Resource Material for Promoting the Bayesian View of Everything

Prepared by Carlton M. Caves

Notes mainly for the use of CMC, Christopher A. Fuchs, and Rüdiger Schack, although anyone is welcome to make use of the notes and even buy in, in whole or in part, if he/she finds the notes convincing. Not everything here is endorsed by Fuchs and Schack, but it is all part of a joint program aimed at producing a Bayesian synthesis. Fuchs, in particular, probably disagrees with every statement about Hamiltonians being the objective part of quantum theory. Nonetheless, nothing should be attributed to me alone until Fuchs and Schack have disavowed it.

Version 2.2: 2001 August 15 (plain TEX files of V1, V2, and V2.1 available on request)

1. Introduction. Science involves investigating the properties of the real world and processing and using what we know about the world. To make progress, it is crucial to separate what's really out there (ontology) from what we know about what's really out there (epistemology). The language for dealing with our (necessarily incomplete) knowledge of the world is Bayesian probability theory, which holds that probabilities are subjective, based on what we know. Here I contrast the Bayesian view of probabilities with other interpretations and consider two natural applications of Bayesian probabilities in physics: statistical physics and quantum theory.

Because physicists believe their science is the most fundamental, they have an ingrained tendency to attribute ontological status to the mathematical objects in their theories, including probabilities. Statistical physics provides a cautionary example of the hazards of this tendency: it leads to the notion that thermodynamic entropy is an objective quantity and thus to fruitless efforts to derive the Second Law of Thermodynamics from the time-symmetric laws of physics. It is now well established—though still not accepted by many practitioners—that entropy is a subjective quantity, based on what one knows. This leads to effortless derivations of the Second Law and also to a deep understanding of the operation of intelligent agents (Maxwell demons) whose objective is to circumvent the Second Law. Quantum theory is tougher: the cut between ontology and epistemology is notoriously hard to identify, because of the intrinsic indeterminism of quantum mechanics. For this reason, we believe that it is in quantum theory that a consistent application of Bayesian probabilities has the most to offer.

2. Interpretations of probabilities.

a. *Empirical (actual) frequentism*, i.e., defining probability as the frequency of occurrence in an *actual* sequence of trials or *real* ensemble. There are immediate problems with this approach. How does one deal with probabilities for single cases, e.g., Laplace's determination of the mass of Saturn or the openness vs. closedness of the Universe, where a real ensemble doesn't exist? Even in cases where a real ensemble can be identified, why should the probability for an individual member depend on what other members of the ensemble do? Moreover, there are serious problems with finite ensembles: What about irrational probabilities? How big an ensemble does one need to get *the* probability? These problems force one to infinite ensembles, but they present insuperable problems: there being no real infinite ensembles, this approach floats free from its initially attractive empiricism, and the frequency is ultimately undefinable (ordering problem).

b. *Mathematical frequentism (law of large numbers)*, i.e., defining probability as the limiting occurrence frequency in an imaginary (hypothetical) sequence (an ensemble). This is the mathematical way of trying to do frequentism; the emphasis here is entirely on mathematical rigor, with no reference to empiricism (i.e., actual ensembles). Despite its rigor—more precisely, because of its rigor—this approach has fundamental flaws. Because the infinite ensemble is purely hypothetical, one must introduce additional *mathematical* structure to characterize the ensemble. What is introduced, right from the start, is the notion of probability. One uses the notion of probability for each member of the ensemble and the notion of the statistical independence of different members of the ensemble to provide the mathematical structure of the i.i.d. (distribution that *i*ndependent and *i*dentically *d*istributed). Thus this approach is circular since it relies on the notion of probability to define the limiting frequency, which is supposed to define probability.

   Moreover, what is proved is that the "right frequency" occurs with probability-one in the infinite ensemble. This result cannot be interpreted without reference to probabilities. The frequentist hopefully asserts that probability-one means certainty, which would allow him to escape reference to probabilities in interpreting the result. Yet this identification can't be justified. Though probability-one does mean certainty for finite sets, this strict connection can't be maintained in the infinite limit. To see this, imagine deleting a set of measure zero from the right-frequency set; the modified set still has probability-one, so are we to think that the deleted sequences are certain not to occur? This is not very reasonable, since any particular sequence with the "right frequency" could be deleted in this way, so we would be forced to conclude that all sequences are certain not to occur. (Another way of saying the same thing is that each infinite sequence has probability zero; should one then conclude that each sequence is certain not to occur?) What this demonstrates is that the only way to *interpret* what probability-one means in the infinite limit is to have already at hand a notion of the meaning of probability and to apply that notion to a limit as the sequence becomes infinitely long.

   Finally, suppose one could define probability as a limiting frequency in a hypothetical ensemble. One would still be left without the ability to make any quantitative statements about finite (real) ensembles. One would not be able to say, for example, that in 1,000 tosses of a fair coin, the number of heads will be near 500 with high probability, for this probability would mean nothing without referring it to yet another hypothetical infinite ensemble, each member of which is a sequence of 1,000 tosses.

One lesson taught by the above two approaches is that it is very important in any discussion of probabilities to determine whether one is discussing real or hypothetical ensembles, because the problems with the two approaches are different. The problem with real ensembles is that usually we don't have them and we certainly don't want to be required to have them. The problem with hypothetical ensembles as fundamental, rather than derived objects is that one must use the probability measure provided by the i.i.d. to give structure to the ensemble, so one can hardly be using the ensemble to define probabilities.

A more important lesson is that trying to define probabilities in terms of frequencies is hopeless. One ends up referring probabilities to an unempirical limit that is undefinable and uninterpretable unless one already has the structure of probability theory in hand. One is trying to define probabilities in terms of concepts derived from probabilities. To make progress, one must get serious about defining probability for a single trial, since we evidently use probabilities in that situation. The leading objectivist candidate is propensity (objective chance).

c. *Propensity or objective chance*, i.e., a probability that is asserted to be an objective property of a physical system or situation. Yet how can propensity be an objective property of a coin when the chance of heads clearly depends on the method of tossing? The Bayesian has no trouble admitting that the probability assignment depends on what one knows. This could include what one knows about the physical properties of the coin, the method of tossing it, the properties of the medium through which the coin passes, the properties of the surface on which it lands, and any other relevant factors, or it could mean knowing nothing about any of these or anything else, except that the coin has two faces. Any rational observer, asked to make a bet on the toss of a coin, can see that the probability assignment on which he bases his bet ought to depend on what he knows about all these factors. The propensitist is thus backed into a corner: which of the Bayesian probabilities deserves to be the propensity, and what should one do in those situations where the propensity is not the best probability assignment?

   From a Bayesian perspective, the question of what propensitists are talking about is easy to answer and depends on what one knows about a sequence of trials. A propensity is a single-trial probability that arises from a robust prior on probabilities, or in other words, a situation where the probabilities on the multi-trial hypothesis space are well approximated by an i.i.d. for a very large number of trials, or in yet other words, where one's probabilistic predictions for future trials change from the original i.i.d. predictions based on the propensity only after one has gathered data from a very large number of trials.

d. *Principal principle*, i.e., use ignorance probabilities till objective ones are identified. This is a desperate retreat in the face of the logic of Bayesianism, hoping against hope that some basis for objective probabilities will eventually emerge even as one admits that most probabilities are Bayesian.

   In the example of the coin, if one knows everything about the coin and its environment, then given that a classical, deterministic description is sufficient, there is no chance at all; the outcome of the toss can be predicted with certainty using the laws of classical mechanics. This conclusion is quite general; as Giere (1973a) notes, there are no objective probabilities, except 0 or 1, in a realistic/deterministic world. All probabilities in a realistic/deterministic world are Bayesian probabilities, reflecting ignorance of the exact conditions that hold in such a world.

Despite the inescapability of this conclusion, cunning attempts are made to get around it. Such attempts should always be examined by asking what happens when one has complete information. For example, it is asserted that for two successive coin flips the probability for head-tail is *objectively* the same as the probability for tail-head. These equal probabilities are then used to build up "objective" probabilities of any magnitude. There has to be something wrong with the conclusion that head-tail and tail-head probabilities are objectively the same, independent of what one knows, for if one knows everything about the two coin tosses, then one can say with certainty what the outcomes of the tosses will be; the head-tail and tail-head probabilities are the same only if both are zero. So what gives? The equality of the head-tail and tail-head probabilities follows from assuming the exchangeability of the two tosses, a condition that is violated in the case of complete, but opposite knowledge for the two tosses. Exchangeability is clearly a part of the state of knowledge on which a Bayesian bases a probability assignment; it is not an objective property of the world. The equal "objective" probabilities are thus an example of assigning Bayesian probabilities based on what one knows, in this case, that one's state of knowledge is symmetric under exchange of successive tosses.

e. *The Bayesian view*, i.e., that probabilities are not a state of the world, but are based on one's state of knowledge. A simple, but compelling argument that probabilities are not objective properties is that one cannot determine them by addressing the alternatives (no multiple trials here: we've seen that that's a dead end, because it requires another probability assignment on the bigger multi-trial hypothesis space); to find out what probabilities have been assigned, you have to ask the assigner.

The mathematical foundations for the Bayesian view come from (i) the Dutch-book argument, which shows that if probabilities are regarded as betting odds that determine one's willingness to place bets, consistency in placing the bets yields the probability rules; (ii) the Cox analysis, which shows that if probabilities are measures of credible belief, then consistency with deductive logic gives the probability rules; (iii) decision theory, which is the application of Bayesian, single-case probabilities to rational decision making (a generalization of betting odds); and (iv) the de Finetti representation theorem (Caves, Fuchs, and Schack 2001a), which shows that probabilities on probabilities are in one-to-one correspondence with exchangeable probability assignments on infinite sequences. The de Finetti theorem banishes the concept of an unknown probability in favor of a primary probability assignment on infinite sequences.

