Everettian Rationality: defending Deutsch's approach to probability in the Everett interpretation

David Wallace*

An analysis is made of Deutsch's recent claim to have derived the Born rule from decision-theoretic assumptions. It is argued that Deutsch's proof must be understood in the explicit context of the Everett interpretation, and that in this context, it essentially succeeds. Some comments are made about the criticism of Deutsch's proof by Barnum, Caves, Finkelstein, Fuchs, and Schank; it is argued that the flaw which they point out in the proof does not apply if the Everett interpretation is assumed.

A longer version of this paper, entitled *Quantum Probability and Decision Theory, Revisted*, is available online (Wallace 2002). The present paper will appear in *Studies in the History and Philosophy of Modern Physics*; confusingly, when it does it will also bear the title *Quantum Probability and Decision Theory, Revisited*.

Keywords: Interpretation of Quantum Mechanics — Everett interpretation; Probability; Decision Theory

1 Introduction

In recent work on the Everett (Many-Worlds) interpretation of quantum mechanics, it has increasingly been recognized that any version of the interpretation worth defending will be one in which the basic formalism of quantum mechanics is left unchanged. Properties such as the interpretation of the wave-function as describing a multiverse of branching worlds, or the ascription of probabilities to the branching events, must be emergent from the unitary quantum mechanics rather than added explicitly to the mathematics. Only in this way is it possible to save the main virtue of Everett's approach: having an account of quantum mechanics consistent with the last seventy years of physics, not one in which the

(*December 26, 2002*)
* Magdalen College, Oxford University, Oxford OX1 4AU, U.K.
(*e-mail:* david.wallace@magdalen.ox.ac.uk).

edifice of particle physics must be constructed afresh (Saunders 1997, p. 44).[1]

Of the two main problems generally raised with Everett-type interpretations, the preferred-basis problem looks eminently solvable without changing the formalism. The main technical tool towards achieving this has of course been decoherence theory, which has provided powerful (albeit perhaps not conclusive) evidence that the quantum state has a *de facto* preferred basis and that this basis allows us to describe the universe in terms of a branching structure of approximately classical, approximately non-interacting worlds. I have argued elsewhere (Wallace 2001a, 2001b) that there are no purely conceptual problems with using decoherence to solve the preferred-basis problem, and that the inexactness of the process should give us no cause to reject it as insufficient. In particular, the branching events in such a theory can be understood, literally, as replacement of one classical world with several — so that in the Schrödinger Cat experiment, for instance, after the splitting there is a part of the quantum state which should be understood as describing a world in which the cat is alive, and another which describes a world in which it is dead. This multiplication comes about not as a consequence of adding extra, world-defining elements to the quantum formalism, but as a consequence of an ontology of macroscopic objects (suggested by Dennett 1991) according to which they are treated as patterns in the underlying microphysics.

This account applies to human observers as much as to cats: such an observer, upon measuring an indeterminate event, branches into multiple observers with each observer seeing a different outcome. Each future observer is (initially) virtually a copy of the original observer, bearing just those causal and structural relations to the original that future selves bear to past selves in a non-branching theory. Since (arguably; see Parfit (1984) for an extended defence) the existence of such relations is all that there is to personal identity, the post-branching observers can legitimately be understood as future selves of the original observer, and he should care about them just as he would his unique future self in the absence of branching.

This brings us on to the other main problem with the Everett interpretation,

[1] This is by no means *universally* recognized. Everett-type interpretations can perhaps be divided into three types:

(i) Old-style "Many-Worlds" interpretations in which worlds are added explicitly to the quantum formalism (see, e. g. , DeWitt (1970) and Deutsch (1985), although Deutsch has since abandoned this approach; in fact, it is hard to find any remaining defendants of type (i) approaches).

(ii) "Many-Minds" approaches in which some intrinsic property of the mind is essential to understanding how to reconcile indeterminateness and probability with unitary quantum mechanics (see, e. g. , Albert and Loewer (1988), Lockwood (1989, 1996), Donald (1997), and Sudbery (2000)).

(iii) Decoherence-based approaches, such as those defended by myself (Wallace 2001a, 2001b), Saunders (1995, 1997, 1998), Deutsch (1996, 2001), Vaidman (1998, 2001) and Zurek (1998).

For the rest of this paper, whenever I refer to "the Everett interpretation", I shall mean specifically the type (iii) approaches. This is simply for brevity, and certainly isn't meant to imply anything about what was intended in Everett's original (1957) paper.

the concept of probability. Given that the Everettian description of measurement is a deterministic, branching process, how are we to reconcile that with the stochastic description of measurement used in practical applications of quantum mechanics? It has been this problem, as much as the preferred basis problem, which has led many workers on the Everett interpretation to introduce explicit extra structure into the mathematics of quantum theory so as to make sense of the probability of a world as (for instance) a measure over continuously many identical worlds. Even some proponents of the Many-Minds variant on Everett (notably Albert and Loewer 1988 and Lockwood 1989, 1996), who arguably have no difficulty with the preferred-basis problem, have felt forced to modify quantum mechanics in this way.

It is useful to identify two aspects of the problem. The first might be called the *incoherence problem*: how, when every outcome actually occurs, can it even make sense to view the result of a measurement as uncertain? Even were this solved, there would then remain a *quantitative problem*: why is that uncertainty quantified according to the quantum probability rule (i. e. , the Born rule), and not (for instance) some other assignment of probabilities to branches?

Substantial progress has also been made on the incoherence problem. In my view, the most promising approach is Saunders' 'subjective uncertainty' theory of branching: Saunders argues (via the analogy with Parfittian fission) that an agent awaiting branching should regard it as *subjectively* indeterministic. That is, he should expect to become one future copy or another but not both, and he should be uncertain as to which he will become. (Saunders' strategy can be found in Saunders (1998), and in the longer version of the current paper.) An alternative strategy has been suggested by Vaidman (1998, 2001): *immediately* after the branching event (before we actually see the result of the measurement) the agent knows that he is determinately in one branch or another but is simply ignorant as to which one.

If progress is being made on the incoherence problem, the quantitative problem is all the more urgent. In this context it is extremely interesting that David Deutsch has claimed (Deutsch 1999) to derive the quantum probability rule from decision theory: that is, from considerations of pure rationality. It is rather surprising how little attention his work has received in the foundational community, though one reason may be that it is very unclear from his paper that the Everett interpretation is assumed from the start.[2] If it is tacitly assumed that his work refers instead to some more orthodox collapse theory, then it is easy to see that the proof is suspect; this is the basis of the criticisms levelled at Deutsch by Barnum *et al*, (2000). Their attack on Deutsch's paper seems to have been influential in the community; however, it is at best questionable whether or not

[2] Nonetheless it *is* assumed:

> However, in other respects he [the rational agent] will not behave as if he believed that stochastic processes occur. For instance if *asked* whether they occur he will certainly reply 'no', because the non-probabilistic axioms of quantum theory require the state to evolve in a continuous and deterministic way. [Deutsch 1999, pp. 13; emphasis his.]

it is valid when Everettian assumptions are made explicit. (This matter will be discussed further below.)

