Probability in the Everett interpretation

Hilary Greaves*
Rutgers University

Abstract
The Everett (many-worlds) interpretation of quantum mechanics faces a *prima facie* problem concerning quantum probabilities. Research in this area has been fast-paced over the last few years, following a controversial suggestion by David Deutsch that decision theory can solve the problem. This article provides a non-technical introduction to the decision-theoretic program, and a sketch of the current state of the debate.

Full text:

1 The problem of probability

1.1 Introduction

Quantum mechanics, as usually understood, is an indeterministic theory; it makes probabilistic predictions. For example: if one feeds into a Stern-Gerlach device an electron in a nontrivial superposition of spin-up and spin-down, $|\psi\rangle = a|\uparrow\rangle + b|\downarrow\rangle$, the 'textbook' predictions are that the outcome will be either spin-up or spin-down, with *probabilities* given by $|a|^2, |b|^2$ respectively.

The Everett, or many-worlds, interpretation gives a different account of quantum measurement. Indeterminism enters the 'textbook' story at the point of *wavefunction collapse*: when the measurement is performed, the wavefunction is supposed to collapse (indeterministically) into one or the other eigenstate of the observable being measured. But the many-worlds interpretation denies that collapse occurs. Instead, according to the many-worlds interpretation, when the measurement is performed, the apparatus, and the observer, and the laboratory, and indeed the entire world, will split into two copies, or 'branches'. There will be one branch on which the spin-up outcome occurs (and a future copy of the observer records spin-up), and there will be another branch on which spin-down occurs (and a future copy of the observer records spin-down). (For an accessible introduction to quantum mechanics and the measurement problem, see Albert (1992). For brief introductions to the many-worlds interpretation, see Vaidman (2002) and Barrett (2003).)

The motivations for the many-worlds interpretation include (i) the fact that (if defensible) it solves the measurement problem, (ii) compatibility with the spirit of special relativity, and (iii) Ockham's razor. (The latter remark will be explained below.) But the interpretation has a *prima facie* problem in making sense of quantum probabilities, for the following reason: according to the many-worlds interpretation, *all outcomes actually occur*. That is: for each outcome that the textbooks would describe as 'possible', the many-worlds interpretation predicts that there will (certainly!) be some branch on which that outcome occurs. It is then difficult to see how one can make sense of assertions like 'the probability of spin-up will be 2/3.' But neither does it seem acceptable just to throw out the probabilistic part of the theory: strip quantum mechanics of its quantitative probabilistic assertions, and one has neither a theory that will be any help in guiding expectations, nor a theory that can be subjected to empirical test.

It will be useful to distinguish two aspects of the problem more sharply. We actually have *two* problems of probability:

The incoherence problem: How can it make sense to talk of probabilities (other than 0 and 1) at all, since all 'possible' outcomes actually occur?

The quantitative problem: Insofar as it does make sense to talk of nontrivial probabilities for branches, how can the probabilities in a many-worlds interpretation agree with those of textbook quantum mechanics?

1.2 A brief history

Since the seminal (1957) paper by Hugh Everett III, 'the Everett interpretation' has been developed in many ways; correspondingly, many different approaches to Everettian probability have been investigated. It will be useful to review three of these, before turning to the interpretation favored by most modern-day 'Everettians'.

Everett (1957). Everett noted that it will often be useful to talk about what happens on 'most' branches[1], and that such talk presupposes the existence of a preferred measure over the set of branches. Everett's proposal is that the measure one should use, for the purposes of such talk, is the "mod-squared measure", i.e. precisely the measure that agrees with the usual quantum-mechanical probabilities:

Mod-squared measure: if ψ is a Hilbert space vector of unit norm, and $a\psi$ is a branch ($a \in C$), then the mod-squared measure of $a\psi$ is $|a|^2$.

However, nothing is said about the *meaning* of this measure (the incoherence problem).

Deutsch (1985). David Deutsch suggests that, to solve the incoherence problem, the ontology of the many-worlds interpretation needs to be supplemented. In addition to the quantum state of the universe, we are to postulate a continuously infinite set of universes, together with a preferred measure on that set. The measure is such that, when a measurement occurs, the proportion of universes in the original branch that end up on a given branch is given by the mod-squared measure of that branch. Observers can then be *uncertain* about which outcome will occur *in the universe they are in*.

Albert and Loewer (1988). David Albert and Barry Loewer explore the possibility of a 'many minds' (as opposed to 'many worlds') interpretation of quantum mechanics. They postulate, not a continuous infinity of worlds, but a continuous infinity of *minds* associated with 'every sentient physical system, every observer' (*ibid*, p.206). Albert and

Loewer then suggest that the time-evolution of an individual mind is probabilistic: for each mind of the observer preparing a Stern-Gerlach experiment, there is a chance of $|a|^2$ that *that mind* will end up in a state of seeming to have observed spin-up. Again, this solves the incoherence problem.

(For a helpful summary of the history of Everettian probability, marginally less brief than the one I present here, see Saunders (1993), pp.134-6.)

1.3 Many worlds via decoherence

We have reviewed this history only for contrast; it is time now to the modern-day 'many-worlds interpretation'.

The idea here is to add nothing, at the level of fundamental ontology, to the quantum state of the universe ('universal wavefunction') – no continuous infinity of minds, and no worlds either. Instead, we have a 'two-level' proposal:

Fundamental ontology: The quantum state of the universe is represented by a Hilbert space vector, $|\psi\rangle$.

Supervenience of the classical on the quantum: The way to relate the 'quasiclassical' worlds of familiar experience to the quantum-mechanical ontology is as follows:

1. Decompose the quantum state into a set of (approximately) *consistent histories*, or, equivalently, expand the quantum state at each time as a superposition (in some basis).
2. Identify quasiclassical worlds with individual histories, or, equivalently, individual elements of the expansion.

Step (2) has the consequence that, after the Stern-Gerlach experiment, it is *not* the case that we have *a* world in which *the* apparatus is in an (improper) mixed state of pointing to 'up' and pointing to 'down'. Rather, we have *two* worlds, in each of which the pointer is in a definite state, *either* of pointing to 'up', *or* of pointing to 'down'. This is how the many-worlds interpretation reconciles our experience (specifically: the fact that we never see pointers in mixed states of pointing to the left and pointing to the right) on the one hand, with the absence of collapse on the other.