3. Application of the Bayesian view to statistical physics.

a. *Entropy is subjective*, defined relative to a probability assignment that is based on what one knows about a system. The subjectivity of entropy lies in the fact that it is the "missing information" required to specify a system's microstate. The relation to physics is that entropy—more precisely, free energy—quantifies available work: each bit of missing information reduces the available work by $k_B T \ln 2$. The subjectivity of entropy is natural in this context, for different observers, knowing different things about a system and thus assigning different entropies, will devise different procedures that extract different amounts of work from the system. Anyone who thinks entropy is objective should be asked how to extract work from a Szilard engine before observing which side the molecule occupies; there is no chance of success in such an endeavor, of course, because success would violate the Second Law.

b. *How does entropy change?* Since entropy represents the amount of information missing toward a maximal description of a system, it changes only when information about the system is obtained or discarded. Entropy does not change under Hamiltonian dynamical evolution, because Hamiltonian evolution is the rule for updating maximal information.

   Thermodynamics is the science of macrostates, i.e., states specified by system probability distributions defined by consistently discarding all information about a system—most importantly, information about a system's past—except information about a few macroscopic parameters that determine the system's macroscopic behavior. Jaynes's (1957a,1957b) derivation of the Second Law is based on discarding everything but the values of a few macroscopic parameters and then maximizing the entropy relative to the constraints imposed by the macroscopic parameters. Indeed, like Jaynes's derivation, *all* derivations of an increase in entropy proceed by discarding information about the system: (i) Boltzmann's H-theorem is based on discarding information about two-particle correlations; (ii) projection-operator techniques throw away information about the microscopic probability distribution, replacing it with a coarse-grained distribution; (iii) master equations are based on discarding information about the correlation between the system and a heat bath or environment.

c. *Entropy decreases when one observes a system.* Though this increases the available work, the increase is offset by the energy required to erase the record of the observation. This *Landauer erasure cost*, amounting to $k_B T \ln 2$ per bit, is required, from the inside view, for consistency with the Second Law and is, from the outside view, simply the free energy that must be dissipated to reduce the entropy of a demon/memory as it is returned to a standard state.

   The Landauer erasure cost leads to the notion of *total entropy*, the sum of the system entropy and the information required to specify the system state. When a demon/memory observes the system and stores a record of its observation, the total entropy does not change; likewise, if the demon/memory discards its record, the total entropy again remains constant. Thus the process of observing and discarding information can be done at no thermodynamic cost. Since the erasure requires the Landauer cost, however, there is no overall thermodynamic cost only if the demon/memory extracts the work made available by its observation before discarding its record.

   Under Hamiltonian evolution the total entropy is nearly constant: the system entropy remains exactly constant, but the information needed to specify the system state increases by the amount of information, $\log t$, needed to give the time t in some appropriate units. Another way of saying this is that the evolved state does not become much more complex than the initial state as a system evolves under Hamiltonian evolution, because the information needed to specify the evolved state consists of the initial state and Hamiltonian—these are background information—and the time. The near constancy of the total entropy means that one does not lose available work as a system evolves. One way to extract the work is to time reverse the system—this is usually difficult to do physically (but note spin echo), but is easy to describe and thus is simple algorithmically—and then use the $\log t$ bits of information to tell one how much time t to wait before extracting the work from the initial state.

d. *Why does one discard any information at all*, since it always has a cost in available work? In particular, why does one discard information about a system's past? The standard explanations fall into three classes: (i) *Coarse graining.* One discards fine-grained information because it is irrelevant to the future behavior of the coarse-grained macroscopic parameters of interest; (ii) *Complexity of evolved state.* The evolved state is so complex that it requires an increasing amount of information about the initial conditions keep track of it; (iii) *Leakage of information to an environment.* The information about the system's initial state leaks into an environment, where it is lost. All of these putative justifications deserve discussion.

(i) *Coarse graining.* This is a powerfully good explanation, which underlies all methods for justifying the increase of entropy in isolated systems. It leads to an entropy increase in the following way: divide phase space into coarse-grained cells j, with phase-space volume Γ_j; at each coarse-grained time $t_n \equiv n\tau$, where τ is a macroscopic coarse-graining interval, replace the evolved phase-space distribution from the previous step with a coarse-grained one, smeared out to be uniform on each coarse-grained cell, with probabilities $p_j(n)$; the resulting coarse-grained (Gibbs) entropy at the nth time step is

$$S(n) = k_B \left(-\sum_j p_j(n) \log p_j(n) + \sum_j p_j(n) \log \Gamma_j \right) ,$$

where the first term is the information needed to specify which cell the system is in and the second term is the average over cells of the information needed to specify a point (microstate) within the cell. This entropy increases monotonically as information about the fine-grained distribution is discarded and, for a mixing system, approaches the microcanonical entropy on the accessible region of phase space. The chief question to be investigated in this approach is whether the behavior of the macroscopic parameters of interest is indeed insensitive to the coarse graining. Though this is a very reasonable approach, it begs the deeper question of why entropy increases, because it discards information for convenience in describing the behavior of macroscopic parameters without addressing why one is willing to accept the corresponding decrease in available work.

(ii) *Complexity of evolved state.* Even though this explanation is the most widely accepted one for isolated systems undergoing Hamiltonian evolution, it is wholly misguided. These days it is usually phrased in terms of chaotic dynamics: to predict the trajectory of a chaotic system requires an exponentially increasing amount of information about the system's initial conditions. Where this explanation goes wrong is in thinking that trajectories are the relevant concept for entropy. Entropy has to do with phase-space distributions, made up of many trajectories, not with individual trajectories. It does not require an exponentially increasing amount of information to predict the evolved phase-space distribution. Indeed, as we have already seen, the complexity of phase-space distributions increases hardly at all under Hamiltonian evolution. Another way of saying this is that phase-space distributions evolve according to the Liouville equation, which is linear and preserves overlaps between distributions. A small error in the initial distribution remains a small error in the evolved distribution.

(iii) *Leakage of information to an environment.* This is the best explanation for the increase of entropy, because *no* system that we have access to is completely isolated from its environment. Nonetheless, it also begs the deepest question, for it simply posits, with no ultimate justification, that the environment is somehow so complicated that information which leaks into it is unrecoverable. The ultimate question is avoided: why does one discard information that leaks into the environment? This leads to the Schack-Caves hypersensitivity-to-perturbation program (Caves 1994a, Caves and Schack 1997a): a system is *hypersensitive to perturbation* when the environmental information required to reduce the system entropy far exceeds the entropy reduction; when a system is hypersensitive to perturbation, the discarding of information in the environment can be justified on strict grounds of relative thermodynamic cost. Schack and Caves have shown that classical chaotic (mixing) systems have an *exponential* hypersensitivity to perturbation, in which the environmental information required to purchase a system entropy reduction increases exponentially with time, and they have accumulated numerical evidence that quantum versions of classically chaotic systems display a similar exponential hypersensitivity to perturbation. Perhaps the most succinct way to describe the Schack-Caves program provides a way to rescue explanation (ii): the evolved phase-space distribution (or quantum state) is not itself algorithmically complex, but for chaotic (mixing) systems it lies close to highly complex distributions into which a perturbing environment can easily push it.

e. *Ergodicity.* Though ergodicity has nothing to do with the approach to thermodynamic equilibrium, since the exploration of all of accessible phase space occurs on too long a time scale to be relevant for the approach to equilibrium, it nonetheless plays a central role in the application of Bayesian probabilities to dynamical systems. If one knows only the energy of a system and that it is constrained by certain external parameters, then one should assign a time-invariant distribution since one's state of knowledge is time-invariant. Ergodicity and the conservation of phase-space volume under Hamiltonian evolution imply that the only time-invariant distribution is the microcanonical distribution.

f. *Why does the canonical distribution lead to predictions of frequencies*, such as the angular distribution of effusion of gas through a small hole, when the Bayesian view does not assert any necessary connection between probabilities and frequencies? Imagine drawing sequences from a probability distribution $\mathbf{p}$. In N trials the number of sequences whose probability exceeds some threshold is proportional to $e^{NH(\mathbf{p})}$. Thus, of all the probability distributions consistent with the mean-value constraints, the canonical (MAXENT) distribution is the one that generates the most high-probability sequences; indeed, for many trials, the high-probability sequences generated by the

MAXENT distribution are essentially *all* the sequences that are consistent with the constraints. If the constraints that went into the MAXENT distribution are truly the only thing you know about the system, then an i.i.d. based on the MAXENT distribution is your most unprejudiced way of predicting frequencies.

This idea can be used to put MAXENT in a broader context, where instead of assigning the MAXENT distribution, one weights the probability on probabilities by the number of high-probability sequences in some number of trials that one contemplates doing; i.e., one assigns an exchangeable multi-system distribution whose probability on probabilities is chosen to be proportional to $e^{NH(\mathbf{p})}$ times delta functions that enforce the mean-value constraints. In this formulation, the parameter N characterizes one's confidence in the MAXENT predictions, which can be overridden by conflicting data in a sufficiently large number of trials.

g. *The Lebowitz program* (Lebowitz 1999a). Lebowitz explains the increase of entropy using what he characterizes as a modern presentation of Boltzmann's approach. He divides phase space into coarse-grained cells j, with volume Γ_j. The cells are defined by the values of some macroscopic parameters and are not of equal size, there being one cell that dominates, the one that has equilibrium values for the macroscopic parameters. He associates with each microstate (point) within cell j a "Boltzmann entropy" $S_B = \log \Gamma_j$. The dominant cell has a Boltzmann entropy that approximates the thermodynamic equilibrium entropy. His argument is that if the system is initially confined to a cell M, defined by some macroscopic constraints, then "typical" initial conditions within that cell end up after a short while in the dominant cell, thus leading to the entropy given by thermodynamic equilibrium.

Probabilities are introduced to characterize "typical" initial conditions—i.e., almost all initial conditions selected randomly from an initial probability distribution that is uniform on M with respect to the standard phase-space measure. Even though it is recognized that probabilities are absolutely essential for this purpose—"any meaningful statement about probable or improbable behavior of a physical system has to refer to some agreed upon measure (probability distribution)"—they are introduced apologetically, with a defensive tone, because the system is at all times actually in a particular microstate, or as Boltzmann is quoted as putting it, "The applicability of probability theory to a particular case cannot of course be proved rigorously." Indeed, the probabilities are never spelled out precisely—they are " a measure which assigns (at least approximately) equal weights to the different microstates consistent with the 'initial' macrostate M"—and are never given a symbol, for to do so would give too much legitimacy to what are obviously Bayesian probabilities based on one's lack of knowledge of the exact initial conditions within M.