If the Everettian context is made explicit, Deutsch's strategy can be reconstructed as follows. Assuming that the outcome of a measurement can in some sense be construed as uncertain (that is, that Saunders', Vaidman's, or some other strategy resolves the incoherence problem), then the 'quantitative problem' splits into two halves:

1. What justifies using probabilities to quantify the uncertainty *at all*?
2. Why use those specific probabilities given by the Born rule?

Fairly obviously, the first of these is not really a quantum-mechanical problem at all but a more general one — and one which decision theory is designed to answer. In decision theory, we start with some general assumptions about rationality, and deduce that any agent whose preferences between actions satisfies those assumptions must act as if they allocated probabilities to possible outcomes and preferred those actions that maximized expected utility with respect to those probabilities. Roughly speaking, this is to *define* the probability assigned to X by the agent as the shortest odds at which the agent would be prepared to bet on X occurring.

Deutsch's strategy is to transfer this strategy across to quantum theory: to start with axioms of rational behavior, apply them to quantum-mechanical situations, and deduce that rational agents should quantify their subjective uncertainty in the face of splitting by the use of probability. What is striking about the quantum-mechanical version of decision theory, though, is that rational agents are so strongly constrained in their behavior that not only must they assign probabilities to uncertain events, they must assign precisely those probabilities given by the Born Rule. This discovery might be called Deutsch's theorem, since it is the central result of Deutsch's paper.

The structure of the paper is as follows. Section 2 gives an unambiguous definition of Deutsch's quantum games, and derives some preliminary results about them; section 3 describes the decision-theoretic assumptions Deutsch makes. In section 4 I run through Deutsch's proof of the Born rule; section 5 gives an alternative proof of my own, from slightly different assumptions. Sections 6 and 7 deal with possible criticisms of Deutsch's approach (section 6 reviews the criticisms made by Barnum *et al*; section 7 describes a possible problem with the proof not discussed either by Barnum *et al* or by Deutsch). Section 8 is the conclusion.

An extended version of this paper (Wallace 2002) is available online.

2 Quantum measurements and quantum games

In this section, I will define Deutsch's notion of a 'quantum game' — effectively a bet placed on the outcome of a measurement. Though I follow Deutsch's definition of a game, my notation will differ from his in order to resolve some ambiguities in the definition (first identified by Barnum *et al*, 2000).

Informally, a (quantum) game is to be a three-stage process: a system is prepared in some state; a measurement of some observable is made on that state; a reward, dependent on the result of the measurement, is paid to the player. Formally, we will define a **game** (in boldface) thus:

> A **game** is an ordered triple $\langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle$, where:
>
> - $|\psi\rangle$ is a state in some Hilbert space $\mathcal{H}$ (technically the Hilbert space should also be included in the definition, but has been omitted for brevity);
> - $\widehat{X}$ is a self-adjoint operator on $\mathcal{H}$;
> - $\mathcal{P}$ is a function from the spectrum of $\widehat{X}$ into the real numbers.

Technically this makes a **game** into a mathematical object; but obviously we're really interested in physical processes somehow described by that object. We say that a given process *instantiates* some **game** $\langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle$ if and only if that process consists of:

1. The preparation of some quantum system, whose state space is described by $\mathcal{H}$, in the state (represented by) $|\psi\rangle$;
2. The measurement, on that system, of the observable (represented by) $\widehat{X}$;
3. The provision, in each branch in which result 'x' was recorded, of some payment of cash value $\mathcal{P}(x)$.

We'll define a game (not in boldface) as any process which instantiates a **game**.

The distinction between **games** and games may seem pedantic: the whole strategy of mathematical physics is to use mathematical objects to represent physical states of affairs, and outside the philosophy of mathematics there is seldom if ever a need to distinguish between the two. However, it's crucial to an understanding of Deutsch's proof to notice that the instantiation relation, between **games** and games is not one-to-one. Quite the reverse, in fact: many **games** can instantiate a given game (unsurprisingly: there are many ways to construct a measuring device), and (perhaps more surprisingly) a single game instantiates many **games**. We can define an equivalence relation $\simeq$ between **games**: $\mathcal{G} \simeq \mathcal{G}'$ iff $\mathcal{G}$ and $\mathcal{G}'$ are instantiated by the same game.

To explore the properties of $\simeq$, we need to get precise about what physical processes count as measurements. Since we are working in the Everett framework, we can model a measurement as follows: let $\mathcal{H}_s$ be the Hilbert space of some subsystem of the Universe, and $\mathcal{H}_e$ be the Hilbert space of the measurement device;[3] let $\widehat{X}$ be the observable to be measured.

Then a *measurement procedure* for $\widehat{X}$ is specified by:

[3] In practice, the Hilbert space $\mathcal{H}_e$ would probably have to be expanded to include an indefinitely large portion of the surrounding environment, since the latter will inevitably become entangled with the device.

1. Some state $|\mathcal{M}_0\rangle$ of $\mathcal{H}_e$, to be interpreted as its initial (pre-measurement) state; this state must be an element of the preferred basis picked out by decoherence.

2. Some basis $|\lambda_i\rangle$ of eigenstates of $\widehat{X}$, where $\widehat{X}\,|\lambda_i\rangle = x_i\,|\lambda_i\rangle$.

3. Some set $\{|\mathcal{M}; x_i; \alpha\rangle\}$ of "readout states" of $\mathcal{H}_s \otimes \mathcal{H}_e$, also elements of the decoherence basis, *at least* one for each state $|\lambda_a\rangle$. The states must physically display x_i, in some way measurable by our observer (e. g., by the position of a needle).

4. Some dynamical process, triggered when the device is activated, and defined by the rule

$$|\lambda_i\rangle\otimes|\mathcal{M}_0\rangle \longrightarrow \sum_\alpha \mu(\lambda_i;\alpha)\,|\mathcal{M}; x_i; \alpha\rangle \tag{1}$$

   where the $\mu(\lambda_i;\alpha)$ are complex numbers satisfying $\sum_\alpha |\mu(\lambda_i;\alpha)|^2 = 1$.