In defense of this 'two-level' proposal, we note that the key idea is one that is already present in classical physics: *at the level of fundamental ontology* in classical physics, there are no tables – there are only particles – but still there is a perfectly good sense in which tables 'emerge', or count as objects at a higher level of description, in a classical-mechanical universe.

The interpretive program sketched above, however, faces a preferred *basis problem*: in virtue of what is one set of (approximately consistent) histories, or one basis for expansion, any better than the others? The modern-day many-worlds theorist appeals, at this point, to the theory of decoherence. The idea here is to argue along the following lines: given the sorts of Hamiltonians that are realized in the actual world, we expect one (approximately defined) basis to be ‘preferred’ in the sense that only when the universal state is expanded in that basis will it be possible for stable higher-level structures (such as tables and human beings) to emerge within individual elements of the superposition. It remains formally possible to expand the universal state in a basis that differs significantly from the decoherence-preferred basis, but there will be no non-gerrymandered objects localized in elements of such a purely formal expansion; in particular, since *we* are non-gerrymandered objects, there is no puzzle about why we experience only objects that are approximately localized relative to the decoherence-preferred basis. (The technical details of the theory of decoherence, and the arguments concerning *precisely* how these details are supposed to solve the preferred basis problem, are somewhat involved: see, e.g., Wallace (2002, 2003a) and references therein.)

The decoherence-based many-worlds interpretation enjoys significant advantages over earlier Everett-style interpretations. Objections to Deutsch's 'many-worlds' and Albert and Loewer's 'many-minds' proposals include charges of ontological extravagance (all these worlds/all these minds!), vagueness (where are the *axioms* specifying the basis in which 'splitting' occurs? (Kent 1990)), and, in the case of the many-minds interpretation, mind-body dualism. All of these charges evaporate once one moves to the decoherence-based approach. There is no 'ontological extravagance' in any offensive sense, since nothing is being added at the fundamental level. (Indeed, the fundamental ontology of this sort of many-worlds interpretation is actually more parsimonious than that of the rival 'hidden-variables' approaches; cf. the remark about Ockham's razor in section 1.1.) Vagueness is acceptable since the 'many worlds' appear at the emergent level rather than the fundamental level, and supervenience relations are always vague. 'Minds' play no special role in the theory, so dualism is not required.

All of this is great progress. But it comes at a price: in the move from Deutsch's or Albert and Loewer's proposals to the decoherence-based approach, the problem of probability has returned. We have an (emergent) account of worlds 'splitting' in the future direction of time, each outcome will occur (on some branch), and there doesn't seem to be anything that the probabilities can be probabilities *of*.

2 The Everettian Representation Theorem

The problem of probability in the (decoherence-based) many-worlds interpretation seemed utterly intractable until David Deutsch (1999) suggested understanding probability in terms of rational action. Specifically, Deutsch claimed to 'prove', via decision theory, that the 'rational' agent who believes she lives in an Everettian multiverse will nevertheless 'make decisions as if' the mod-squared measure gave chances for outcomes.

This claim is highly controversial, and continues to receive a significant amount of attention in the philosophy of physics literature. Advocates claim that Deutsch's proposal removes the last obstacle to accepting a many-worlds interpretation; critics contend either that the decision-theoretic result is unsound, or that it is insufficient to give the Everettian everything she needs from 'quantum probability'.[2] In this section, we sketch the spirit of decision theory in general, and Deutsch's proposal in particular.

Decision theory (Savage). Decision theory is a theory designed for the analysis of rational decision-making under conditions of uncertainty. One considers an agent who is uncertain what the *state* of the world is: for example, she is uncertain whether or not it will rain later today. The agent faces a choice of *acts*: for example, she is going for a walk, and she has to decide whether or not to take an umbrella. She knows, for each possible state of the world and each possible act, what will be the *consequence* if the state in question obtains and the act in question is performed: for example, she knows that, if she elects not to take the umbrella and it rains, she will get wet. The agent is therefore able to describe each of her candidate acts as a function from the set of possible States to the set of Consequences.[3]

Decision theory then places a set of *rationality constraints* over the set of Acts. For example, the preferences of a rational agent must be transitive: if an agent prefers act *A* to act *B*, and prefers act *B* to act *C*, then the same agent must prefer act *A* to act *C*. Less trivially, we require the agent's preferences to satisfy, for example, an axiom of Dominance: roughly, that if act *A* and act *B* give the same consequence on some subset of possible states, and act *A* results in better consequences than act *B* on the remaining subset ('better' according to the agent's own preferences), then the rational agent prefers *A* to *B*.

The decision theorist then proves a *representation theorem*: it can be shown that, for any agent whose preferences over acts satisfy the given rationality constraints, there exists a unique probability measure *p* on the set of States, and a utility function *U* on the set of Consequences (unique up to positive linear transformation), such that, for any two acts *A*, *B*, the agent prefers *A* to *B* iff the expected utility of *A* is greater than that of *B*. Here, the expected utility of an act *A* is defined by:

$$EU(A) := \sum_{s \in S} p(s) \cdot U(A(s)) .$$

This result guarantees an operational role for subjective probability: any *rational* agent will (at least) *act as if* she is maximizing expected utility with respect to some probability measure.

Everettian application of decision theory (Deutsch). Deutsch's idea was to apply a similar approach to the case of a rational agent choosing among quantum games (or quantum bets). The conceptual side of Savage's decision theory must be adjusted slightly:

- 'States': To specify a quantum bet, we consider, not the set of possible states of the world, but the set of future branches that will (certainly!) result from a given quantum measurement, on the assumption that Everettian quantum mechanics is true. For example, if an Everettian agent knows that a Stern-Gerlach experiment is to be performed on an electron initially in a superposition $|\psi\rangle = a|\uparrow\rangle + b|\downarrow\rangle$, then she knows that after the measurement, there will be one branch (of weight $|a|^2$) on which spin-up ($\uparrow$) occurs, and one branch (of weight $|b|^2$) on which spin-down ($\downarrow$) occurs. The set of states would then be given by $\{\uparrow, \downarrow\}$.