A Bayesian sees through this smokescreen immediately. The initial uniform probability distribution does apply to an individual system and is justified on the grounds stated above. The Lebowitz program is a species of coarse graining [(i) above], the problem being that the refusal to use Bayesian probabilities renders the coarse graining unnecessarily mysterious and the explanation of what is going on nearly nonsensical. The Boltzmann entropy of a microstate has no physical meaning, in contrast to the

Gibbs entropy of a distribution, which quantifies the amount of available work. Indeed, the Boltzmann entropy is not a property of the microstate at all, but a property of the coarse graining. For a particular trajectory, the Boltzmann entropy bobs up and down like a cork as the trajectory moves from one coarse-grained cell to another, but this bobbing up and down has no physical meaning whatsoever. To take advantage of a decrease in the Boltzmann entropy, say to extract additional work, you would need to know what trajectory the system is on, but you don't know that and if you did know it, you wouldn't be messing around with the coarse graining, because you would know exactly what the system is doing at all times. The Lebowitz program is a perfect example of the contortions that come from insisting that physics is only about what's really out there, when it is evident here that what one can do depends on what you know about what's really happening.

A Bayesian, asked to fix the Lebowitz program, might do so in three steps. First, he would point out that since one doesn't know which trajectory the system is on, if the program is to make any sense at all, it must deal with the *average* Boltzmann entropy

$$\bar{S}_B \equiv \sum_j p_j(t) \log \Gamma_j \ ,$$

where $p_j(t)$ is the probability to be in cell j at time t, given an initial uniform distribution on M. Lebowitz's refusal to use the average Boltzmann entropy and his insistence on using the Boltzmann entropy for individual trajectories are probably based on his distrust of probabilities for a single system, which after all does have a particular trajectory. Second, since the average Boltzmann entropy is only part of the missing information, the Bayesian would replace it with the closely related Gibbs entropy of the coarse-grained distribution, which quantifies the total missing information and is directly related to the available work:

$$S(t) = k_B \left(-\sum_j p_j(t) \log p_j(t) + \sum_j p_j(t) \log \Gamma_j \right) \ .$$

Third, the Bayesian would notice that the procedure used to get $p_j(t)$ and, hence, $S(t)$ does not consistently discard information: $p_j(t)$ comes from coarse graining at each time t the exact distribution that evolves from the initial distribution, instead of coarse graining the distribution that comes from the coarse graining at the preceding time step. As a result, $S(t)$ occasionally decreases. Although the bobbing up and down of $S(t)$ is very much suppressed relative to the bobbing up and down of the Boltzmann entropy for a particular trajectory, a decrease of $S(t)$, no matter how small, describes a decrease in the missing information, even though one has not acquired any new information about the system. This doesn't make any sense, so a Bayesian would replace $p_j(t)$ with the probability $p_j(n)$ that comes from consistently discarding fine-grained information. The resulting coarse-grained Gibbs entropy,

$$S(n) = k_B \left(-\sum_j p_j(n) \log p_j(n) + \sum_j p_j(n) \log \Gamma_j \right) \ ,$$

is the one introduced earlier to describe coarse graining. It increases monotonically and has a direct physical interpretation in terms of available work.

The result of a fix of the Lebowitz program is to disenthrone the Boltzmann entropy and to replace it with the coarse-grained Gibbs entropy that lies at the heart of any coarse-graining strategy. What an irony it is that Lebowitz heaps scorn on the Gibbs entropy, saying that it can't be the right entropy for nonequilibrium situations, because unlike the Boltzmann entropy, it "does not change in time even for time-dependent ensembles describing (isolated) systems not in equilibrium."

How do people fool themselves that the Lebowitz program is sensible when its underlying principles are so fundamentally flawed? There are two reasons. First, as we have seen, the Lebowitz program can be placed on a sensible Bayesian foundation simply by reinterpreting it as a standard coarse-graining procedure that uses the coarse-grained probabilities $p_j(n)$ and the corresponding Gibbs entropy $S(n)$. Second, for the coarse graining used by Lebowitz, the first term in the coarse-grained Gibbs entropy, which is the missing information about which coarse-grained cell the system occupies, is negligible compared to the second term, and after a short time, the second term is dominated by a single cell, which has nearly unity probability and nearly all the phase-space volume. After a short time, the coarse-grained Gibbs entropy is given approximately by the Boltzmann entropy for the dominant cell, thus making the Lebowitz program a very good approximation to a well founded Bayesian coarse graining, even though its justification doesn't make sense.

This means that almost all of what is done in the Lebowitz program can be given a sensible Bayesian reinterpretation. For example, Lebowitz notes, "The microstates in Γ_{M_b}, which have come from Γ_{M_a} through the time evolution during the time interval from t_a to t_b, make up only a very small fraction of the volume of Γ_{M_b}, call it Γ_{ab}. Thus we have to show that the overwhelming majority of points in Γ_{ab} (with respect to the Liouville measure on Γ_{ab}, which is the same as the Liouville measure on Γ_{M_a}) have *future* macrostates like those typical of Γ_b—while still being very special and unrepresentative of Γ_{M_b} as far as their *past* macrostates are concerned." When reinterpreted, this is simply the statement that one desires that the future behavior of macroscopic variables be insensitive to the coarse graining.

Lebowitz says that the "big question" is why are the initial conditions so special and concludes, along with many others, that one must posit that the Universe was originally in a much more ordered state than it is now. We have seen above that this conclusion simply cannot be supported, but that it can be replaced by the conclusion that the evolved state, though not complex itself, is close to very complex states.

4. Bayesian or information-based interpretation of quantum mechanics. Much of the material in Secs. 4.a–c is contained in Caves, Fuchs and Schack (2001a, 2001b).

Let's begin with motivation provided by E. T. Jaynes (1990a), the great physicist and Bayesian:

> Let me stress our motivation: if quantum theory were not successful pragmatically, we would have no interest in its interpretation. It is precisely *because* of the enormous suc-

cess of the QM mathematical formalism that it becomes crucially important to learn what that mathematics means. To find a rational physical interpretation of the QM formalism ought to be considered the top priority research problem of theoretical physics; until this is accomplished, all other theoretical results can only be provisional and temporary.

> This conviction has affected the whole course of my career. I had intended originally to specialize in Quantum Electrodynamics, but this proved to be impossible. Whenever I look at any quantum-mechanical calculation, the basic craziness of what we are doing rises in my gorge and I have to try to find some different way of looking at the problem, that makes physical sense. Gradually, I came to see that the foundations of probability theory and the role of human information have to be brought in, and so I have spent many years trying to understand them in the greatest generality.
>
> . . .
>
> Our present QM formalism is a peculiar mixture describing in part laws of Nature, in part incomplete human information about Nature—all scrambled up together by Bohr into an omelette that nobody has seen how to unscramble. Yet we think the unscrambling is a prerequisite for any further advance in basic physical theory, and we want to speculate on the proper tools to do this.

The information-based or Bayesian interpretation of quantum mechanics is founded on the notion that quantum states, both pure and mixed, represent states of knowledge and that all the probabilities they predict are Bayesian probabilities.

This point of view, particularly as it applies to the probabilities that arise from pure states, seems crazy at first. The probabilities that come from a pure state are intrinsic and unavoidable. How can they not be objective properties of a quantum system when they are prescribed by physical law? How can they be ignorance probabilities when one knows everything possible about the quantum system? Indeed, as Giere (1973a) notes, if one is to find objective probabilities, one must look outside the determinism of the classical world, and quantum mechanics, with its intrinsic indeterminism, seems to be just the place to look. Many physicists, even Bayesian ones, have assumed instinctively that quantum probabilities are different from the ignorance probabilities of a realistic/deterministic world. Nonetheless, our view is that *all* probabilities—even quantum probabilities—are Bayesian, i.e., based on what one knows, the Bayesian view being the only consistent way to think about probabilities. The probabilities of quantum mechanics—even those that arise from a pure state—are based on what the describer knows. Let's give E. T. Jaynes (1990a) another hearing:

> For some sixty years it has appeared to many physicists that probability plays a fundamentally different role in quantum theory than it does in statistical mechanics and analysis of measurement errors. It is a commonly heard statement that probabilities calculated within a pure state have a different character than the probabilities with which different pure states appear in a mixture, or density matrix. As Pauli put it, the former represents "... eine prinzipielle *Unbestimmtheit*, nicht nur *Unbekanntheit*". But this viewpoint leads to so many paradoxes and mysteries that we explore the consequences of the unified view, that all probability signifies only human information.

a. *Why and how probabilities? Kochen-Specker and Gleason.* We adopt the Hilbert-space structure of quantum questions, which in its finest-grained form deals with questions described by orthogonal one-dimensional projectors.

The Kochen-Specker theorem says that there is no way (in three or more dimensions) to assign truth or falsity to every one-dimensional projector (only finite sets of projectors are required for the proof) in such a way that every quantum question has a definite, predictable answer. As a consequence, quantum mechanics cannot be reduced to certainties, but rather must deal with probabilities. The crucial assumption in the Kochen-Specker theorem, called *noncontextuality*, is that the truth or falsity of a projector does not depend on which orthogonal set it is a member of. This assumption, unreasonable for a hidden variable theory, which ought to be able to snub its nose at the Hilbert-space structure of quantum questions, is a reasonable, even necessary assumption for our purpose of demonstrating that quantum mechanics must deal with probabilities. Noncontextual truth assignments ignore the Hilbert-space structure; that being the only input from quantum mechanics, if one ignores it, one can't hope to find out anything about quantum mechanics.