What justifies calling this a 'measurement'? The short answer is that it is the standard definition; a more principled answer is that the point of a measurement of $\widehat{X}$ is to find the value of $\widehat{X}$, and that whenever the value of $\widehat{X}$ is definite, the measurement process will successfully return that value. (Of course, if the value of $\widehat{X}$ is not definite then the measurement process will lead to branching of the device and the observer; but this is inevitable given linearity.)

Observe that:

1. We are not restricting our attention to so-called "non-disturbing" measurements, in which $|\mathcal{M}; x_i\rangle = |\lambda_i\rangle\otimes|\mathcal{M}'; x_i\rangle$. In general measurements will destroy or at least disrupt the system being measured, and we allow for this possibility here.

2. The additional label α allows for the fact that many possible states of the measurement device may correspond to a single measurement outcome. Even in this case, of course, an observer can predict that whenever $|\psi\rangle$ is an eigenstate of $\widehat{X}$, all his / her future copies will correctly learn the value of $\widehat{X}$. In practice most realistic measurements are likely to be of this form, because the process of magnifying microscopic data up to the macro level usually involves some random processes.

3. Since a readout state's labelling is a matter not only of physical facts about that state but also of the labelling conventions used by the observer, there is no physical difference between a measurement of $\widehat{X}$ and one of $f(\widehat{X})$, where f is an arbitrary one-to-one function from the spectrum of $\widehat{X}$ onto some subset of $\Re$: a measurement of $f(\widehat{X})$ may be interpreted simply as a measurement of $\widehat{X}$, using a different labelling convention. More accurately, there *is* a physical difference, but it resides in the brain state of the observer (which presumably encodes the labelling convention in some way) and not in the measurement device.

To save on repetition, let us now define some general conventions for **games**: we will generally use $\widehat{X}$ for the operator being measured, and denote its eigenstates by $|\lambda_i\rangle$; the eigenvalue of $|\lambda_i\rangle$ will be x_i. (We allow for the possibility of degenerate $\widehat{X}$, so that may have $x_i = x_j$ even though $i \neq j$.) We write $\sigma(\widehat{X})$ for the spectrum of $\widehat{X}$, and $\widehat{P}_X(x)$ for the projector onto the eigensubspace of $\widehat{X}$ with eigenvalue x; thus,

$$\widehat{X} = \sum_{x \in \sigma(X)} x \, \widehat{P}_X(x). \tag{2}$$

For a given **game** $\mathcal{G} = \langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle$, we also define the *weight map* $W_{\mathcal{G}} : \Re \rightarrow \Re$ by

$$W_{\mathcal{G}}(c) = \sum_{x \in \mathcal{P}^{-1}\{c\}} \langle\psi| \, \widehat{P}_X(x) \, |\psi\rangle \tag{3}$$

(that is, the sum ranges over all $x \in \sigma(\widehat{X})$ such that $\mathcal{P}(x) = c$). It is readily seen that for any game instantiating $\mathcal{G}$, $W_{\mathcal{G}}(c)$ is the weight of the payoff c: that is, the sum of the weights of all the branches in which payoff c is given. Because of this we can refer without confusion to $W_{\mathcal{G}}(c)$ as the weight of c. (Recall that the *weight* of a branch is simply the squared modulus of the amplitude of that branch (relative to the pre-branching amplitude, of course); thus if the state of a measuring device following measurement is

$$\sum_i \alpha_i \, |\mathcal{M}; x_i\rangle \, , \tag{4}$$

then the weight of the branch in which result x_i occurs is $|\alpha_i|^2$.)

We can now state and prove the:

Equivalence Theorem

1. Payoff Equivalence (PE):

$$\langle |\psi\rangle, \widehat{X}, \mathcal{P} \cdot f\rangle \simeq \langle |\psi\rangle, f(\widehat{X}), \mathcal{P}\rangle \tag{5}$$

   where $f : \sigma(\widehat{X}) \rightarrow \Re$.

2. Measurement Equivalence (ME):

$$\langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle \simeq \langle \widehat{U} \, |\psi\rangle, \widehat{X}', \mathcal{P}'\rangle \tag{6}$$

   where
   - $\widehat{U}$ is a unitary transformation;
   - $\widehat{X}$ and $\widehat{X}'$ satisfy $\widehat{U}\widehat{X} = \widehat{X}'\widehat{U}$;
   - $\mathcal{P}$ and $\mathcal{P}'$ agree on $\sigma(\widehat{X})$.

(Note that we allow $\widehat{U}$ to connect different Hilbert spaces here. If $\widehat{U}$ transforms a fixed Hilbert space, the result simplifies to

$$\langle |\psi\rangle , \widehat{X}, \mathcal{P}\rangle \simeq \langle \widehat{U} |\psi\rangle , \widehat{U}\widehat{X}\widehat{U}^{\dagger}, \mathcal{P}\rangle. \quad) \tag{7}$$

3. General Equivalence (GE): $\mathcal{G} \simeq \mathcal{G}'$ iff $W_{\mathcal{G}} = W_{\mathcal{G}'}$.

Proof:

1. Recall that our definition of a measurement process involves a set of states $|\mathcal{M}; x_i\rangle$ of the decoherence-preferred basis, which are understood as readout states — and that the rule associating an eigenvalue x_i with a readout state $|\mathcal{M}; x_i\rangle$ is just a matter of convention. Change this convention, then: regard $|\mathcal{M}; x_i\rangle$ as displaying $f(x_i)$ — but also change the payoff scheme: replace a payoff $\mathcal{P} \cdot f(x)$ upon getting result x with a payoff $\mathcal{P}(x)$. These two changes replace the **game** $\langle |\psi\rangle , \widehat{X}, \mathcal{P}\cdot f\rangle$ with $\langle |\psi\rangle , f(\widehat{X}), \mathcal{P}\rangle$ — but no physical change at all has occurred, just a change of labelling convention. Hence $\langle |\psi\rangle , \widehat{X}, \mathcal{P} \cdot f\rangle \simeq \langle |\psi\rangle , f(\widehat{X}), \mathcal{P}\rangle$.

2. For simplicity, let us assume that $\widehat{X}$ and $\widehat{X}'$ act on different Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$. (This assumption can be relaxed either by a trivial change of the proof, or directly by realizing $\widehat{U}$ in two steps, via an auxiliary Hilbert space.)