- 'Consequences': Naive application of Savage's theory to a case of Everettian branching might suggest that the 'Consequence' of an action is the entire branching structure to the future of the decision. Crucially, this is not the route taken by Deutsch's approach: instead, we identify the 'Consequences' of decision theory with things that happen to individual future copies of the agent, *on particular branches*.
- 'Acts': An act is still a function from States to Consequences. Here, corresponding to our identifications of States with branches and Consequences with rewards-on-branches, this means: an act is an assignment of rewards to branches. For example, if the agent pays \$1 to enter an agreement on which she receives \$3 on the spin-up branch and \$0 on the spin-down branch, then the act she performs is given by:

$$A(\uparrow) = \$3 - \$1 = \$2$$
$$A(\downarrow) = \$0 - \$1 = -\$1.$$

We further introduce the notions of a *chance setup* and a *quantum game*:

- 'Chance setup': A *chance setup* is a pair $\langle |\psi\rangle, \hat{X} \rangle$, where $|\psi\rangle$ is the quantum state of the object system to be measured, and $\hat{X}$ is the hermitian operator representing the physical property to be measured. (So: a chance setup generates a set of branches, together with a natural measure on that set, given by the mod-squared rule. For example, if $|\psi\rangle$ is the spin superposition $|\psi\rangle \equiv a|\uparrow\rangle + b|\downarrow\rangle$ and $\hat{X}$ is the spin measurement operator $\hat{X} \equiv \frac{1}{2}|\uparrow\rangle\langle\uparrow| - \frac{1}{2}|\downarrow\rangle\langle\downarrow|$, then the chance setup will generate a set containing a spin-up branch of natural measure $|a|^2$, and a spin-down branch of natural measure $|b|^2$.)
- 'Quantum game' or 'quantum bet': A *quantum game* consists of a chance setup, together with an act or 'payoff function' *P* associating a Consequence with each eigenvalue of the operator $\hat{X}$ (i.e., with each future branch). A quantum game is therefore specified by a triple $\langle |\psi\rangle, \hat{X}, P \rangle$.

By imposing a set of rationality constraints on agents' preferences among such quantum games, Deutsch is able to prove a representation theorem that is analogous in many respects to Savage's: the preferences of a rational agent are representable by a probability measure over the set of States (branches) for every possible chance setup, and a utility function on the set of Consequences (rewards-on-branches), such that, for any two Everettian acts *A*, *B*, the agent prefers *A* to *B* iff *EU(A)>EU(B)*. (Expected utility is defined via the same formal expression as above.) This result, if sound, seems (perhaps) to solve the incoherence problem: the rational Everettian agent *acts as if* she regarded her multiple future branches as multiple possible futures.

There is, however, a further, quantitative feature of Deutsch's result that has no analog in Savage's representation theorem. For Savage, two ideally rational agents may be represented by *different* probability measures on the set of States. (For any given agent, the probability measure is uniquely fixed by that agent's preferences. But two rational agents may have different sets of preferences, each consistent with the rationality constraints.) This freedom to disagree is not permitted by Deutsch's result: according to the latter, once the chance setup is specified (so that the set of branches and the mod-squared measure of each future branch have been fixed), *the probability measure over that set of branches that represents the preferences of any rational agent must be the mod-squared measure*. Call this result the **Everettian Representation Theorem (ERT)**.

Deutsch suggests that this result captures, within the Everett interpretation, everything that is important about quantum probability:

> "I shall prove that [the rational agent] necessarily makes decisions as if [the assertion that the mod-squared measure gives probabilities] were true. I take this to be the effective meaning of [that assertion]." (Deutsch (1999), p.2)

3 Problems with the decision-theoretic approach

This result strikes many people as a 'rabbit out of a hat' trick – how did it happen, and how can it be right? There are three (*prima facie*) reasons to be suspicious of the Everettian Representation Theorem:

1. Decision theory is supposed to apply to decision-making under *uncertainty*. How, then, can its application to deterministic branching, in which the agent is certain that all 'possible' outcomes will actually occur, be justified?
2. As noted above, decision theory does not usually fix a *unique* probability measure over the set of States; it is usually possible for two ideally rational agents to have different probability measures. To obtain this feature in the Everettian case, Deutsch and Wallace must therefore be adding some axioms that have no analogs in Savage's theory. How can these additional axioms be justified?
3. The decision-theoretic result, even if sound, guarantees only that a rational agent who believes the Everett interpretation will *act as if* the Everettian branch weights are probabilities. Is this really all we need from quantum probability?

We will discuss each of these in turn.

3.1 Justifying the application of decision theory

Let us elaborate on our first problem. In conventional applications of decision theory, one starts from the assumption that the agent is uncertain which State obtains; one then applies the theory, and concludes that (if the agent is rational, then) this uncertainty is quantifiable by some probability measure. It is then (relatively) clear how the probability measure is to be interpreted: the probability measure gives the agent's *degrees of belief*, for each subset of S, that some state in that subset obtains.

The problem is that, at first sight, none of this seems to make sense in Deutsch's application of decision theory to Everettian branching. The agent (it seems) is *not* 'uncertain which outcome will occur'; she *knows* that *all* outcomes will occur. Correspondingly, even if (as seems implausible, in the absence of the appropriate interpretation) we could find some reason for crediting Deutsch's axioms with the status of 'rationality constraints', it is unclear what the resulting probability measure could mean. It cannot be quantifying the agent's degrees of belief in the corresponding outcomes, since all such degrees of belief are 1; how, then, could we understand the prescription to maximize 'expected utility' with respect to such a measure?

3.1.1 The 'subjective uncertainty' (SU) program

The response of the 'subjective uncertainty' (SU) program is to assert that, contrary (perhaps) to initial appearances, the set of future branches in a 'deterministic' branching situation *is* a locus of uncertainty for the agent. The agent performing a Stern-Gerlach measurement, for example, knows that (in some sense) all branches will be physically real, but still (in another sense) is *uncertain about whether the result of the experiment will be spin-up or spin-down*.