The rule for assigning probabilities comes from Gleason's Theorem. A frame function assigns to each one-dimensional projector $\Pi = |\psi\rangle\langle\psi|$ a number between 0 and 1 inclusive, with the property that the function sums to 1 on orthogonal projectors. A frame function makes a noncontextual probability assignment to all quantum questions (noncontextual because the probability assigned to a projector does not depend on which orthogonal set it is a member of). Gleason's Theorem shows that (in three or more dimensions) any frame function can be derived from a density operator ρ according to the standard quantum rule, $\mathrm{tr}(\rho\Pi) = \langle\psi|\rho|\psi\rangle$. Thus in one stroke, Gleason's theorem establishes that density operators provide the state-space structure of quantum mechanics and gives the rule for calculating probabilities from states. For the same reason as above, the assumption of a noncontextual probability assignment is perfectly reasonable here (although perhaps less convincing because probability assignments can be tweaked slightly, whereas truth assignments cannot), where we are trying to determine the consequences of quantum mechanics for probability assignments. From a Bayesian perspective, what Gleason's theorem says is that the only way for someone to assign probabilities to quantum questions in a way that doesn't correspond to a density operator is to make a contextual assignment.

In a realistic/deterministic world, maximal information corresponds to knowing which of a set of mutually exclusive, exhaustive alternatives is the true one. It provides a definite, predictable answer for all questions, including the finest-grained ones, i.e., those that ask which alternative is true. The slogan is that in a realistic/deterministic world, "maximal information is complete," providing certainty for all questions. Quantum mechanics is different, since no density operator (quantum state) gives certainty for all questions. Mixed states cannot provide certainty for any fine-grained question. Only pure states—themselves one-dimensional projectors—can provide certainty for some fine-grained questions. Thus they are the states of maximal information. The quantum slogan is that "maximal information is not complete and cannot be completed," thus giving rise to Bayesian probabilities even when one knows as much as possible about a quantum system.

States of maximal information correspond to well defined preparation procedures both in a realistic/deterministic world and in quantum mechanics, the procedure being the one that renders certain the answers to the appropriate questions. Other states of knowledge do not correspond to well defined procedures. There are many different kinds of situations where one assigns probabilities in a realistic/deterministic world or a mixed state in quantum mechanics, but one such situation is where the describer knows that the system has been prepared by a well defined procedure, but does not know which procedure.

A complete theory must have a rule for assigning probabilities in the case of maximal information: if one has maximal information, there is no other information that can be brought to bear on the probability assignment; thus, if the theory itself is complete, it must supply a rule. In a realistic/deterministic world, maximal information corresponds to certainty, and the Dutch book argument requires all probabilities to be 0 or 1. In quantum mechanics, where maximal information is not complete, the Dutch book argument only prescribes probabilities for those questions whose outcome is certain. Fortunately, Gleason's theorem comes to the rescue, providing the unique rule for assigning probabilities that is consistent with the Hilbert-space structure of questions. For this essential service, *Gleason's theorem can be regarded as the greatest triumph of Bayesian reasoning.*

Perhaps the most compelling argument for the subjectivity of quantum probabilities comes from the multiplicity of ensemble decompositions of a density operator. The ensemble probabilities are clearly Bayesian, reflecting ignorance of which pure state in the ensemble, for all the reasons cited for classical probabilities. The probabilities derived from the pure states in the ensemble are natural candidates for "objective probabilities." The problem with this idea is that the multiplicity of ensemble decompositions means that is impossible to separate cleanly the subjective and objective probabilities. Here's how Jaynes (1957b) put it long ago in one of his pioneering articles on the foundations of statistical physics:

> A density matrix represents a fusion of two different statistical aspects; those inherent in a pure state and those representing our uncertainty as to which pure state is present. If the former probabilities are interpreted in the objective sense, while the latter are clearly subjective, we have a very puzzling situation. Many different arrays, representing different combinations of subjective and objective aspects, all lead to the same density matrix, and thus to the same predictions. However, if the statement, "only certain specific aspects of the probabilities are objective," is to have any operational meaning, we must demand that some experiment be produced which will distinguish between these arrays.

The multiplicity of decompositions implies that all the probabilities must be given the same interpretation. Since some of the probabilities are clearly subjective, one is forced to acknowledge that all quantum probabilities are Bayesian. The mixed-state decomposition problem in quantum mechanics tell us that in quantum mechanics, even more than in a realistic/deterministic world, you have to have a consistent view of probabilities and stick with it, and the only consistent view of probabilities is the Bayesian view. Anyone who promotes an interpretation of quantum mechanics without first declaring his interpretation of probabilities should be sent back to the starting gate.

b. *Quantum states as states of knowledge.* The simplest argument for why quantum states are states of knowledge is the same as the argument for probabilities: If you want to know the state of a quantum system, you cannot ask the system; it doesn't know its state. If you want to know the state, you must ask the describer.

Given maximal information, the Dutch book argument requires assignment of a unique pure state, so once you obtain the maximal information of the describer, you must assign the same pure state. If the describer tells you that the state is mixed, then you are free to assign a different mixed state based on the same information or to acquire privileged information that permits you to assign a different state, pure or mixed (the Dutch-book argument does require your mixed state to have support the lies within the support of the describer's mixed state).

The notion that a pure state corresponds to maximal information already says that the properties whose statistics are given by quantum probabilities cannot be objective properties of a quantum system. If they were, then the purportedly maximal information would not be maximal and should be completed by determining the values of all these properties.

A curious feature of this simple argument is that it doesn't have to change at all to accommodate hidden-variable theories. Indeed, Bayesian probabilities are the natural way to think about quantum states in a hidden-variable theory. In a hidden-variable theory, all the properties of the system are objective, having actual values, but values that are determined by "hidden variables" that are for the present inaccessible. The quantum probabilities are naturally regarded as Bayesian probabilities, but now reflecting ignorance of the hidden variables, and the quantum state, as a summary of those probabilities, is a state of knowledge. As long as the hidden variables remain inaccessible, it is still not possible to determine the state by asking the system. The only difference is that in a hidden variable theory, the purportedly maximal information corresponding to a pure state isn't maximal at all, even though it might be impossible in principle to complete it. Thus the slogan for hidden-variable theories is, "Apparently maximal information is not maximal, but might or might not be completable."

The virtue of Bell inequalities is that they show that the hidden variables must be nonlocal, if they are to duplicate the statistical predictions of quantum mechanics, and that they provide experimental tests that distinguish quantum mechanics from local hidden variable theories. For this reason, variants of the "you can't ask the system" argument for entangled states are perhaps more convincing. In particular, there is Chris's favorite: by making an appropriate measurement on one member (Alice's) of an entangled pair, you can make the pure state of the other member (Bob's) be a state chosen randomly from any orthonormal basis. This is accomplished without in any way interacting with Bob's particle. Different von Neumann measurements on Alice's particle will leave Bob's particle in a state chosen randomly from incompatible bases. This is a cogent argument for pure states being states of knowledge instead of states of the world. Nonetheless, the argument can be put in the context of *nonlocal* hidden-variable theories, where there would be a real nonlocal (potentially acausal) influence of Alice's choice of measurement on Bob's particle. Thus the only real advantage of this argument over the simple you-can't-ask-the-system argument—perhaps this is a considerable advantage—is that it forces the hidden variables to be nonlocal if they are to be capable of providing an ontology underneath the quantum state.

We can summarize our position as follows: A pure state described by a state vector corresponds to a state of maximal information, for which there is a well defined, repeatable preparation procedure. That is the reason one assigns the product state to many copies of a system identically prepared, for one then has maximal information about the composite system and must assign a state that gives certainty for repeated yes/no questions corresponding to the state. On the other hand, one shouldn't fall into the trap of regarding the state vector as real. It corresponds to a state of knowledge. The proof is in the fact that even though one can prepare a pure state reliably, using the maximal information, one can't determine it, which one ought to be able to do if it is real. Someone else cannot determine the state of a system, because it is not out there, but in the mind of the describer. If you want to know the state I assign, who do you ask? The system or me? The state vector can change as a consequence of my obtaining information, and this also argues strongly for its subjective status. A pure state, rather than being objective, is *intersubjective*, because of the reproducibility of maximal information and the Dutch-book-enforced agreement on assigning a pure state.

c. *Principle of quantum indeterminism and the quantum de Finetti theorem.* A guiding principle that we use in assigning probabilities is that one should never make a probability assignment that prohibits learning from data, by which we mean using data to update probabilistic predictions for situations from which data has not yet been gathered, *unless* one already has maximal information, in which case there is nothing further to learn. If you do not have maximal information, there is always room for hypotheses about things you do not know, so you should allow for these hypotheses in your probability assignments. When you do have maximal information, your probability assignment should not allow learning from data.

In a realistic/deterministic world, maximal information means certainty, so all probabilities are 0 or 1. In contrast, in a quantum world, "maximal information is not complete and cannot be completed" and thus gives rise to Bayesian probabilities even when one knows as much as possible about a quantum system.

The classical and quantum de Finetti theorems (Caves, Fuchs, and Schack 2001a), which deal with exchangeable probabilities or density operators, provide a setting for applying this guiding principle. For exchangeable sequences, i.i.d. probability assignments are the unique ones that do not allow any learning. The guiding principle implies that one should never assign the i.i.d., except when one has maximal information. In a realistic/deterministic world, where maximal information is complete, this means that one should assign the i.i.d. only in the trivial case of certainty. This refusal to assign the i.i.d. leaves open the possibility of using frequency data from initial trials to update probabilistic predictions for further trials. Things are different in a quantum-mechanical world. Since maximal information is not complete, one can assign (nontrivial) i.i.d.'s to exchangeable trials when all the systems are described by the same pure state. Notice that the data gathered from such trials comes from "the bottomless quantum well of information" and is useless for updating the state assignment for additional systems.

Since it is easy to fail to appreciate the content of our guiding principle in the classical case, it is worth discussing that situation in some detail. Should someone not see the point of the guiding principle, ask if he would assign equal probabilities to further tosses of an allegedly fair coin after getting 100 consecutive heads. If he continues to accept even bets, you've found a gold mine. If he doesn't, then point out that any time one doesn't continue to assign probabilities based on the initial single-trial probability, it means one didn't assign the i.i.d. to the multi-trial hypothesis space. This is because the data from initial trials cannot be used to update via Bayes's rule the probabilities for further trials when the multi-trial assignment is an i.i.d.