   Because $\widehat{U}\widehat{X} = \widehat{X}'\widehat{U}$, it must be possible to label the eigenstates $|\mu_1\rangle , \ldots |\mu_{n'}\rangle$ of $\widehat{X}'$ so that for $a \leq n$, $\widehat{U} |\lambda_i\rangle = |\mu_i\rangle$ and $\widehat{X}' |\mu_i\rangle = x_i |\mu_i\rangle$. Now, without loss of generality take $|\psi\rangle = \sum_{i=1}^{n} \alpha_i |\lambda_i\rangle$, and consider the following physical process:

   (a) Prepare the system represented by $\mathcal{H}$ in state $|\psi\rangle$, and the system represented by $\mathcal{H}'$ in some fixed state $|0'\rangle$, so that the overall quantum state is
   $$|\psi\rangle \otimes |0'\rangle \otimes |\mathcal{M}_0\rangle \tag{8}$$
   where $|\mathcal{M}_0\rangle$ is the initial state of some measurement device for $\mathcal{H}'$.

   (b) Operate on $\mathcal{H}\otimes\mathcal{H}'$ with some unitary transformation realizing $|\phi\rangle \otimes |0'\rangle \to |0\rangle \otimes \widehat{U} |\phi\rangle$, where $|\phi\rangle$ is an arbitrary state of $\mathcal{H}$ and $|0\rangle$ is some fixed state of $\mathcal{H}$. (That such a transformation exists is trivial.)

   (c) Discard the system represented by $\mathcal{H}$. (This step is just for notational convenience.) The system retained is now in state $\widehat{U} |\psi\rangle \otimes |\mathcal{M}_0\rangle$.

   (d) Measure $\widehat{X}'$ using the following dynamics:
   $$|\mu_i\rangle \otimes |\mathcal{M}_0\rangle \longrightarrow |\mathcal{M}; x_i; a\rangle \tag{9}$$
   where for each i, $|\mathcal{M}; x_i; i\rangle$ is a readout state giving readout x_i. (The extra i index is only there to allow for degeneracy, and can be dropped if $\widehat{X}'$ is non-degenerate.)

(e) The final state is now

$$\sum_{i=1}^{n} \alpha_i \left|\mathcal{M}; x_i; i\right\rangle . \tag{10}$$

In the branches where result x_i is recorded, give a payoff $\mathcal{P}'(x_i)$.

This process can be described as follows: in steps (a)–(c) we prepare the state $\widehat{U}\left|\psi\right\rangle$ of $\mathcal{H}'$, using an auxiliary system represented by $\mathcal{H}$. In step (d) we measure the operator $\widehat{X}'$ on that state, and in step (e) we provide a payout $\mathcal{P}'$. This is an instantiation of the **game** $\langle \widehat{U}\left|\psi\right\rangle, \widehat{X}', \mathcal{P}'\rangle$.

However, suppose we just treat steps (b)–(d) as a black box process. That process realizes the transformation

$$\left(\sum_{i=1}^{n} \alpha_i \left|\lambda_i\right\rangle\right) \otimes \left|0'\right\rangle \otimes \left|\mathcal{M}_0\right\rangle \longrightarrow \sum_{i=1}^{n} \alpha_i \left|\mathcal{M}; x_i; i\right\rangle , \tag{11}$$

which — by definition of measurement — is a measurement of $\widehat{X}$ on the state $\left|\psi\right\rangle$, using a measurement device with initial state $\left|0\right\rangle \otimes \left|\mathcal{M}_0\right\rangle$.

This observation means that the process (a)–(e) can *also* be described in another way: in step (a) we prepare the state $\left|\psi\right\rangle$ of $\mathcal{H}$; in steps (b)–(d) we measure the operator $\widehat{X}$ on that state (using an auxiliary system represented by $\mathcal{H}'$); in step (e) we provide a payout $\mathcal{P}$. (Note that the measurement of $\widehat{U}\left|\psi\right\rangle$ gives, with certainty, some result $x_1, \ldots x_n$, so there is no physical difference between providing payoff $\mathcal{P}$ and payoff $\mathcal{P}'$.) Thus the process is an instantiation of the **game** $\langle \left|\psi\right\rangle, \widehat{X}, \mathcal{P}\rangle$.

There is no physical difference between the two descriptions of (a)– (e); there is simply a change in how we choose to describe the process. It follows that $\langle \left|\psi\right\rangle, \widehat{X}, \mathcal{P}\rangle \simeq \langle \widehat{U}\left|\psi\right\rangle, \widehat{X}', \mathcal{P}'\rangle$.

3. For each n, let $\mathcal{H}_0^n$ be some n-dimensional Hilbert space with self-adjoint operator $\widehat{K}$, having eigenstates $\left|\kappa_1\right\rangle, \ldots \left|\kappa_n\right\rangle$ with $\widehat{K}\left|\kappa_i\right\rangle = i\left|\kappa_i\right\rangle$. (Technically we should distinguish between the $\widehat{K}$ for different n, but no ambiguity will result from this abuse of notation.)

   If $\mathcal{G} = \langle \left|\psi\right\rangle, \widehat{X}, \mathcal{P}\rangle$ is any **game** with n distinct payoffs — that is, elements in the range of $\mathcal{P}$ — $c_1, \ldots c_n$ with non-zero weights $w_1, \ldots w_n$ (plus any number of 'possible' payoffs with zero weight), we will show that $\mathcal{G}$ is equivalent to the 'canonical' **game** $\langle \left|\psi_0\right\rangle, \widehat{K}, \mathcal{P}_0\rangle$, where

   - $\left|\psi_0\right\rangle$ is a state in $\mathcal{H}_0^n$;
   - $\left|\psi_0\right\rangle = \sum_{i=1}^{n} \sqrt{w_i}\left|\kappa_i\right\rangle$;
   - $\mathcal{P}_0(i) = c_i$.

   This will be sufficient to prove GE. We proceed as follows:

(a) Let $\mathcal{H}$ be the Hilbert space of $\mathcal{G}$ (i. e., the Hilbert space on which $\widehat{X}$ acts) and let $\mathcal{S}$ be the direct sum of all eigenspaces of $\widehat{X}$ which have nonzero overlap with $|\psi\rangle$. If $\widehat{U}$ is the embedding map of $\mathcal{S}$ into $\mathcal{H}$, and $\mathcal{P}|_{\mathcal{S}}$ is the restriction of $\mathcal{P}$ to the spectrum of $\widehat{X}|_{\mathcal{S}}$, then it follows from ME that

$$\mathcal{G} \simeq \langle |\psi\rangle, \widehat{X}|_{\mathcal{S}}, \mathcal{P}|_{\mathcal{S}}\rangle. \tag{12}$$

We may therefore assume, without loss of generality, that $\mathcal{S} = \mathcal{H}$; that is, that $|\psi\rangle$ has non-zero overlap with all eigenstates of $\widehat{X}$.