There are two ways of fleshing out this proposal:

- **Subjective-uncertainty (splitting).** There is one world before the measurement is performed, and this splits into two copies when the measurement occurs. (In the language of the 'consistent histories' formalism: a world, at a given time, is represented by a projector associated with that time.) However, when a world at a given time has multiple futures, with different things happening in different futures, there is *no fact of the matter* about what will happen in 'the' future of the world in question. In our example, it is not true that spin-up will occur, and it is not true that spin-down will occur. Our agent knows this: so, she does not believe that spin-up will occur, and she does not believe that spin-down will occur. However, she *does* believe that spin-up-or-spin-down will occur: this occurs on every branch to her future, and she knows that. She is *uncertain* as to whether the result will be spin-up or spin-down. (This proposal is much in the spirit of the Aristotelian attitude to future contingents: "there will be a sea-battle tomorrow" is neither true nor false, yet I am uncertain as to whether there will be a sea-battle tomorrow.)

- **Subjective-uncertainty (divergence**). There are two worlds all along: a world is a maximal sequence of macroscopically distinct world-stages. (In consistent histories language: a world is represented by a *history*, i.e. a time-ordered sequence of projections, extending both into past and future.) In particular, there are two copies of our agent all along, even before the measurement. However, while those agents both know all of the above, they both have self-locating ignorance: neither of them knows which of the two agents she is. Each of them, that is, is uncertain as to whether she is the copy in the spin-up world, or the copy in the spin-down world. It follows, of course, that each of them is uncertain about whether it is spin-up or spin-down that will occur (in the world she is in).

For further discussion of precisely how the subjective-uncertainty proposal might be filled out, see Wallace (2005c).

If (either version of) the subjective-uncertainty thesis is true, then our first problem dissolves: decision theory applies to cases of uncertainty, Everettian branching is a case of uncertainty, and so decision theory applies to Everettian branching.

3.1.2 Justification without subjective uncertainty

This quick response only works, however, if (some version of) the subjective-uncertainty thesis *is* true. Two points should be noted:

- The subjective-uncertainty thesis is highly controversial. (For objections, see Greaves 2004 and Lewis 2006. For a helpful review of most of the debate over SU, see Wallace (2005b, secs. 3.3-3.5). For a more extended exposition of Wallace's own argument for SU, see Wallace (2005c).)
- Some of the Everettian claims (specifically: the claim that a *better* account of probability can be provided on the Everett interpretation than on other interpretations; see sections 3.2 and 4.2 below) rely crucially on the *differences* between a branching universe and a non-branching one. These differences are obliterated by the language of subjective-uncertainty. It therefore seems that, even if the SU thesis is correct, any claim that Everettian probability is in a better state than non-Everettian probability must rely on the fact (if it is a fact) that the decision-theoretic approach can be justified independently of SU.

We must therefore ask whether or not the decision-theoretic approach *can* be justified without appealing to subjective-uncertainty.

The issue is whether or not our usual reasons for accepting the decision-theoretic axioms, in cases of uncertainty, have equally compelling analogs in the branching case. So we must try to identify exactly what those reasons are. Consider, for example the Dominance axiom alluded to above. Recall: This says, roughly, that if act *A* and act *B* give the same consequence on some subset ($S_1 \subset S$, say) of possible states, and act *A* results in better consequences than act *B* on the remaining subset $\overline{S_1} = S - S_1$ ('better' according to the agent's own preferences), then the rational agent prefers *A* to *B*.

Why does rationality require conforming to Dominance, when the set of States is a locus of uncertainty? Well: it is difficult to 'justify' such obvious truths, but one might reason as follows. Either the actual state *s* is in S_1, or *s* is in $\overline{S_1}$. In the former case I get the same reward regardless of which act I perform; in the latter case I will do better if I choose act *A* than if I choose act *B*. So, in choosing *A* over *B*, I will not lose, and I might gain.

Suppose now that the set of States is interpreted, not as a locus of uncertainty, but as a set of coexisting branches. Then an exactly parallel justification for Dominance can be given, as follows. The world of which I am now a part has multiple futures, given by the elements of S. In some of these (viz. those in S_1), I get the same reward regardless of which act I perform. In others (viz. those in $\overline{S_1}$), I do better if I chose act *A* than if I chose act *B*. So, in choosing *A* over *B*, I will not lose on any branch, and I will gain on some branches.

Is the justification of Dominance more compelling, or in any other important sense better, in the uncertainty case than in the branching case? As far as I can see, the answer is 'no'; in that case, the decision theory required for the ERT is just as well-justified without subjective-uncertainty as with. The result is that, even if she does *not* regard the future as uncertain, the rational Everettian agent maximizes expected utility with respect to some probability measure over branches.

It does not, of course, make sense to call this measure a measure of the agent's *uncertainty* if we are rejecting subjective-uncertainty, so, talk of 'degrees of belief' seems inappropriate. We might instead call it the agent's 'caring measure', since the measure quantifies the extent to which (for decision-making purposes) the agent cares about what happens on any given branch. (This proposal is advocated by Greaves (2004); a similar suggestion is made by Vaidman (2002).)

3.2 Justifying the additional axioms

3.2.1 Measurement Neutrality (MN) and Equivalence

Suppose, then, that (with or without subjective uncertainty) we accept the applicability in principle of decision theory to cases of deterministic branching. Then it is unsurprising that the rational agent maximizes expected utility with respect to *some* measure over branches (interpreted either as a measure of the agent's degree of belief that any given branch will be actual, or as a measure of the degree to which the agent cares about which reward is received on any given branch). But how can we force the measure in question to be the mod-squared one, where Savage's decision theory achieves no such thing?

This extra feature of the ERT can be traced to the assumption of 'Measurement Neutrality' (or, equivalently, 'Equivalence').[4] These assumptions are as follows:

Measurement Neutrality:
A rational agent is indifferent between any two quantum bets that agree on the state $|\psi\rangle$ on which the measurement is to be performed, the observable $\hat{X}$ to be measured, and the 'payoff function' P from the spectrum of $\hat{X}$ to the set of consequences.