It is an interesting aside to note that many people react in contradictory ways—you might want to test your own reaction—when presented with the same problem in slightly different guises. Handed what is said to be fair coin, they will assert that one ought to stick with 50-50 predictions for future tosses even after many consecutive tosses give heads. On the other hand, given a coin about which they are told nothing is known, they will assert that the the probability for heads in the $(N+1)$th toss is the observed frequency of heads, n/N, in the first N tosses, and they will heap scorn on Laplace's Rule of Succession, which says to use head probability $(n+1)/(N+2)$ for the $(N+1)$th toss. From a Bayesian perspective, these different attitudes reflect different probabilities on probabilities or, using the de Finetti representation, different exchangeable multi-trial probabilities. The desire to stick with an initial 50-50 probability comes from a probability on probabilities that is highly concentrated near 50-50—the robust 50-50 odds are then what was called a propensity above—the consequence being an extreme reluctance to let the data change the initial 50-50 probabilistic predictions. The use of observed frequency to predict future tosses reflects just the opposite prejudice, i.e., a prejudice for letting the data dictate predictions for further trials. Laplace's Rule of Succession lies in between, but much closer to learning from the data than to a stubborn insistence on an initial single-trial probability. An excellent discussion of these questions and their relation to probabilities on single-trial probabilities can be found in Jaynes (1986b).

Hypothesis testing and parameter estimation (continuous variable hypothesis testing) are real-life situations where one uses probabilities on probabilities. Each hypothesis leads to different probabilistic predictions for data that will be used to decide among the hypotheses, and thus each hypothesis represents some state of knowledge about the data. The state of knowledge might include knowledge of the actual value of some objective property, but it cannot include complete knowledge, for then the data could be predicted with certainty. Thus the prior probabilities on the hypotheses are probabilities on probabilities, which in a Bayesian perspective should be banished in favor of primary probability assignments on the data. The de Finetti representation theorem shows how to do this in the case of data that is exchangeable. Notice, however, that if the only difference between hypotheses lies in different, but unknown values of an objective property, then the goal of collecting data is to make the best possible determination of that objective value.

Before returning to quantum mechanics, we have to deal with a couple of objections to our statement that one should never assign the i.i.d. in a realistic/deterministic world. For *any* exchangeable probability assignment for binary variables, if one selects the successive occurrences of head-tail and tail-head, these occurrences have equal probabilities and thus are governed by the 50-50 i.i.d. Recall that this scenario, which is the simplest of many such scenarios, raised its ugly head earlier in the context of an attempt to find objective classical probabilities; easily disposed of in that context, it is more insidious now because within a Bayesian context, it challenges our guiding principle that one should never assign the i.i.d. classically except when there is certainty. The challenge is easily met, however, because the essence of our guiding principle is that one should never assign probabilities that prohibit learning from data except in the case of maximal information. We did not intend, for it is not true, that such a probability assignment prohibit selecting subsets of the data from which nothing can be learned, and that's what happens in the head-tail vs. tail-head scenario. We don't need to change our statement that one should never assign probabilities that forbid learning, except in the case of maximal information; in the case of i.i.d.'s, however, when being utterly precise, we should say that one should never assign the i.i.d. to the base hypothesis space, except in the case of maximal information.

There is another, more troubling objection. In the case of exchangeable multi-trial probabilities, collecting frequency data from many trials allows one to update the probabilities for further trials. In the limit of infinitely many trials, the probability assignment for further trials converges to an i.i.d. whose single-trial probabilities are given by the measured frequencies. So what happens to our statement that you shouldn't assign the i.i.d. except in the case of maximal information?

There are two good answers to this question. The first, easier answer is that you don't converge to the i.i.d. except in the unattainable infinite limit. For finite numbers of trials, your single-trial probabilities will become increasingly robust, but there will always be some remaining doubt. Since we never questioned the idea that one might make arbitrarily robust single-trial probability assignments, there is no contradiction with our guiding principle.

The second, probably better answer takes one outside the arena of exchangeable probabilities. If successive trials do not all yield the same result, then in a realistic/deterministic world, there are undiscovered details about the trials that if known, would give certainty for each trial. A random-number generator or a calculation of the digits of π provides a good example of the kind of underlying realistic mechanism might be at work in generating successive trials. Knowing all the details or some part of them necessarily takes one outside the province of exchangeable probabilities unless all the trials yield the same result. Learning about these details from data requires more than frequency data—it would involve information about correlations between trials—and updating probabilities based on this nonfrequency data requires a nonexchangeable probability assignment. Thus the guiding principle actually says that one should never make a strictly exchangeable probability assignment in a realistic/deterministic world, except in the case of maximal information, and this neatly avoids the possibility of assigning an i.i.d. or converging to an i.i.d.

In quantum mechanics we replace the objectivist's obsession with objective probabilities with a Bayesian attention to the conditions under which one should assign the i.i.d. We argue that one assigns the i.i.d. only in a situation corresponding to maximal information, which can be reproduced reliably from trial to trial. In a realistic/deterministic world this gives the i.i.d. only in the case of certainty, but in a quantum world it gives the i.i.d. for pure states, i.e., maximal information.

One can argue that one never actually has maximal information about a system, either classical or quantum mechanical, and that this means that one never does assign the i.i.d. even in quantum mechanics. Though it is perhaps true that maximal information is an unattainable limit of more and more refined information, the crucial point is that this limit corresponds to certainty in a realistic/deterministic world, whereas it corresponds to a pure state—and the consequent i.i.d.'s—in quantum mechanics.

From the Bayesian perspective, there is no necessary connection between probabilities and frequencies in a realistic/deterministic world. That is the content of our guiding principle when applied to exchangeable sequences. Nonetheless, what we have learned is that in quantum mechanics, there can be a strict connection between single-trial probabilities and observed frequencies. The reason is that the well defined procedure for preparing a pure state allows one to prepare many copies of a system, all in the same state, and this leads to the i.i.d. for repeated measurements of the same quantity on the successive copies. This is Rüdiger's argument, and he has dubbed the fundamental new quantum-based connection between probabilities the *principle of quantum indeterminism*. It is the Bayesian answer to why the probabilities that come from pure states can be used to predict frequencies.

Notice the similarity of our argument to Giere's (Giere 1973a) view of objective chance. Giere admits that in a deterministic world the only objective chance is certainty, but maintains that in a quantum world the quantum probabilities correspond to objective chance. The difference is that Giere has the probabilities change character, from subjective to objective in the limit of maximal information, whereas we regard all probabilities as Bayesian. Even in the limit of maximal information, we think of the probabilities as Bayesian, i.e., based on the maximal information, but we say that the maximal information can be reproduced reliably from trial to trial.

A potential weakness of our argument is that there is no operational difference between our view—assign Bayesian i.i.d.'s only in the case of maximal information—and Giere's view—the only objective probabilities, which give the i.i.d., occur in the situations we call maximal information. Indeed, I think one of the responses to our view will be that maximal information defines those situations where probabilities are objective properties. The difficulty is that we want to retain in quantum mechanics some, but not all of the features of classical maximal information, and there will certainly be disagreement over the features we choose. States of classical maximal information can be prepared and verified reliably; they provide the ontology of a realistic/deterministic world. States of maximal information in quantum mechanics can be prepared reliably, but they cannot be verified (you can't ask the system); because they cannot be verified, we do not accord them objective reality, and we regard the probabilities they generate as Bayesian probabilities.

d. *Ontology.* In the Bayesian interpretation, the states of quantum systems do not have objective reality—they are states of knowledge, not states of the world—and the values of the properties of microscopic systems do not have objective reality. The apparatus of quantum states and associated probabilities is an elaborate theory of inference—a law of thought, in Chris's phrase—in which we put in what we know and get out statistical predictions for things we can observe. In my version of the Bayesian interpretation, the objective parts of the theory—the ontology—lie in the other part of quantum mechanics, i.e., in the physical laws that govern the structure and dynamics of physical systems. These physical laws are encoded in Hamiltonians and Lagrangians.

   What is the evidence that Hamiltonians have ontological status. The most compelling argument is the following: *whereas you can't ask a quantum system for its state, you can ask it for its Hamiltonian.* By careful experimentation, you can deduce the Hamiltonian of a quantum system—that's what physics is all about.

   A second argument comes from dynamics. It is possible to argue that the only kind of dynamics consistent with the convex structure of density operators is given by positive superoperators (linear maps on operators). Among these positive superoperators, a special place is occupied by those maps that preserve maximal information, and it is possible to show that the only positive maps that preserve maximal information and are connected continuously to the identity are generated by Hamiltonians. If you have maximal information about a quantum system and you want to retain such maximal information, you must know the system's Hamiltonian. Where does that Hamiltonian come from? I think it must be an objective property of the system.

   What does it mean to say that Hamiltonians are objective? Does it mean that the formula written on a page is real? That's silly. Does it mean that the positions and momenta and spins in it are real? Certainly their values don't have objective status, but their appearance in the Hamiltonian does determine the kind of Hilbert space that applies to the system and thus dictates the structure and—this is important—to some extent the meaning of the quantum questions that began our discussion of quantum probabilities. Moreover, the form of the Hamiltonian in terms of positions, momenta, spin, and so forth together with the parameters in the Hamiltonian—masses, charges, coupling constants—determines the physical properties of systems. If microscopic systems are to have any real properties, it is these physical properties that are the best candidates, and their objective status is equivalent to the objective status of at least some part of the Hamiltonians governing microscopic systems.

   Again we run up against the question of which aspects of maximal information are to be promoted to objective status. In a realistic/deterministic world, where maximal information leads to certainty, we regard as objective the alternatives that correspond to maximal information. In quantum mechanics, we do not grant objective status to the pure states that correspond to maximal information. Fundamentally, this is because maximal information not being complete, pure states lead to probabilities, and we know that the only consistent way to interpret probabilities is the subjective, Bayesian interpretation. How could the pure state be objective if the probabilities it predicts aren't? When we come to Hamiltonians, which provide the rule for updating maximal information, we face a choice. Classically they can naturally be regarded as objective, but what should we do in quantum mechanics? The choice is by no means clear, but the choice made in this document is to say that they are objective properties of the system being updated. Fundamentally, this choice comes down to the fact that Hamiltonians aren't directly associated with probabilities, so we are free of the prejudice to declare them subjective. The slogan is that *quantum systems know what they're doing, but we often don't.* The hope is that this point of view can be an ingredient in constructing the effective reality that emerges from quantum mechanics.