(b) Also without loss of generality, we may assume the eigenstates of $\widehat{X}$ ordered so that the first n_1 give payoff c_1, the next n_2 give payoff c_2, and so on. (We know each n_i is non-zero, by (a).) Then we can write $|\psi\rangle$ as

$$|\psi\rangle = \sum_{i=1}^{N} \alpha_i |\lambda_i\rangle, \tag{13}$$

where each α_i is non-zero.
Now define the normalized vectors $|\mu_1\rangle, \ldots |\mu_n\rangle$ by

$$|\mu_1\rangle = \frac{\alpha_1 |\lambda_1\rangle + \cdots + \alpha_{n_1} |\lambda_{n_1}\rangle}{\sqrt{|\alpha_1|^2 + \cdots + |\alpha_{n_1}|^2}}, \tag{14}$$

$$|\mu_2\rangle = \frac{\alpha_{n_1+1} |\lambda_{n_1+1}\rangle + \cdots + \alpha_{n_2} |\lambda_{n_2}\rangle}{\sqrt{|\alpha_{n_1+1}|^2 + \cdots + |\alpha_{n_2}|^2}}, \tag{15}$$

etc. Then by definition of $W_{\mathcal{G}}$, $w_i \equiv W_{\mathcal{G}}(c_i)$, we now have

$$|\psi\rangle = \sum_{i=1}^{n} \sqrt{w_i} |\mu_i\rangle. \tag{16}$$

(c) Define f by $f(x) = i$ whenever x is the eigenvalue of an eigenstate leading to payoff c_i. By PE,

$$\mathcal{G} \simeq \langle |\psi\rangle, f(\widehat{X}), \mathcal{P} \cdot f^{-1}\rangle. \tag{17}$$

$f(\widehat{X})$ is an operator which has the $|\mu_i\rangle$ as eigenstates: $f(\widehat{X}) |\mu_i\rangle = i |\mu_i\rangle$.

(d) Finally, let $\widehat{U}$ be a unitary map from $\mathcal{H}_0^n$ to $\mathcal{H}$, given by $\widehat{U} |\kappa_i\rangle = |\mu_i\rangle$. Since $\widehat{U}\widehat{K} = f(\widehat{X})\widehat{U}$, we have by ME

$$\mathcal{G} \simeq \langle \sum_{i=1}^{n} \sqrt{w_i} |\kappa_i\rangle, \widehat{K}, \mathcal{P} \cdot f^{-1}\rangle \tag{18}$$

which is the canonical **game** above.

Before moving on, it's necessary to cover one further ramification to his concept of 'game' (and '**game**'): *compound games.* A compound game is obtained from an existing game by replacing some or all of its consequences with new games. (For instance, we might measure the spin of a spin-half particle, and play one of two possible games according to which spin we obtained.)[4]

3 Decision theory

To complete our goal of deriving the Born rule, we will need to introduce some decision-theoretic assumptions about agents' preferences between games. Following Deutsch, we do so by introducing a *value function*: a map $\mathcal{V}$ from the set of **games** to the reals, such that if some **game**'s payoff function is constant and equal to c, then the value of that **game** is c. (For convenience, we write '$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P})$' in place of '$\mathcal{V}(\langle|\psi\rangle, \widehat{X}, \mathcal{P}\rangle)$'.)

The idea of the function is that a rational agent prefers a **game** $\mathcal{G}$ to another $\mathcal{G}'$ just if $\mathcal{V}(\mathcal{G}) > \mathcal{V}(\mathcal{G}')$. $\mathcal{V}(\mathcal{G})$ can be thought of, in fact, as the 'cash value' of $\mathcal{G}$ to the agent: s/he will be indifferent between playing $\mathcal{G}$, and receiving a reward with a cash value of $\mathcal{V}(\mathcal{G})$. (It follows that a **game** whose payoff function is constant must have $\mathcal{V}(\mathcal{G})$ equal to that constant value; hence the requirement.)

Deutsch now imposes the following restrictions on $\mathcal{V}$.

Dominance: If $\mathcal{P}(x) \geq \mathcal{P}'(x)$ for all x, then

$$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}) \geq \mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}'). \tag{19}$$

Substitutivity: If $\mathcal{G}_{comp}$ is a compound **game** formed from some **game** $\mathcal{G}$ by substituting for its consequences $c_1, \ldots c_n$ **games** $\mathcal{G}_1, \ldots \mathcal{G}_n$ such that $\mathcal{V}(\mathcal{G}_i) = c_i$, then $\mathcal{G}_{comp} \simeq \mathcal{G}$.

Weak additivity: If k is any real number, then

$$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P} + k) = \mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}) + k. \tag{20}$$

Zero-sum: For given payoff $\mathcal{P}$, let $-\mathcal{P}$ be defined by $(-\mathcal{P})(x) = -(\mathcal{P}(x))$. Then

$$\mathcal{V}(|\psi\rangle, \widehat{X}, -\mathcal{P}) = -\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}). \tag{21}$$

As with any set of decision-theoretic assumptions, the idea is that any rational set of preferences between games must be given by some value function which satisfies these constraints: to violate any one of them is to be irrational in some way.

---

[4] Formally,

- a **simple game** is just a **game** as defined above;
- a **compound game** of rank n is a triple $\langle|\psi\rangle, \widehat{X}, \mathcal{P}\rangle$, where $\mathcal{P}$ is a map from $\sigma(\widehat{X})$ into the set of **simple games** and **compound games** of rank $n-1$;
- A compound game is any physical process instantiating a **compound game**;

although in fact we will not ever need to be so formal.

Specifically, **Dominance** says that if one game invariably leads to better rewards than another, take the first game. **Substitutivity** says that, if an agent is indifferent between getting a definite reward c and playing some game, s/he should also be indifferent between a chance of getting c and the same chance of playing that game.

We can motivate **Weak additivity** like this: consider any physical process which first instantiates $\mathcal{G} = \langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle$, and then delivers a reward of value k with certainty. This is physically equivalent to measuring $|\psi\rangle$ and then receiving, sequentially, two rewards on getting result x_a: one of cash value $\mathcal{P}(x_a)$ and one of value k. This reward is equivalent to a single one of value $\mathcal{P}(x_a) + k$ and so the physical process realizes $\langle |\psi\rangle, \widehat{X}, \mathcal{P} + k\rangle$.

Now, suppose that the fixed reward k is received *before* playing (the game instantiating) $\mathcal{G}$. By **Substitutivity**, the agent is indifferent between receiving k then playing $\mathcal{G}$, and receiving k then receiving $\mathcal{V}(\mathcal{G})$. But the latter process is just that of receiving a 'lump-sum' payment of $\mathcal{V}(\mathcal{G}) + k$.