Equivalence:
A rational agent is indifferent between any two quantum bets that agree, for each possible reward, on the mod-squared measure of branches on which that reward is given.

For a (highly illuminating) proof that MN implies Equivalence, see Wallace (2005a, section 7). The converse is a trivial consequence of quantum mechanics. (Two measurements of the same quantity $\hat{X}$ on the same initial state $|\psi\rangle$ generate the same set of branches and the same mod-squared measure on that set, regardless of the details of the measurement process; if the agent is indifferent between *any* two bets that agree on the total mod-squared measure of each reward, she is, in particular, indifferent between any two bets that agree on the total mod-squared measure of each reward and further agree on the chance setup $\left\langle |\psi\rangle, \hat{X} \right\rangle$ that led to the branching event and the reward assignment.)

3.2.2 What MN/Equivalence rules out

The clearest way to assess the plausibility (or otherwise) of these axioms is to consider, by way of a few examples, what they rule out.

Here are some decision-making strategies that conform to the 'maximize expected utility' injunction (and so satisfy the Everettian analogs of the usual Savage axioms), but that employ a probability measure other than the mod-squared measure (and violate Measurement Neutrality/Equivalence):

- **Naive Counting**: All branches are to be assigned equal measure.
- **Eigenvalue Relevance**: The measure is given by assigning, to each branch, the absolute value of the measurement eigenvalue corresponding to the branch in question (and renormalizing). (So, for example: I should always 'have degree of belief zero that the outcome will be eigenvalue 0', or I should always 'have caring measure zero for any eigenvalue-0 branches'.)
- **Relevance of Socks**: The measure of a given branch is to be set to zero if the agent possesses an odd number of socks on that branch 10 minutes after the measurement; otherwise the mod-squared measure is to be used (and the resulting measure renormalized). (So: if, for example, I am almost certain that on the spin-up branch all but one of my socks will be burnt 5 minutes after the measurement, I should 'be almost certain that spin-up will not occur', or I should 'care next to nothing about the spin-up branch'.)

Here is a fourth possibility, of a different character. Rather than attempting to specify some particular alternative measure, we simply assert

- **Full permissivism**: *Any* probability measure on the set of possible branches is rational.

3.2.3 Assessing MN/Equivalence

The question is then: do the above MN/Equivalence-violating policies count as rationally permissible (and hence as counterexamples to the claim that MN/Equivalence is a rationality constraint), or can they be ruled out as irrational (and MN/Equivalence upheld)?

Let us put full permissivism aside for the moment. (I return to it in passing in the last three paragraphs of this subsection; cf. also section 4.1.) Let us address each of our three specific putative counterexamples in turn; there are interesting differences between the three.

Consider, first, **Eigenvalue Relevance**. It is not too difficult to see that this policy is actually incoherent: it would have the agent's preferences between quantum games turn on arbitrary features of particular descriptions of those games, rather than physical features of the games themselves. Any measurement of an operator $\hat{X} = \sum_i a_i |i\rangle\langle i|$ is equally a measurement of any other operator of the form $\hat{X}' = \sum_i b_i |i\rangle\langle i|$; there is no fact of the matter as to whether 'the eigenvalue' attached to the branch $|i\rangle$ is given by a_i or b_i; the eigenvalues are mere arbitrary labels.

Naive Counting is incoherent too, but for subtler reasons.[5] Here it is essential that by 'many-worlds interpretation' we mean the *decoherence*-based MWI, not an old-style MWI in which worlds are added to the formalism as an additional ontological primitive. (In the latter, naive counting would be coherent if, *contra* Deutsch's own (1985) proposal, we postulated a *finite* number of 'worlds'.)

Why does naive counting break down in the decoherence approach? The core of the problem is that naive counting, too, presupposes the existence of a piece of structure that is not in fact present in the theory. We can truly describe (e.g.) the situation after a Stern-Gerlach experiment as one in which there are 'two' branches, and for many purposes it will be useful to do so, but there is no more going on here than when one (truly) says "There are two mutually exclusive and jointly exhaustive subsets of the real line: $(-\infty, 0)$ and $[0, +\infty)$." In both cases, there are arbitrarily many alternative descriptions that are equally true. It is therefore a condition of adequacy on any suggested measure that the suggestion be invariant under such redescriptions; the naive counting measure fails to meet this condition. (To put this point another way: to insist on 'naïve counting' is to apply a Principle of Indifference in a way that leads to paradox.)

Let us elaborate slightly on the preceding paragraph. What we really have (at the quantum level of description) is the quantum state of the universe. We decompose this into a set of branches, for the purposes of relating the quasiclassical level of description to the quantum level of description. But the criteria of adequacy of the decomposition (cf. section 1.3) allow for a certain amount of vagueness as to which is the 'right' set of histories. In particular:

- The history space can be coarse-grained and fine-grained, without losing decoherence. There is no fact of the matter as to which level of coarse-graining 'really' carves the quantum state at the world-joints. But coarse-graining and fine-graining can drastically alter the naive counting measure.
- The decoherence basis can be rotated slightly in Hilbert space, without losing (approximate) decoherence. There is no fact of the matter as to exactly which basis 'really' matches the world-joints. But such slight rotations can drastically alter the naive counting measure.

This point is of crucial importance, and often insufficiently appreciated; it rules out the vast majority of suggested 'counterexamples' to the Born rule in the Everettian context. (For further explanation and discussion, see Wallace (2005a, p.21) and/or Greaves (2004, sec. 5.3.)

It might be objected that these objections to Eigenvalue Relevance and Naïve Counting show only that the policies are as yet *underspecified*; we could complete them, on a case-by-case basis, by stipulating that on *this* occasion we intend to use *that* set of eigenvalues/*that* decomposition into branches for the purposes of defining our subjective probability measure, and note that we have thereby violated MN/Equivalence without implicitly committing ourselves to any contradictions. This is perfectly correct, technically speaking. However, insofar as Eigenvalue Relevance and Naive Counting were supposed to be any advance on general permissivism, the objection is beside the point. It was always known that, for any branching multiverse, there formally *exist* measures other than the mod-squared measure over the set of branches. The potential interest of Eigenvalue Relevance and Naïve Counting lay in the claim that there exist alternative measures that are (i) reasonably natural, and (ii) simply specifiable. The underspecification problem disposes of this claim. (The lingering optimist is invited to try completing the specification!)