Quantum computation provides some support for the notion that Hamiltonians are objective. The standard description of an ideal quantum computation is something like the following. The computer is prepared initially in a pure state in the computational basis; this input state might encode input information. The computer then runs through a prescribed set of unitary steps that leave it in a computational basis state that stores the result of the computation. A measurement reveals the result. The user doesn't know the output state—otherwise he wouldn't need to do the computation—so he assigns a density operator to the output. In a Grover-type search algorithm, the user doesn't know the output state because he is trying to discover the actions of an "oracle," which prepared the output state. In a Shor-type algorithm, however, the user doesn't know the output state even if he knows the prescribed sequence of unitaries, because the consequences of the unitary steps are too complex too be simulated. In this case the output tells the user something objective about the input—the factors in Shor's algorithm—and the unitary steps act as a sort of guarantee that the answer can be trusted. For a quantum computer, even if one knows the Hamiltonian, one cannot retain maximal information, but nonetheless trusts that the computer will output the right answer. This is truly an example where "the system knows what it is doing, even though the user doesn't."

e. *Emergent effective reality.* The realistic/deterministic reality of everyday experience is an emergent property in the Bayesian interpretation. It arises in sufficiently complicated, "macroscopic" systems in which one is able to observe only "macroscopic" variables. The mechanism for its emergence is Zurek-style (1998a) decoherence that comes from not having access to the microscopic variables (formally, one traces over the microscopic variables). The result is hoped to be the "effective reality" of our everyday experience. An important question is the extent to which the properties of the effective reality are dictated by the laws of physics—Hamiltonians and Lagrangians—as opposed to depending on the quantum state of the microscopic variables. It would be nice if the character of the effective reality came mainly from the microscopic ontology, i.e., Hamiltonians, with only minimal dependence on the subjective quantum state.

There is some reason to think this is true. One argument is that since the physical laws involve local interactions, natural selection would favor a local reality, not nonlocal superpositions, which would be difficult to follow because of decoherence unless one had the ability to monitor the details of the environment. This, or some other argument, perhaps not involving natural selection, gives the separation between the ontology and the epistemology in the Bayesian interpretation: the Hamiltonians are the ontology, giving rise to the effective reality, and the structure of quantum states provides rules of inference for making predictions in view of what we know. Even without the whole Bayesian apparatus of quantum inference, active agents can take advantage of the predictability within the almost realist/almost deterministic effective reality.

We are now in a position to be more specific about quantum states as states of knowledge—knowledge about what?—and quantum probabilities as ignorance probabilities—ignorance of what? The probabilities of quantum mechanics—even those that arise from a pure state—are based on what the describer knows. They do not reflect ignorance of some underlying reality, as in a hidden-variable theory, but rather ignorance of results of possible observations. They are ultimately probabilities for "macroscopic" alternatives in the emergent effective reality of the "macroscopic" world. They are ignorance probabilities because the describer cannot predict which of the macroscopic alternatives is realized. They are based on what the describer knows about "microscopic" systems that intervene between himself and the ultimate alternatives, microscopic systems that cannot be described within the approximately realistic, approximately deterministic effective reality, but must be described by quantum mechanics, thereby introducing uncertainty into the ultimate predictions.

This kind of effective reality is the best we can hope for. Einstein emphasized that there are not two presentations of reality: one out there and one constructed from our theories that are tested against what's out there. Rather there is a single reality constructed from our perceptions and our theories about them. This is just what the effective reality provides. It is the reality that is relevant for avoiding a predator, catching prey, or playing baseball *and* for making statistical predictions about the behavior of microscopic systems whose behavior lies outside the almost-realistic/almost-deterministic world of everyday experience.

The notion of an effective reality is clearly a program that is mainly not even started, but it builds on the considerable work already done on decoherence and decoherent histories. There is an important consistency issue in constructing the effective reality from physical law applied to quantum systems and then using that same effective reality as the backdrop for our investigations of quantum systems.

Two further comments on Hamiltonians: (i) Non-Hamiltonian (i.e., nonunitary) evolutions inevitably involve a state assignment to another system that interacts with the primary system, and thus they include a subjective component. Evidence for the subjectivity of nonunitary evolutions thus has no bearing on the reality of Hamiltonians. (ii) Most Hamiltonians are emergent in some sense. The Hamiltonians of classical mechanics, for example, take as given parameters that ultimately come from a quantum description of atomic systems; these Hamiltonians are thus a part of the effective reality, not an ontological input to it. As another example, the Hamiltonians of atomic physics and condensed-matter physics take as input the masses and charges of electrons and protons, which come as input from a more fundamental description like QCD.

A final point before moving on. The Bayesian interpretation takes as given that particular events happen in the effective reality. The almost-realistic/almost-deterministic flow of events that occur wholly within the effective reality can only be pushed so far into the microscopic level; pushed farther, it fails. Even though we operate wholly within the macroscopic level, our description of what happens there sometimes requires us to dip into the quantum world and thus to introduce uncertainty into the consequences within the effective reality; under these circumstances, the responsibility of the theory is to provide probabilities for events that cannot be predicted with certainty. Quantum theory gives us a reliable method for predicting the probabilities of these uncertain events, and with that accomplished, its job is finished. In particular, a detailed explanation of how a particular event is actualized out of the several possibilities is, we contend, outside the provenance of quantum theory.

f. *Other issues.*

We have a Dutch-book argument for why two observers, sharing the same maximal information, cannot assign different pure states. This argument ought to be tested in a variety of circumstances where there seems to be no pure state or there seems to be more than one pure state: (i) two-time situations where there is no consistent state-vector assignment (measure x spin and later z spin; in between, measurements of x and z spin are both determined, when conditioned on pre- and post-measurements; no state vector has these statistics); (ii) relativistic situations where two observers assign different state vectors (not really a problem, because they never run into any betting contradictions); (ii) pure-state assignments in the situation where one observer is part of the state vector of another super-observer (one can get contradictory wave function assignments, although the super-observer will be unable to assign a wave function to the system being considered by the subobserver).

What is a measurement? It is the acquisition of information that is used to update the quantum state (collapse of the wave function, to use less neutral terms). Can dogs collapse the wave function? This is a dumb question, since dogs don't use wave functions. What they do use is the emergent reality, in which they and other agents gather and process information and make decisions based on the results. What we are now doing in quantum information science is getting beyond information processing that occurs entirely within the effective reality of our everyday experience, even though it uses the structures of quantum systems; instead we are doing rudimentary information processing part of which occurs in the quantum world, where the realism of everyday experience doesn't apply. The power of quantum information lies in the fact that it allows us to escape the constraints of realistic classical information. This is something a dog can't do, so it might be said that just as we are beginning to understand the complex operations of our own genome, we are also leaving dogs behind for the first time.

There is a misconception about applying the conclusions of quantum mechanics—that the properties of microscopic systems do not have objectively real values, loosely stated as there being no objectivity at the microscopic level—to the wider world around us. The founders of quantum mechanics—Bohr, Heisenberg, and Pauli—sometimes hinted that their thinking about quantum mechanics could be applied to cultural, political, historical, and social questions. Postmodernists can make the same mistake in a different way: to the extent that they pay any attention to quantum mechanics, they might look at the subjective features of quantum mechanics and say that cultural, political, historical, and social questions inherit the same subjectivity. This is nonsense. It is, ironically, the basic mistake of reductionism—thinking that what occurs at lower levels in the hierarchy of our description of the world is happening at higher levels, when in fact the microscopic, quantum level provides the effective, realistic structure in which the higher levels operate. The objectivity of events is an emergent property, which applies in the effective reality, but this doesn't make them any less objective. The theory provides the stage of an emergent effective reality, and culture, politics, history, and sociology act on that stage.

5. **The four questions.** There are four questions that should be addressed in thinking about any interpretation of quantum mechanics.
   - **Probabilities.** What is the nature of probabilities (ignorance or Bayesian probabilities vs. frequentist or ensemble probabilities)? In particular, what is the nature of quantum probabilities?
   - **Ontology vs. epistemology.** Does the interpretation posit an ontology (underlying objective reality), or is it wholly epistemological? If there is an ontology, what is it?
   - **Which basis and which alternative?** How does the interpretation explain which basis (i.e., which observable) is to be used and how a particular alternative is actualized within that basis?
   - **Classical reality.** How does the interpretation explain the world we observe around us? Does its ontology aid in this explanation?

   For theories that go beyond quantum mechanics, we should include a fifth question:
   - **Different predictions.** How do the the statistical predictions of the new theory differ from the predictions of quantum mechanics? At Växjö in 2001 June, Lucien Hardy suggested to me that this fifth question should be expanded to something like: How much can the theory be tweaked so as to provide a range of alternatives to quantum mechanics? Such theories are extremely valuable even if they turn out to be wrong, because they provide a way of getting at which features are special to quantum mechanics and which are only incidental.

For the Bayesian interpretation outlined in Sec. 3, the answers to these questions—or at least the hope for an answer—should be clear. It is worth emphasizing that the Bayesian interpretation places actualization outside its provenance. The reason for emphasizing this point is that much pointless wrangling over interpretations comes from repeated accusations that an interpretation doesn't deal with actualization. It is best to explicitly eschew actualization if your interpretation doesn't deal with it. Then, though it is a perhaps devastating criticism that your interpretation doesn't deal with actualization, it is a criticism that only needs to be made once.