Zero-sum can be motivated as follows: if I and someone else who shares my exact preferences play some sort of game in which any gain to one is balanced by a loss to the other, it seems reasonable to assume that if one of us actively wants to play (that is, expects to benefit), the other must actively want not to play (that is, expects to lose out).

Now suppose $\mathcal{G} = \langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle$, and that I play $\mathcal{G}' = \langle |\psi\rangle, \widehat{X}, \mathcal{P} - \mathcal{V}(\mathcal{G})\rangle$ with my alter ego acting as banker; he is thus playing $-\mathcal{G}' = \langle |\psi\rangle, \widehat{X}, \mathcal{V}(\mathcal{G}) - \mathcal{P}\rangle$.

But by **Weak additivity**, I am indifferent to playing $\mathcal{G}$' ($\mathcal{V}(\mathcal{G}') = 0$). It follows that my alter ego must be indifferent to playing $-\mathcal{G}'$, and hence (applying the lemma again) that **Zero-sum** holds. (I am grateful to Simon Saunders for this argument.)

In fact, both **Weak additivity** and **Zero-Sum** are special cases of the following general principle:

Additivity:

$$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P} + \mathcal{P}') = \mathcal{V}(\langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle) + \mathcal{V}(\langle |\psi\rangle, \widehat{X}, \mathcal{P}'\rangle). \tag{22}$$

This can be motivated as follows: suppose I know some measurement is to be carried out, and I want to pay to buy tickets entitling me to a bet on that measurement. Each bet is represented by some payoff function $\mathcal{P}$, to which I might imagine assigning a cash value (the largest value I'll pay for the ticket which allows me to make the bet.) If I assume that the price I'd pay for a given ticket doesn't depend on which tickets I've already bought, **Additivity** follows.

Deutsch doesn't in fact assume **Additivity** (though I don't think there's any deep significance to this) but it will allow us to simplify his proof considerably. In practice, **Additivity** is essentially equivalent to the conjunction of **Weak additivity**, **Zero-Sum** and **Substitutivity**: the first two allow us to prove **Additivity** for games with two possible outcomes, and the third allows us to build up multi-outcome games from two-outcome ones.

All of these assumptions are essentially independent of quantum mechanics, and they already allow us to do quite a lot of decision theory: in fact, we can prove the

> **Probability representation theorem:** If $\mathcal{V}$ is a value function which satisfies **Additivity** and **Dominance**, $\mathcal{V}$ is given by
>
> $$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}) = \sum_{x \in \sigma(X)} \Pr_{\psi,X}(x)\mathcal{P}(x) \tag{23}$$
>
> where the $\Pr_{\psi,X}(x)$ are real numbers between 0 and 1 which depend on $|\psi\rangle$ and $\widehat{X}$ but not on $\mathcal{P}$, and where
>
> $$\sum_{x \in \sigma(X)} \Pr_{\psi,X}(x) = 1. \tag{24}$$

(The essential idea of the proof is that we *define* the probability of a measurement outcome as the shortest odds we'd accept on its occurrence, and use **Additivity** to prove this is consistent; see the appendix for the full proof.)

In fact, this result shows that the decision-theoretic axioms we are adopting are actually quite strong: they imply, for instance, that it's rational to bet the mortgage on a one-in-a-million chance of winning the GNP of Europe. They seem reasonable as long as we restrict our attention to betting with small sums, however. (And, as I show in Wallace (2002), it is possible to improve Deutsch's results by substantially weakening his decision-theoretic assumptions.)

Nonetheless, the Representation Theorem is still far short of the Born rule. No link has been made between the probabilities $\Pr_{\psi,X}(x)$ and the weight of the branches, and in fact the Representation Theorem is consistent with different agents (that is, different value functions) assigning very different probabilities to the same event.

The connection to quantum theory comes in entirely through the last assumption made:

Physicality: Two **games** instantiated by the same physical process have the same value; that is, $\mathcal{G} \simeq \mathcal{G}' \rightarrow \mathcal{V}(\mathcal{G}) = \mathcal{V}(\mathcal{G}')$.

The motivation for this, obviously, is that real agents have preferences between games, not **games**. I return to this point in section 7.

4 Deutsch's proof

We are now in a position to state and prove

> **Deutsch's Theorem:** If $\mathcal{V}$ is a value function which satisfies **Physicality**, **Weak additivity**, **Substitutivity**, **Dominance**, and **Zero-sum**, then $\mathcal{V}$ is given uniquely by the Born rule:
>
> $$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}) = \sum_{x \in \sigma(X)} \langle\psi| \widehat{P}_X(x) |\psi\rangle \mathcal{P}(x) \equiv \sum_{c \in \mathcal{P}[\sigma(X)]} cW_{\mathcal{G}}(c). \tag{25}$$

The proof given below follows Deutsch's own proof rather closely (although some minor changes have been made for clarity or to conform to my notation and terminology.) In particular, though Deutsch often uses PE and ME (parts 1 and 2 of the Equivalence Theorem) he never derives the 3rd part, GE. As such, I make no use of it here (though see section 5).

As usual, $|\lambda_a\rangle$ will always denote an eigenstate of $\widehat{X}$ with some eigenvalue x_a. It will be convenient, for each operator $\widehat{X}$, to define the function id_X as the restriction of the identity map $\mathrm{id}(x) = x$ to the spectrum of $\widehat{X}$; note that $\mathrm{id}_{f(X)} \cdot f = f \cdot \mathrm{id}_X$. Because of PE, if we can prove the theorem for $\mathcal{P} = \mathrm{id}_X$ we can prove it for general $\mathcal{P}$:

$$\langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle \simeq \langle |\psi\rangle, \mathcal{P}(\widehat{X}), \mathrm{id}_X\rangle. \tag{26}$$

We will therefore take id_X as the 'default' payoff function, and will write just $\langle |\psi\rangle, \widehat{X}\rangle$ in place of $\langle |\psi\rangle, \widehat{X}, \mathrm{id}_X\rangle$.

Stage 1 *Let* $|\psi\rangle = \frac{1}{\sqrt{2}}(|\lambda_1\rangle + |\lambda_2\rangle)$. *Then* $\mathcal{V}(|\psi\rangle, \widehat{X}) = \frac{1}{2}(x_1 + x_2)$.