What of **Relevance of Socks**? Well, plausible or otherwise, there is nothing incoherent about this policy. The policy's use of the mod-squared measure (as opposed to any implicit reliance on counting) ensures that it is appropriately robust under relabelings of branches, coarse-and fine-graining of the history space, rotations of the decoherence basis, etc.

One might object that **Relevance of Socks** is crazy. This is correct, but it is far from clear that that can be taken as an objection in the present context: according to the weak notion of 'rationality' with which decision theory usually deals, the agent is permitted to have an arbitrarily 'crazy' *utility function*, so why not a 'crazy' probability measure, too?

To answer this question would require examination of the positive arguments that can be offered in defense of Equivalence. A review of these is outside the scope of this article. (For defense of MN and/or Equivalence, see, e.g., Wallace (2005a, secs.7-8), Saunders (2004, sec.5.3).)

For the sake of argument, though, suppose that the arguments in question are unpersuasive. Then we cannot convince an agent who is determined to set his degrees of belief/'caring measure' according to **Relevance of Socks** that such a course of action is irrational. But how bad is that? We cannot convince an agent who is certain that she is being spied upon by aliens that such belief is irrational, either; this does not prevent the rest of us from considering ourselves more rational than she, or from acting on our own beliefs.

The right attitude seems to be the following. Given that the actual physical state of the universe is the physical state, the measure given by using the mod-squared prediction in conjunction with that state is overwhelmingly more *natural* than the measure given by Relevance of Socks (why socks, rather than pants?), or any of the arbitrarily complicated measures allowed by full permissivism. This makes acceptance of the Born rule more rational than acceptance of any rival branch measure. (We had better not spurn such 'naturalness' requirements: without them, decision theory is trivialized.) So *something* can be said in favor of the mod-squared measure. It is true that we cannot win over a really die-hard advocate of **Relevance of Socks**; but this does not prevent the rest of us from considering ourselves more rational than he, and acting according to the measure we ourselves accept. Furthermore, for agents (like the author and, I hope, the reader) who *do* accept the mod-squared measure in both non-Everettian and Everettian contexts, the choice between an Everettian and a non-Everettian interpretation of quantum mechanics will not affect everyday decisions.

3.3 Is decision-theoretic probability enough?

We turn to our third problem for the decision-theoretic approach to Everettian probability: even if it is *true* that a rational Everettian agent maximizes expected utility with respect to the mod-squared measure over branches, is this *enough* to give us everything we need from Everettian probability?

Here is a reason to think it might not be enough. One role that 'physical probability' is supposed to play is, indeed, the practical role captured by decision theory: if one believes a theory that assigns high probability to *X*, then one has high degree of belief that *X* will happen, and one acts accordingly. But there is also a second, epistemic, role: a theory's assignment of probabilities to possible outcomes is intimately involved in the process of *theory confirmation*. We normally consider a theory to be empirically confirmed if we have observed results that the theory in question would have predicted *with high chance*; Everettian quantum mechanics would have predicted the results we have in fact observed with high 'branch weight'; it is not immediately obvious that this is good enough. One might worry, therefore, that while the decision-theoretic result establishes how one should act if one has (somehow!) come to believe that Everettian quantum mechanics is true, it says nothing about whether or not one *should* believe that Everettian quantum mechanics *is* true. Call this the 'epistemic problem' for Everettian quantum mechanics.

Conditional on the assumption that the subjective-uncertainty program is viable, the worry that Everettian probability might play the practical role that 'physical probability' normally plays, while not also playing the epistemic role, is rather far-fetched. According to subjective-uncertainty, there is no relevant difference between the Everett interpretation and any other probabilistic theory: both sorts of theory give us a set of physically possible futures, and give the rational subjective likelihoods of each possibility's being actual. It is then extremely hard to see how the link between the practical and epistemic roles of probability – whatever exactly that link is! – can hold in non-Everettian cases, while not also holding in the Everettian case.

If the subjective-uncertainty program is rejected, however, this quick response is not available. It does not follow, of course, that the epistemic problem *cannot* be solved without subjective-uncertainty; what follows is that further argument is required. (See

Wallace (2005b), sec. 4.2, for a suggestion that the prospects look bleak; see Greaves (2007) for an attempt to provide the required further argument.)

4 The goalposts

4.1 Many worlds are no worse than one

In a theory that postulates non-branching universes (e.g., in the quantum context, a stochastic collapse or hidden-variable theory), the treatment of objective probability has the following structure:

1. The theory postulates a set of possible histories.
2. The theory places a measure on that set, and calls the measure 'chance'.
3. Rational agents accept the Principal Principle: roughly, that their degree of belief in a given event, conditional on the truth of the theory in question, is to be equal to the chance assigned by that theory to that event.

In discussions of the allegedly 'problematic' status of Everettian probability, it has been repeatedly pointed out[6] that there is in any case no substantive justification for the Principal Principle: the principle is accepted as a primitive principle of rationality. It follows that, if we are not to set the bar higher for Everettian quantum mechanics than we normally do for non-branching-universe theories, we should not reject the Everett interpretation merely because nothing substantive can be said about the *quantitative* problem of probability (if indeed that is the case). Once we have a solution to the incoherence problem – so, either we accept ('subjective-uncertainty') that it makes sense for an Everettian agent to have non-trivial degrees of belief about future measurement outcomes, or we accept the idea of a 'caring measure' over branches, with respect to which the agent maximizes 'expected' utility – we should, if we are content with non-Everettian theories, be content with the following:[7]

4. The theory postulates a set of histories (all of which are supposed physically real).
5. The theory places a measure on that set, and calls the measure 'branch weight'.
6. Rational agents accept a branching-universe version of the Principal Principle: roughly, that their degree of belief (or 'caring measure') for a given set of branches, conditional on the truth of the theory in question, is to be equal to the branch weights assigned by that theory to that event.