6. Other interpretations.

a. *Copenhagen interpretations.* The Copenhagen interpretation has acquired so much baggage—classical apparatuses, classical language, quantum/classical cuts, necessity of a classical world, complementarity, uncertainty-principle limits, forbidden questions, Bohrian obscurity—that it is next to impossible to sort out what it is. At this stage in history, it is probably best just to wipe the slate clean and give the Copenhagen interpretation a new, more informative name that highlights its essential feature. The *information-based interpretation*, with its insistence that quantum states are states of knowledge, *is* the new Copenhagen interpretation.

b. *Ensemble interpretations.* These aren't really interpretations. They are misconceptions about probabilities that are taken directly over into quantum mechanics because quantum mechanics necessarily deals with probabilities. If you believe in an ensemble interpretation, you need first to get your notion of probabilities straight before proceeding to quantum mechanics. Again quoting E. T. Jaynes (1986a):

> We think it unlikely that the role of probability in quantum theory will be understood until it is generally understood in classical theory and in applications outside of physics. Indeed, our fifty-year-old bemusement over the notion of state reduction in the quantum-mechanical theory of measurement need not surprise us when we note that today, in all applications of probability theory, basically the same controversy rages over whether our probabilities represent real situations, or only incomplete human knowledge.

The traditional ensemble interpretation, mainly encountered in old quantum mechanics texts, means the idea that the wave function must be interpreted as describing a *real* ensemble of identical systems. It has fallen out of favor because we apply the wave function (more generally, the density operator) routinely (in quantum optics, for example) to individual systems subjected to repeated measurements. But the real problem is that it takes a misconception about probabilities—that they can't be applied to single cases and thus must be referred to a real ensemble—and infects quantum mechanics with the same misconception. The right approach, as Jaynes says, is to get the interpretation of probabilities straight before proceeding to quantum mechanics.

What about hypothetical ensembles? Here the main candidate is the Hartle (1968a) approach to getting quantum probabilities as limiting frequencies. Hartle argues from the perspective of a (mathematical) frequentist who believes that probabilistic statements acquire meaning as limiting frequencies in an infinite ensemble. He attempts to show, by taking limits as the number of systems goes to ∞, that an infinite product of a given state is an eigenstate of the frequency operator, with eigenvalue given by the usual quantum probability. He uses the eigenstate hypothesis to conclude that a measurement of the frequency operator must yield the quantum probability, thus establishing the quantum probability law as a consequence of the weaker eigenstate hypothesis. In this quantum derivation, one does have an advantage over trying to get classical probabilities as limiting frequencies, because the inner-product structure of Hilbert space provides a measure. Nonetheless, one still has the probability-one problem: even though an infinite-product state is an eigenstate of the frequency operator, this doesn't mean that the limiting frequency occurs with certainty (i.e., the eigenstate hypothesis fails for continuously infinite sets). Thus one can't escape the need to have a preëxisting notion of probability to interpret that being an eigenstate of the frequency operator means that the eigenvalue occurs with probability-one. Furthermore, one still has the problem of making quantitative statements about finite (actual) ensembles, since each finite ensemble will require yet another infinite hypothetical ensemble.

Note that Zurek's (1998a) objection to the Hartle argument misses the mark. Zurek contends that the frequency operator is a collective observable whose measurement has nothing to do with frequencies of "events" for the individual systems (I believe this was Chris's original objection). Zurek's conclusion applies to a direct measurement of the frequency operator on the joint system. The result of such a measurement is a particular frequency, and the measurement projects the joint system into the *subspace* corresponding to that frequency. In such a measurement there are no measurement results for the individual systems: it isn't known how the individual systems are ordered so as to give rise to the observed frequency. This is not, however, the only way to measure the frequency operator. Suppose one measures the individual systems in the relevant basis; in this case, if the state of the system is a frequency eigenstate, but not necessarily a product state in the relevant basis, then the frequency constructed from the measurement results will definitely be the frequency eigenvalue. Measurements on the individual systems provide more information than a direct measurement of the frequency operator—they give an ordered sequence of results, not just the frequency—but this should not obscure the fact that the frequency operator has a direct connection to frequencies of "events" on the individual systems. Formally, what is going on is that the frequency operator commutes with the product observable for the joint system; measuring the product observable removes the degeneracies in the frequency operator, thus giving more information than a direct measurement of the frequency operator, but certainly providing a measurement of the frequency operator.

c. *Consistent-histories and decoherent-histories interpretations.* Consistent histories (Omnès 1992a) and decoherent histories generalize the third question from being about which basis and which alternative in a basis to being about about which set of histories and which history within the set. As such, they are not interpretations so much as useful and instructive generalizations of the framework in which interpretations are considered. They generalize the incompatible bases that go with different sets of commuting observables to the different, incompatible sets of consistent histories.

Consistent- and decoherent-histories interpretations don't provide compelling answers to any of the four questions, particularly the third, which is just generalized from bases to histories. The consistent historians, Griffiths and Omnès, seem to grant simultaneous existence to all the incompatible sets of consistent histories and to hold that one history within each of the sets is realized. They explicitly eschew the need to explain actualization, and they don't seem to care whether all the sets of consistent histories correspond to worlds like our classical experience, being content to know that some set of consistent histories corresponds to a world like ours. The decoherent historians, Gell-Mann and Hartle, originally thought that decoherence would be sufficient to restrict decoherent histories to those that match our experience. When it didn't, they were left a bit out at sea, with no clear answer to any of the questions.

d. *Realistic, deterministic, causal interpretations. The Bohm interpretation.* In Bohmian mechanics, the phase of the wave function defines trajectories for particle position; these trajectories obey a Hamilton-Jacobi equation that is modified from the classical equation by the addition of a "quantum potential" that is determined by the magnitude of the wave function. This is just a re-write of wave mechanics. The Bohm *interpretation* promotes particle position to ontological status as the (hidden) reality underlying quantum mechanics, with the absolute square of the wave function giving probabilities for particle position.

   The interpretation of probabilities presents a problem for the Bohm interpretation. As discussed above, the natural probability interpretation for hidden-variable theories is the Bayesian interpretation, but then you have to explain how probabilities that are states of mind can push particles around via the quantum potential. A frequentist (ensemble) interpretation has the same problem: how is it that other members of the ensemble affect the motion of a particle through the action of the quantum potential? Notice that there are problems whether the ensemble is thought to be real or a theoretical construct, although they are different in the two cases. These problems have led Bohmians to speculate about how the "actual" probabilities might "relax" to the quantum probabilities, much as in relaxation to thermodynamic equilibrium, but these efforts have not been very convincing and, from a Bayesian perspective, are wholly misguided anyway. The best strategy for a Bohmian might be to adopt the idea that the probabilities are objective propensities; then they could push particles around.

   Though the ontology of the Bohm interpretation is superficially attractive, the realistic particle trajectories in the case of many interacting particles are highly nonlocal (they must be nonlocal for Bohmian mechanics to agree with the predictions of quantum mechanics for entangled states). The nonlocal influences will be along spacelike separations in a relativistic version of the theory and thus acausal. The whole world is thus connected together in an acausal web. This is Eastern reality stretched to nightmarish proportions—a reality that is completely disconnected from the reality of our everyday perceptions. Though this picture is inherent in the Bohmian reality, it remained an abstraction whose impact was not fully appreciated till the work of Englert, Scully, Süssmann, and Walther (1992a) [for me the most accessible version of these ideas has been given by Aharonov and Vaidman (1996a)]. They showed conclusively, for a simple example of two interacting systems, that the nonlocality of Bohmian mechanics means that the Bohmian trajectories have nothing whatsoever to do with the reality of everyday life.

   Now the Bohmian has a problem. The initially attractive ontology turns out to be useless for understanding everyday experience, so he will have to do as much hard work as anybody else—probably more to give an actual Bohmian account—to construct an emergent reality for the macroscopic world. In fact, he must carefully exclude the nonlocal, acausal aspects of the ontology from the emergent reality. Far from being an aid, the underlying trajectories are a serious nuisance, for he has to wipe out any trace of them and substitute in their place something that looks like the macroscopic world. In the Bayesian interpretation the problem is how to get an effective reality to emerge from a microscopic theory that doesn't have the objective properties of our everyday experience, whereas in the Bohm interpretation the problem is how to get a local, causal reality to emerge from the nonlocal web of Bohmian trajectories.

The Bohm interpretation accords position a special ontological status because our perceptions are so closely connected to location, but once one realizes that the Bohmian trajectories are irrelevant to our perceptions, the choice of position begins to smell like an arbitrary choice. Bohm-type theories can be constructed in other bases, e.g., the momentum representation (Brown and Hiley 2001a). As a result, Bohmian mechanics is far from unique as a foundation for a realistic interpretation, and the Bohm interpretation becomes just one possibility out of many.

The Bohmian trajectories are an example of what I call a *gratuitous reality*: they are pasted onto the theory because of an ingrained need for an ontology and the resulting habit of naïvely assigning reality to mathematical objects in the theory, but they are irrelevant to constructing and understanding the reality of our everyday experience. The defenses of Bohmian mechanics against the attack of Englert *et al.* (Englert 1992a) have focused on pointing out that the statistical prediction of Bohmian mechanics agree with those of quantum theory, but this misses the point. The point is that Englert *et al.* demonstrate that the ontology of Bohmian trajectories is a gratuitous reality that helps not at all in constructing the emergent effective reality of everyday experience.

e. *Many-worlds interpretations* (Vaidman 1999a, Wallace 2001a). Find wave-function collapse distasteful, so banish it. Make the most naïve realistic assumption: declare the wave function to be objectively real, and then—damn the torpedos!—plow straight ahead, undaunted by the mind-boggling consequences (well, actually, revel in them a bit; they make good press and great science fiction). That's the spirit of many-worlds interpretations. If you want to be perceived as a deep thinker without actually having to do any thinking, this is your interpretation.

The many-worlds interpretation posits a single, objective wave function for the entire Universe. Superpositions within some (arbitrarily chosen) basis correspond to branchings into different worlds, all of which are actualized. Fundamentally there are no probabilities in the theory, since all possibilities are actualized, but for subsystems where quantum probabilities are used to make predictions, there are attempts to derive those probabilities objectively from the Universal wave function. The world of everyday experience is the way it is because that's the way it is (thanks, Walter) on the branch we're on.

