontological and a purely operationalist usage of such words as 'have,' 'is,' 'objective,' and the rest.

We find ourselves very much in sympathy with d'Espagnat.

ACKNOWLEDGMENTS

We thank the Aspen Institute for their hospitality in the formative stages of this work. F. D. was supported by DOE and NASA grant NAGW-81 at Fermilab and by NSF grant no. PHY-9008502; A. K. was supported by a Royal Society University Research Fellowship. It is a pleasure to thank Murray Gell-Mann, Bob Griffiths, and Roland Omnès for invaluable discussions of their work. We would particularly like to thank Jim Hartle for taking the time to explain his ideas to us and for his generous support and constructive criticisms in many discussions over the last two years. We are very grateful to Trevor Samols for many helpful discussions and to Shelly Goldstein for many thoughtful criticisms and insights. We have also been greatly helped by discussions with Andy Albrecht, Arley Anderson, Jeremy Butterfield, Matthew Donald, Jerome Gauntlett, Nico Giulini, Peter Goddard, Jonathan Halliwell, Chris Isham, Achim Kempf, Raymond LaFlamme, Klaas Landsman, Noah Linden, Jim McElwaine, Don Page, Michael Redhead, Lee Smolin, Rafael Sorkin, John Taylor, Bill Unruh, Gérard Watts, and Wojciech Zurek.

REFERENCES

1. R. B. Griffiths, *J. Stat. Phys.* **36**:219 (1984).
2. R. B. Griffiths, *Found. Phys.* **23**:1601 (1993).
3. R. Omnès, *J. Stat. Phys.* **53**:893, 933, 957 (1988); **57**:357 (1989).
4. R. Omnès, *Rev. Mod. Phys.* **64**:339 (1992).
5. M. Gell-Mann and J. B. Hartle, In *Complexity, Entropy, and the Physics of Information*, W. Zurek, ed. (Addison-Wesley, Reading, Massachusetts, 1990).
6. M. Gell-Mann and J. B. Hartle, In *Proceedings of the 3rd International Symposium on the Foundations of Quantum Mechanics in the Light of New Technology*, S. Kobayashi, H. Ezawa, Y. Murayama, and S. Nomura, eds. (Physical Society of Japan, Tokyo, 1990).