The whole Everettian decision-theoretic enterprise – once we have a solution to the incoherence problem (i.e. once we have accepted that rational agents must have *some* credence distribution or 'caring measure' over branches) – is actually an attempt to go beyond this minimal solution to the quantitative problem. It is an attempt to provide a substantive justification of the 'Everettian Principal Principle', and thus to exceed the achievements of non-branching accounts of objective probability.

4.2 Are many worlds **better** *than one?*

The 'many worlds are no worse than one' claim of section 4.1 is *relatively* uncontroversial (although the emphasis is intended seriously!). Some authors, however, make a stronger claim: that Everettian QM (post-ERT) actually provides a *more* satisfactory treatment of objective probability than has been possible on the basis of any other physical theory. Thus, for example[8], Saunders writes:

> "It is ironic that the interpretation of probability in the Everett interpretation has always been thought to be its weakest link. On the contrary, it seems that it is one of the strongest points in its favour." (Saunders 2004, p.26)

Whether or not this is correct depends on the extent to which the additional axioms – the axioms that enable Deutsch and Wallace to establish the uniqueness of the mod-squared measure as the rational 'uncertainty' or 'caring' measure, where Savage could only establish the existence of *some* probability measure for any given rational agent – have more force in a branching-multiverse situation than in a parallel-worlds situation. This is an open question (cf. section 3.2 above).

5 Many worlds and Sleeping Beauty

I conclude with some brief remarks on another open problem, whose solution may offer significant illumination.

When one finds uncertainty in an Everettian context, it is often of the *self-locating* variety. For example, according to one version of the subjective-uncertainty program, the *pre*-measurement agent is unsure *which world she is in* – but 'world' here does not mean 'everything that is physically real', and both worlds exist in the same possible multiverse. Less contentiously, with or without subjective-uncertainty, an Everettian agent *after* a measurement is often in a situation of self-locating uncertainty: if she has not yet looked at the measuring apparatus, then she does not know which branch she is in.

This suggests the possibility of analogies between the Everettian case and a puzzle surrounding the notion of self-locating belief more generally: the Sleeping Beauty problem (Elga (2000), Lewis (2000)). Exploration of possible analogies may shed light on both Everettian epistemology and the nature of rational *de se* belief in general. For example, Peter Lewis (2006) sketches and advocates a particular way in which the Everettian and Sleeping Beauty cases might be regarded as analogous, and argues that, pending explication of any relevant *dis*analogy, "the dominant 'thirder' solution to the Sleeping Beauty paradox [is] incompatible with the tenability of the many-worlds interpretation" (*ibid*, p.1). This issue deserves further attention.

6. Biography

Hilary Greaves is a philosopher of physics. Her recent research focusses on the issue of probability in the Everett (many-worlds) interpretation of quantum mechanics; she defends the Everett interpretation against the charge that it cannot make adequate sense of quantum probabilities. She has also published closely related work in mainstream Bayesian epistemology, where she provides (with David Wallace) a justification of the updating rule of conditionalization from the point of view of cognitive decision theory.

Greaves is currently working on the conceptual status of symmetries in physics, and the interpretation of quantum field theory.

Greaves is a graduate student at Rutgers University. She has published in *Mind* and in *Studies in History and Philosophy of Modern Physics*, and has held visiting scholarships at UC Irvine, the University of Pittsburgh, and the University of Sydney (Centre for Time). She holds a BA in physics and philosophy from the University of Oxford (2003).

7. References

Albert, David (1992). *Quantum mechanics and experience*. Harvard University Press: Cambridge, MA/London.

Albert, David and Barry Loewer (1988). Interpreting the many-worlds interpretation. *Synthese* 77, 195-213.

Barnum, H., C. M. Caves, J. Finkelstein, C. A. Fuchs and R. Schack (2000). Quantum probability from decision theory? Proceedings of the Royal Society of London A456, 1175-1182. Available online at http://www.arxiv.org/abs/quant-ph/9907024.

Barrett, Jeffrey (2003). Everett's Relative-State Formulation of Quantum Mechanics, *The Stanford Encyclopedia of Philosophy (Spring 2003 Edition)*, Edward N. Zalta (ed.). Online at http://plato.stanford.edu/archives/spr2003/entries/qm-everett/.

Deutsch, David (1985). Quantum theory as a universal physical theory. *International Journal of Theoretical Physic*s 24(1), 1-41.

Deutsch, David (1999). Quantum theory of probability and decisions. *Proceedings of the Royal Society of London* A455, 3129-3137. Available online at http://www.arxiv.org/abs/quant-ph/9906015; page numbers refer to the online version.

DeWitt, Bryce and Neill Graham (1973). *The many-worlds interpretation of quantum mechanics.* Princeton: Princeton University Press.

Everett, Hugh III (1957). Relative state formulation of quantum mechanics. *Review of Modern Physics* 29, 454-62. Reprinted in DeWitt and Graham (1973).

Graham, Neill (1973). The measurement of relative frequency. In DeWitt and Graham (1973).

Greaves, Hilary (2004). Understanding Deutsch's probability in a deterministic multiverse. *Studies in History and Philosophy of Modern Physics* 35, 423-56. Available online at http://www.arxiv.org/abs/quant-ph/0312136 and http://philsci-archive.pitt.edu/archive/00001742.

Greaves, Hilary (2007). On the Everettian epistemic problem. To appear in *Studies in History and Philosophy of Modern Physics*, 2007. Available online from http://philsci-archive.pitt.edu/archive/00002953/ .

Jeffrey, Richard (1965). *The logic of decision*, 2nd edition (1983). Chicago/London: University of Chicago Press.

Kent, Adrian (1990). Against many-worlds interpretations. *International Journal of Theoretical Physics* A5, 1745-62. Available online at http://www.arxiv.org/abs/gr-qc/9703089.

Lewis, David (2004). How many lives has Schrodinger's Cat? *Australasian Journal of Philosophy*, 82(1), pp.3-22, March 2004.

Lewis, Peter J. (2005). Probability in Everettian quantum mechanics. Available online at http://philsci-archive.pitt.edu/archive/00002716.

Lewis, Peter J. (2006a). Uncertainty and probability for branching selves. Available online at http://philsci-archive.pitt.edu/archive/00002636.