Talk about a gratuitous reality! This is the granddaddy of them all. To avoid a physical wave-function collapse, the many-worlds interpretation pastes onto quantum theory an unobserved and unobservable infinity of worlds that explains nothing about—it simply posits—the world we actually live in. As far as I can see, many-worlds interpretations provide no insight into any of the four questions, particularly the third, because the branching occurs in a basis chosen for no other reason than to give worlds that mirror our our macroscopic experience (Vaidman 1998a, Steane 2000a). Many-worlds interpretations, far from providing deep insights into the nature of quantum reality, are really founded on an inability to imagine anything except a naïve realistic world view, which is to apply in each branch.

Let's not be be too negative. There are deep thinkers who work on the many-worlds interpretation, and it has led to important insights, notably David Deutsch's idea of the quantum parallelism that is thought to provide the power behind quantum computers.* Still, I think that much of the current popularity of many-worlds is an example of what Richard Dawkins calls an "argument from personal incredulity": "I thought about it for a little while and couldn't figure out how to reconcile wave-function collapse with the Schrödinger equation, so I signed up with the many-worlders."

The attitude of many physicists might be summarized in the following way. Schrödinger formulated his equation, which is the essence of quantum mechanics; Born introduced the probability rule, and then von Neumann tacked on his ugly "collapse of the wave function," which interrupts the beautiful flow of Schrödinger evolution. We are taught that there are two kinds of evolution in quantum mechanics: the pristine evolution governed by Schrödinger's differential equation, and the *ad hoc* and unjustified collapse introduced by von Neumann. Phrased in these ways, our job as physicists is clear: find a way to get rid of the collapse. The Bayesian interpretation offers a useful corrective. From the Bayesian view the apparatus of quantum probabilities, including the updating that results from observation and that goes under the name of collapse, lies at the very heart of quantum mechanics. It's not the ugly part; it's the main part of the theory and certainly a beautiful part. Schrödinger's equation tells us how to update maximal information when a system is isolated from the rest of the world. That's important, too, as the place where physical laws find expression and thus perhaps as the objective part of the theory, but certainly not more important than the extraordinarily successfully quantum prescription for reasoning when "maximal information is not complete."

f. *Other interpretations.* Comments on other interpretations will be added as I learn enough about them to make the comments sensible.

7. Actualization and indeterminism vs. determinism. Which is better? An indeterministic world is certainly more interesting than a deterministic one, whose history is just a progressive revelation of the initial conditions. Moreover, if the world is intrinsically indeterministic, it means that the problem of actualization must lie outside the province of our theories. Omnès (1992a) provides a powerful and poetic account of this:

> Perhaps the best way to see what it is all about is to consider what would happen if a theory were able to offer a detailed mechanism for actualization. This is, after all, what the advocates of hidden variables are asking for. It would mean that everything is deeply

* To be fair, one should note that quantum parallelism might be a misleading way to think about quantum computation (Steane 2000a) and also that the other great advances in quantum computation—the Shor factoring algorithm, the Cirac-Zoller proposal for a semirealistic quantum computing system involving trapped ions, and the Shor-Steane realization that quantum error correction is possible—don't seem to have been motivated by a many-worlds perspective.

determined. The evolution of the universe would be nothing but a long reading of its initial state. Moreover, nothing would distinguish reality from theory, the latter being an exact copy of the former. More properly, nothing would distinguish reality from logos, the time-changing from the timeless. Time itself would be an illusion, just a convenient ordering index in the theory. . . . Physics is not a complete explanation of reality, which would be its insane reduction to pure unchanging mathematics. It is a *representation* of reality that does not cross the threshold of actuality. . . . It is wonderful how quantum mechanics succeeds in giving such a precise and, as of now, such an encompassing description of reality, while avoiding the risk of an overdeterministic insanity. It does it because it is probabilistic in an essential way. This is not an accident, nor a blemish to be cured, since probability was found to be an intrinsic building block of logic long before reappearing as an expression of ignorance, as empirical probabilities. Moreover, and this is peculiar to quantum mechanics, theory ceases to be identical with reality at their ultimate encounter, precisely when potentiality becomes actuality. This is why one may legitimately consider that the inability of quantum mechanics to account for actuality is not a problem nor a flaw, but the best mark of its unprecedented success.

8. A final note. Why even a statistical order in an indeterministic world? The Bayesian version of this question might be the following: Why should an intrinsically indeterministic, but supposedly complete theory—i.e., one in which maximal information is not complete—supply a unique rule for assigning probabilities in the case of maximal information? One can argue that it would be very unsatisfactory to have an indeterministic, but complete theory that failed to supply such a probability rule in the case of maximal information, there being no place outside the theory—it's complete!—to look for the rule. Unsatisfactory though it might be, however, it is hard to see why all theories would have this property.

Quantum mechanics, of course, obliges with the standard quantum probability rule, which follows just from applying Dutch-book consistency to probabilities that are faithful (i.e., noncontextual) to the Hilbert-space structure of quantum questions. The fact that the Hilbert-space structure so tightly constrains quantum probabilities that it gives a unique rule in the case of maximal information is certainly trying to tell us something very basic. Perhaps this tight constraint is the key feature of quantum mechanics, indeed the key to unlocking the ontology of quantum mechanics. In the Bayesian interpretation, there is a quantum reality that we are describing, in the best way possible, using the rules of quantum mechanics, but that reality is more subtle than the realist's direct, one-to-one correspondence between the theory (our model) and reality (what's out there). Perhaps the surprisingly constrained quantum probability rule is the first element of this Bayesian reality. The emergent effective reality would then be the second aspect. Demonstrating the emergence and consistency of the effective reality is a long-term goal of the Bayesian program.

REFERENCES (completely incomplete)

1. Y. Aharonov and L. Vaidman 1996a, "About position measurements which do not show the Bohmian particle position," in *Bohmian Mechanics and Quantum Theory: An Appraisal*, edited by J. T. Cushing, A. Fine, and S. Goldstein (Kluwer Academic, Dordrecht, The

Netherlands), pp. 141–154 (Proceedings of a Conference on Quantum Theory Without Observers, Bielefeld, Germany, July 24–28, 1995).

2. M. R. Brown and B. J. Hiley, "Schrödinger revisited: An algebraic approach" (`arXiv.org e-print quant-ph/0005026`).

3. C. M. Caves 1994a, "Information, entropy, and chaos," in *Physical Origins of Time Asymmetry,* edited by J. J. Halliwell, J. Pérez-Mercader, and W. H. Zurek (Cambridge University Press, Cambridge, England), pages 47–89 (Proceedings of the NATO Advanced Research Workshop on the Physical Origins of Time Asymmetry, Mazagón, Huelva, Spain, September 29–October 4, 1991).

4. C. M. Caves and R. Schack 1997a, "Unpredictability, information, and chaos," *Complexity* **3**(1), 46–57.

5. C. M. Caves, C. A. Fuchs, and R. Schack 2001a, "Unknown quantum states: The quantum de Finetti representation," submitted to *American Journal of Physics* (`arXiv.org e-print quant-ph/0104088`).

6. C. M. Caves, C. A. Fuchs, and R. Schack 2001b, "Making good sense of quantum probabilities" (`arXiv.org e-print quant-ph/0106133`).

7. B.-G. Englert, M. O. Scully, G. Süssmann, and H. Walther 1992a, "Surrealistic Bohm trajectories," *Zeitschrift für Naturforschung A* **47a**, 1175–1186.

8. R. N. Giere 1973a, "Objective single-case probabilities and the foundations of statistics," in *Logic, Methodology and Philosophy of Science IV*, edited by P. Suppes, L. Henkin, A. Joja, and Gr. C. Moisil (North-Holland, Amsterdam), pp. 467–483.

9. J. B. Hartle 1968a, "Quantum mechanics of individual systems," *American Journal of Physics* **36**, 704–712.

10. E. T. Jaynes 1957a, "Information Theory and Statistical Mechanics," *Physical Review* **106**, 620–630.

11. E. T. Jaynes 1957b, "Information Theory and Statistical Mechanics. II," *Physical Review* **108**, 171–190.

12. E. T. Jaynes 1986a, "Predictive statistical mechanics," *Frontiers of Nonequilibrium Statistical Physics*, edited by G. T. Moore and M. O. Scully (Plenum Press, New York), pp. 33–55 (Proceedings of a NATO Advanced Study Institute on Frontiers of Nonequilibrium Statistical Physics, Santa Fe, New Mexico, June 3–16, 1984).

13. E. T. Jaynes 1986b, "Monkeys, kangaroos, and N," in *Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by J. H. Justice (Cambridge University Press, Cambridge, England, 1986), pages 26–58 (Proceedings of the Fourth Maximum Entropy Workshop, University of Calgary, August 1984).

14. E. T. Jaynes 1990a, "Probability in quantum theory," in *Complexity, Entropy and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Redwood City, CA), pp. 381–403 (Proceedings of the 1988 Workshop on Complexity, Entropy, and the Physics of Information, St. John's College, Santa Fe, New Mexico, May 29–June 2, 1988; Santa Fe Institute Studies in the Sciences of Complexity, Vol. VIII).

15. J. L. Lebowitz 1999a, "Statistical mechanics: A selective review of two central issues," *Reviews of Modern Physics* **71**, S346–S357.

16. R. Omnès 1992a, "Consistent interpretations of quantum mechanics," *Reviews of Modern Physics* **64**, 339–382.

17. A. M. Steane 2000a, "A quantum computer only needs one universe," unpublished (`arXiv.org e-print quant-ph/0003084`).

18. L. Vaidman 1999a, "On schizophrenic experiences of the neutron or why we should believe in the many-worlds interpretation of quantum theory," *International Studies in the Philosophy of Science* **12**, 245–261.

19. D. Wallace 2001a, "Worlds in the Everett interpretation" (`arXiv.org e-print quant-ph/0103092`).

20. W. H. Zurek 1998a, "Decoherence, einselection and the existential interpretation (The rough guide)," *Philosophical Transactions of the Royal Society A* **356**, 1793–1821.

My two heroes of Bayesian probabilities are Bruno de Finetti and, especially, Edwin T. Jaynes, the former a fascist and the latter a very conservative Republican. I certainly don't equate the two, but this does give me pause, though it probably shouldn't. It just confirms that a person's politics are a very poor guide to a his/her value as a scientist or even as a person. To those concerned, Fuchs, Schack, and I offer Bayesian role models from other parts of the political spectrum.