7. M. Gell-Mann and J. B. Hartle, In *Proceedings of the 25th International Conference on High Energy Physics, Singapore, August 2-8, 1990*, K. K. Phua and Y. Yamaguchi, eds. (South East Asia Theoretical Physics Association and Physical Society of Japan, distributed by World Scientific, Singapore, 1990).
8. M. Gell-Mann and J. B. Hartle, In *Proceedings of the NATO Workshop on the Physical Origins of Time Asymmetry, Mazagón, Spain, September 30-October 4, 1991*, J. Halliwell, J. Pérez-Mercader, and W. Zurek, eds. (Cambridge University Press, Cambridge, 1994).
9. M. Gell-Mann and J. B. Hartle, *Phys. Rev. D* **47**:3345 (1993).
10. F. Dowker and A. Kent, *Phys. Rev. Lett.* **75**:3038 (1995).
11. B. DeWitt and R. N. Graham, eds., *The Many Worlds Interpretation of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1973).
12. J. S. Bell, The measurement theory of Everett and de Broglie's pilot wave, in *Quantum Mechanics, Determinism, Causality and Particles*, M. Flato *et al.*, eds. (Reidel, Dordrecht, 1976); reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1987).
13. H. Stein, *Noûs* **18**:635 (1984).
14. A. Kent, *Int. J. Mod. Phys. A* **5**:1745 (1990).
15. J. S. Bell, Quantum mechanics for cosmologists, in *Quantum Gravity 2*, C. Isham, R. Penrose, and D. Sciama, eds. (Clarendon Press, Oxford, 1981), pp. 611–637.
16. J. B. Hartle, In *Quantum Cosmology and Baby Universes*, S. Coleman, J. Hartle, T. Piran, and S. Weinberg, eds. (World Scientific, Singapore, 1991).
17. B. d'Espagnat, *J. Stat. Phys.* **56**:747 (1989).
18. W. Zurek, *Prog. Theor. Phys.* **89**:281 (1993).
19. D. Dürr, S. Goldstein, and N. Zanghí, *J. Stat. Phys.* **67**:843 (1992).
20. A. Albrecht, *Phys. Rev. D* **46**:5504 (1992); **48**: 3768 (1993).
21. J. Paz and W. Zurek, *Phys. Rev. D* **48**:2728 (1993).
22. D. Bohm, *Phys. Rev.* **85**:166 (1952).
23. T. M. Samols, A stochastic model of a quantum field theory, Cambridge preprint DAMTP/94-39; *J. Stat. Phys.* to appear.
24. J. Cushing, A. Fine, and S. Goldstein, eds., *Bohmian Mechanics and Quantum Theory: An Appraisal* (Kluwer, Dordrecht, to be published).
25. Y. Aharonov, P. Bergmann, and J. Lebovitz, *Phys. Rev. B* **134**:1410 (1964).
26. J. B. Hartle, *Phys. Rev. D* **44**:3173 (1991).
27. C. J. Isham, In *Integrable Systems, Quantum Groups and Quantum Field Theories*, L. A. Ibort and M. A. Rodriguez (eds.) (Kluwer, London, 1993); C. J. Isham, *J. Math. Phys.* **23**:2157 (1994); C. J. Isham and N. Linden, *J. Math. Phys.* **35**:5452 (1994).
28. M. Gell-Mann and J. B. Hartle, Equivalent sets of histories and multiple quasiclassical domains, Preprint UCSBTH-94-09, gr-qc/9404013, submitted to gr-qc 8 April 1994.
29. S. Goldstein and D. Page, *Phys. Rev. Lett.* **74**:3715 (1995).
30. J. S. Bell, *Physics* **1**:195 (1964); *Rev. Mod. Phys.* **38**:447 (1966).
31. D. Bohm, *Quantum Theory* (Prentice-Hall, Englewood Cliffs, New Jersey, 1951), Chapter 22.
32. E. Joos and H. D. Zeh, *Z. Phys. B* **59**:2 (1985).
33. W. Zurek, *Phys. Rev. D* **24**:1516 (1981); **26**:1862 (1982).
34. A. Caldeira and A. Leggett, *Physica* **121A**:587 (1983).
35. J. McElwaine, Approximate and exact consistency of histories, University of Cambridge preprint DAMTP/95-32, grant-ph/9506034, Submitted to *Phys. Rev. A*.
36. P. H. Gosse, *Omphalos: An Attempt to Untie the Geological Knot* (1857).
37. R. Omnès, Private communication.

38. R. Griffiths, Private communication.
39. S. Saunders, The quantum block universe, Harvard Department of Philosophy preprint (1992); Decoherence, relative states, and evolutionary adaptation, Harvard Department of Philosophy preprint (1993).
40. H. Everett, *Rev. Mod. Phys.* **29**:454 (1957).
41. M. Gell-Mann and J. B. Hartle, Equivalent sets of histories and multiple quasiclassical domains, preprint UCSBTH-94-09, revised version as of 26 April 1995.
42. M. Gell-Mann, *The Quark and the Jaguar* (Little, Brown and Co., London, 1994).
43. J. B. Hartle, Private communication.
44. A. Whitaker, *Einstein, Bohr and the quantum world*, to be published.
45. J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955).
46. E. P. Wigner, Remarks on the mind–body question, in *The Scientist Speculates*, I. J. Good ed. (Heinemann, London, 1961). pp. 284–302.
47. R. Penrose, The nature of space and time, Isaac Newton Institute debate with S. W. Hawking (May 1994).
48. D. Page, *Phys. Rev. Lett.* **70**:4034 (1993).
49. S. Goldstein, Private communication.
50. R. D. Sorkin, Quantum mechanics as quantum measure theory, Syracuse preprint SU-GP-93-12-1, gr-qc/9401003.
51. G. Ghirardi, A. Rimini, and T. Weber, *Phys. Rev. D* **34**:470 (1986).
52. R. Omnès, *Phys. Lett. A* **187**:26 (1994).
53. N. Gisin, *Helv. Phys. Act.* **62**:363 (1989).
54. I. Percival, *Proc. Roy. Soc. Lond. Ser. A* **447**:189 (1994).
55. R. Penrose, *Shadows of the Mind* (Oxford University Press, Oxford, 1994).