Lewis, Peter J. (2006b). Quantum sleeping beauty. Available online at http://philsci-archive.pitt.edu/archive/00002715.

Saunders, Simon (1996). Relativism. In Rob Clifton (ed.), *Perspectives on quantum reality*. Dordrecht/London: Kluwer.

Saunders, Simon (2002). Is the zero-point energy real? In *Ontological aspects of quantum field theory*, M. Kuhlmann, H. Lyre and A. Wayne (eds.), Singapore; World Scientific, 2002. Available online at http://philsci-archive.pitt.edu/archive/00002013; page numbers refer to the online version.

Saunders, Simon (2004). What is probability? To appear in Elitzur, A., S. Dolev and N. Kolenda (eds.), *Quo Vadis Quantum Mechanics*, Springer-Verlag. Available online at http://philsci-archive.pitt.edu/archive/00002011; page numbers refer to the online version.

Savage, Leonard J. (1972). *The foundations of statistics*. New York: Dover.

Papineau, David (1996). Many minds are no worse than one. *British Journal for the Philosophy of Science* 47 (2), 233-40.

Papineau, David (2004). David Lewis and Schrodinger's Cat. *Australasian Journal of Philosophy* 82(1), pp.153-69, March 2004.

Vaidman, Lev (2002). Many-Worlds Interpretation of Quantum Mechanics. *The Stanford Encyclopedia of Philosophy (Summer 2002 Edition)*, Edward N. Zalta (ed.). Online at http://plato.stanford.edu/archives/sum2002/entries/qm-manyworlds/.

Wallace, David (2002). Worlds in the Everett interpretation. *Studies in History and Philosophy of Modern Physics* 33 (2002), pp. 637--661. Available online at http://arxiv.org/abs/quant-ph/0103092 and http://philsci-archive.pitt.edu/archive/00000208.

Wallace, David (2003a). Everett and structure. *Studies in History and Philosophy of Modern Physics 34 (2003), pp. 86--105.* Online at http://arxiv.org/abs/quant-ph/0107144 and http://philsci-archive.pitt.edu/archive/00000681.

Wallace, David (2003b). Everettian Rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studies in History and Philosophy of Modern Physics* 34, 415-38. Available online at http://arxiv.org/abs/quant-ph/0303050 and http://philsci-archive.pitt.edu/archive/00001030.

Wallace, David (2005a). Quantum Probability from Subjective Likelihood: improving on Deutsch's proof of the probability rule. Forthcoming in *Studies in History and Philosophy of Modern Physics.* Available online at http://arxiv.org/abs/quant-ph/0312157 or from http://philsci-archive.pitt.edu.

Wallace, David (2005b). Epistemology Quantized: Circumstances in which we should come to believe in the Everett interpretation. Forthcoming in *British Journal for the Philosophy of Science.* Available online at http://philsci-archive.pitt.edu/archive/00002368.

Wallace, David (2005c). Language use in a branching universe. Available online at http://philsci-archive.pitt.edu/archive/00002554/02/branch_dec05.pdf.

Notes:

*Correspondence address: Philosophy Department, Davison Hall, 26 Nichol Ave, New Brunswick, NJ 08901-2882, USA. Email: hgreaves@rci.rutgers.edu

[1] Everett writes of a 'relative state interpretation' rather than a 'many-worlds interpretation'; correspondingly, we writes of 'almost all observer states' rather than 'almost all branches'.

[2] A brief history of the early part of this discussion is as follows. The original suggestion was made by Deutsch (1999). Deutsch's argument was criticized by Barnum, Caves, Finkelstein, Fuchs and Schack (2000). Wallace (2003b) gave a particularly clear reconstruction of Deutsch's argument, and argued that (*pace* Barnum *et al*) it can be seen to be sound after all, once one recognizes two crucial premises that Deutsch himself had left implicit: the Everett interpretation itself, and 'Measurement Neutrality'. Wallace (2005a) presents a simpler form of the argument, using weaker decision-theoretic axioms. Either of these papers by Wallace provides a good entry point into the literature. (Wallace's (2003b) is illuminating, but perhaps requires greater familiarity with the quantum formalism than does his (2005a).)

[3] This is the decision-theoretic framework developed by Savage (1972). An alternative framework is proposed by Jeffrey (1965). It would be interesting to explore the application of the Jeffrey framework to the Everettian argument.

[4] I skirt over a large amount of technical detail here. Deutsch's and Wallace's proofs do not, in fact, proceed by first citing the usual Savage proof, and then adding an assumption of Measurement Neutrality/Equivalence to restrict the permissible probability measures; rather, that 'extra' assumption is built in right from the beginning. (It is actually encoded in the notation: Wallace and Deutsch write, e.g.,

$\left\langle \psi, \widehat{X}, \mathrm{P} \right\rangle$ for a quantum bet, and this notation would be an insufficient specification of the bet if MN/Equivalence was not assumed.) The reader who wishes to pursue the details of the proofs is directed to the excellent papers by Wallace, cited above.

[5] Historically, the thought that the 'naive counting measure' is not only *a possible alternative* to the mod-squared measure, but is actually the *unique* rational measure for an Everettian, has been a major source of objections to Everettian solutions to the quantitative problem. Cf. Neill Graham's remark:

"It is extremely difficult to see what significance [the mod-squared] measure can have when its predictions are completely contradicted by a simple count of the worlds involved, worlds that Everett's own work assures us must be on an equal footing." (Graham 1973, p.236)

Even if the naive counting measure were coherent, though, the objection would remain a curious one: it presupposes a dubious application of the principle of indifference (and might just as well be levelled against a single-universe collapse interpretation, in which all outcomes are 'equally possible').

[6] See, e.g. Papineau (1996), Lewis (2004). (Papineau structures the case in terms of a 'decision-theoretic link' between probabilities and rational action, and an 'inferential link' between probabilities and frequencies, rather than directly in terms of the Principal Principle.)

[7] This way of setting out the analogy is due to Wayne Myrvold (in correspondence).

[8] Cf. also Papineau (2004, secs. IIC,D); Wallace (2005b, sec.3.6).