From **Weak additivity** and PE, we have

$$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathrm{id}_X) + k = \mathcal{V}(|\psi\rangle, \widehat{X}, \mathrm{id}_X + k) = \mathcal{V}(|\psi\rangle, \widehat{X} + k, \mathrm{id}_X) \tag{27}$$

Similarly, **Zero-Sum** together with another use of PE gives us

$$\mathcal{V}(|\psi\rangle, -\widehat{X}) = -\mathcal{V}(|\psi\rangle, \widehat{X}), \tag{28}$$

and combining (27) and(28) gives

$$\mathcal{V}(|\psi\rangle, -\widehat{X} + k) = -\mathcal{V}(|\psi\rangle, \widehat{X}) + k. \tag{29}$$

Now, let f be the function of reflection about the point $1/2(x_1 + x_2)$. Then $f(x) = -x + x_1 + x_2$. Provided that $\widehat{X}$ is non-degenerate and that the spectrum of $\widehat{X}$ is invariant under the action of f, the operator $\widehat{U}_f$, given by $\widehat{U}_f \widehat{X} \widehat{U}_f^\dagger = f(\widehat{X})$ is well-defined and leaves $|\psi\rangle$ invariant. ME then gives us

$$\mathcal{V}(|\psi\rangle, -\widehat{X} + x_1 + x_2) = \mathcal{V}(|\psi\rangle, \widehat{X}). \tag{30}$$

Combining this with (29), we have

$$\mathcal{V}(|\psi\rangle, \widehat{X}) = -\mathcal{V}(|\psi\rangle, \widehat{X}) + x_1 + x_2, \tag{31}$$

which solves to give $\mathcal{V}(|\psi\rangle, \widehat{X}) = \frac{1}{2}(x_1 + x_2)$, as required.

In the general case where $\widehat{X}$ is degenerate, or has a spectrum which is not invariant under the action of f, let $\mathcal{S}$ be the span of $\{|\lambda_1\rangle, |\lambda_2\rangle\}$ and let $\mathcal{V} : \mathcal{S} \to \mathcal{H}$ be the embedding map. ME then gives us

$$\langle |\psi\rangle, \widehat{X}|_{\mathcal{S}}\rangle \simeq \langle |\psi\rangle, \widehat{X}\rangle, \tag{32}$$

and the result follows.

Deutsch refers to this result, with some justice, as 'pivotal': it is the first point in the proof where a connection has been *proved* between amplitudes and probabilities. Note the importance in the proof of the symmetry of $|\psi\rangle$ under reflection, which in turn depends on the equality of the amplitudes in the superposition; the proof would fail for $|\psi\rangle = \alpha\,|\lambda_1\rangle + \beta\,|\lambda_2\rangle$, unless $\alpha = \beta$.

Stage 2 *If $N = 2^n$ for some positive integer n, and if $|\psi\rangle = (1/\sqrt{N})(|\lambda_1\rangle + \cdots + |\lambda_N\rangle)$, then*

$$\mathcal{V}(|\psi\rangle\,, \widehat{X}, \mathcal{P}) = (1/N)(x_1 + \cdots + x_N). \tag{33}$$

The proof is recursive on n, and I will give only the first step (the generalization is obvious). It relies on the method of forming composite games, hence on **Substitutivity**. Define:

- $|\psi\rangle = (1/2)(|\lambda_1\rangle + |\lambda_2\rangle + |\lambda_3\rangle + |\lambda_4\rangle)$;
- $|A\rangle = (1/\sqrt{2})(|\lambda_1\rangle + |\lambda_2\rangle)$; $|B\rangle = (1/\sqrt{2})(|\lambda_3\rangle + |\lambda_4\rangle)$;
- $y_A = (1/2)(x_1 + x_2)$; $y_B = (1/2)(x_3 + x_4)$.
- $\widehat{Y} = y_A\,|A\rangle\,\langle A| + y_B\,|B\rangle\,\langle B|\,.$

Now, the **game** $\mathcal{G} = \langle |\psi\rangle\,, \widehat{Y}\rangle$ has value $1/4(x_1 + x_2 + x_3 + x_4)$, by Stage 1. In the y_A branch, a reward of value $1/2(x_1 + x_2)$ is given; by **Substitutivity** the observer is indifferent between receiving that reward and playing the **game** $\mathcal{G}_A = \langle |\psi\rangle\,, \widehat{X}\rangle$, since the latter **game** has the same value. A similar observation applies in the y_B branch.

So the value to the observer of measuring $\widehat{Y}$ on $|\psi\rangle$ and then playing either G_A or G_B according to the result of the measurement is $1/4(x_1 + x_2 + x_3 + x_4)$. But the physical process which instantiates this sequence of **games** is just

$$\left(\sum_{i=1}^{4} \frac{1}{2}\,|\lambda_i\rangle\right) \otimes |\mathcal{M}_0\rangle \rightarrow \sum_{i=1}^{4} \frac{1}{2}\,|\mathcal{M}; x_i\rangle\,, \tag{34}$$

which is also an instantiation of the **game** $\langle |\psi\rangle\,, \widehat{X}\rangle$; hence, the result follows.

Stage 3 *Let $N = 2^n$ as before, and let a_1, a_2 be positive integers such that $a_1 + a_2 = N$. Define $|\psi\rangle$ by $|\psi\rangle = \frac{1}{\sqrt{N}}(\sqrt{a_1}\,|\lambda_1\rangle + \sqrt{a_2}\,|\lambda_2\rangle)$. Then*

$$\mathcal{V}(|\psi\rangle) = \frac{1}{N}(a_1 x_1 + a_2 x_2). \tag{35}$$

Without loss of generality (because of ME) assume $\mathcal{H}$ is spanned by $|\lambda_1\rangle$, $|\lambda_2\rangle$. Let $\mathcal{H}'$ be an $N-$dimensional Hilbert space spanned by states $|\mu_1\rangle\,, \ldots |\mu_N\rangle$, and define:

- $\widehat{Y} = \sum_{i=1}^{N} i\,|\mu_i\rangle\,\langle\mu_i|$.

- $f(i) = x_1$ for i between 1 and a_1, $f(i) = x_2$ otherwise.
- $\widehat{V} : \mathcal{H} \to \mathcal{H}'$ by

$$\widehat{V} \left|\lambda_1\right\rangle = \frac{1}{\sqrt{a_1}} \sum_{i=1}^{a_1} \left|\mu_i\right\rangle \quad \text{and} \quad \widehat{V} \left|\lambda_2\right\rangle = \frac{1}{\sqrt{a_2}} \sum_{i=a_1+1}^{N} \left|\mu_i\right\rangle . \tag{36}$$

Then since $f(\widehat{Y})\widehat{V} = \widehat{V}\widehat{X}$, we have

$$\langle \left|\psi\right\rangle, \widehat{X}, \mathcal{P}\rangle \simeq \langle \widehat{V} \left|\psi\right\rangle, f(\widehat{Y}), \mathrm{id}_{f(Y)}\rangle \simeq \langle \widehat{V} \left|\psi\right\rangle, \widehat{Y}, f \cdot \mathrm{id}_Y . \rangle \tag{37}$$

Since in fact $\widehat{V} \left|\psi\right\rangle$ is an equal superposition of all of the $\left|\mu_i\right\rangle$, the result now follows from Stage 2.

Deutsch then goes on to prove the result for arbitrary N (i. e., not just $N = 2^n$); however, that step can be skipped from the proof without consequence.

Stage 4 *Let a be a positive real number less than 1, and let $\left|\psi\right\rangle = \sqrt{a} \left|\lambda_1\right\rangle + \sqrt{1-a} \left|\lambda_2\right\rangle$. Then $\mathcal{V}(\left|\psi\right\rangle) = ax_1 + (1-a)x_2$.*

Suppose, without loss of generality, that $x_1 \leq x_2$, and make the following definitions:

- $\mathcal{G} = \langle \left|\psi\right\rangle \rangle$.
- $\{a_n\}$ is a decreasing sequence of numbers of form $a_n = A_n/2^n$, where A_n is a positive integer, and such that $\lim_{n\to\infty} a_n = a$. (This will always be possible, as numbers of this form are dense in the positive reals.)
- $\left|\psi_n\right\rangle = \sqrt{a_n} \left|\lambda_1\right\rangle + \sqrt{1-a_n} \left|\lambda_2\right\rangle$.
- $\left|\phi_n\right\rangle = (1/\sqrt{a_n})(\sqrt{a} \left|\lambda_1\right\rangle + \sqrt{a_n - a} \left|\lambda_2\right\rangle$.
- $\mathcal{G}_n = \langle \left|\psi_n\right\rangle \rangle$.
- $\mathcal{G}'_n = \langle \left|\phi_n\right\rangle \rangle$.

Now, from Stage 3 we know that $\mathcal{V}(\mathcal{G}_n) = a_n x_1 + (1-a_n)x_2$. We don't know the value of $\mathcal{G}'_n$, but by **Dominance** we know that it is at least x_1. Then, by **Substitutivity**, the value to the observer of measuring $\left|\psi_n\right\rangle$, then receiving x_2 euros if the result is x_2 and playing $\mathcal{G}'_n$ if the result is x_1, is at least as great as the $\mathcal{V}(\mathcal{G}_n)$.

But this sequence of games is just an instantiation of $\mathcal{G}$, for its end state is one in which a reward of x_1 euros is given with amplitude a and a reward of x_2 euros with amplitude $\sqrt{1-a}$. It follows that $\mathcal{V}(\mathcal{G}) \geq \mathcal{V}(\mathcal{G}_n)$ for all n, and hence that $\mathcal{V}(\mathcal{G}) \geq ax_1 + (1-a)x_2$.

A similar argument with an increasing sequence establishes that $\mathcal{V}(\mathcal{G}) \leq ax_1 + (1-a)x_2$, and the result is proved.

Stage 5 *Let α_1, α_2 be complex numbers such that $|\alpha_1|^2 + |\alpha_2|^2 = 1$, and let $|\psi\rangle = \alpha_1 |\lambda_1\rangle + \alpha_2 |\lambda_2\rangle$. Then $\mathcal{V}(|\psi\rangle) = |\alpha_1|^2 x_1 + |\alpha_2|^2 x_2$.*

This is an immediate consequence of ME and Stage 4: let $\widehat{U} = \sum_i \exp(i\theta_i) |\lambda_i\rangle \langle\lambda_i|$; then $\widehat{U}$ leaves $\widehat{X}$ invariant and so $\langle \widehat{U} |\psi\rangle, \widehat{X}\rangle \simeq |\psi\rangle, \widehat{X}\rangle$; but the eigenstate $\widehat{U} |\psi\rangle$ has only positive real coefficients, and so its value is given by Stage 4.

Stage 6 *If $|\psi\rangle = \sum_i \alpha_i |\lambda_i\rangle$, then $\mathcal{V}(|\psi\rangle) = \sum_i |\alpha_i|^2 x_i$.*

This last stage of the proof is simple and will not be spelled out in detail. It proceeds in exactly the same way as the proof of Stage 2: any n-term measurement can be assembled by successive 2-term measurements, using **Substitutivity**.

5 Alternate form of Deutsch's proof

A slight change of Deutsch's assumptions allows us to simplify the theorem and its proof. In this section we will be concerned with:

> **Deutsch's Theorem (variant form):** If $\mathcal{V}$ is any value function satisfying **Physicality**, **Dominance** and **Additivity**, it will be given by the Born Rule.

The proof proceeds via part 3 of the Equivalence Theorem (General Equivalence), which Deutsch did not use in his own proof. We define the *expected utility* of a **game** by $EU(\mathcal{G}) = \sum_c W_{\mathcal{G}}(c)\, c$, where the sum ranges over the distinct payoffs made.

As with Deutsch's own proof, we hold fixed the observable $\widehat{X}$ to be measured, and suppose $\mathcal{P}(x) = \mathrm{id}_X$ by default: this allows us to write $\langle |\psi\rangle \rangle$ for $\langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle$. In this case, we also write $EU(|\psi\rangle)$ for $EU(\mathcal{G})$.

Stage 1 *If $\mathcal{G}$ is an equally-weighted superposition of eigenstates of $\widehat{X}$, $\mathcal{V}(|\psi\rangle) = EU(\psi)$.*

Without loss of generality, suppose $|\psi\rangle = (1/N)(|\lambda_1\rangle + \cdots + |\lambda_N\rangle)$. Assume first that all the x_a are distinct, let π be an arbitrary permutation of $1, \ldots, N$, and define $\mathcal{P}_\pi$ by $\mathcal{P}_\pi(x_a) = x_{\pi(a)}$. Then by **Additivity**,

$$\sum_\pi \mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}_\pi) = \mathcal{V}(|\psi\rangle, \widehat{X}, \sum_\pi \mathcal{P}_\pi) = (n-1)! \sum_i x_i \tag{38}$$

since $\sum_\pi \mathcal{P}_\pi$ is just the constant payoff function that gives a payoff of $(n-1)!(x_1 + \cdots + x_N)$ irrespective of the result of the measurement.

But each of the $n!$ **games** $\langle |\psi\rangle, \widehat{X}, \mathcal{P}_\pi\rangle$ is a **game** in which each payoff x_i occurs with weight $1/N$. Hence, by GE, all have equal value, and that value is just $\mathcal{V}(|\psi\rangle)$. Thus, $n!\mathcal{V}(|\psi\rangle) = (n-1)!(x_1 + \cdots + x_N)$, and the result follows.

If the x_i are not all distinct, construct a sequence of operators $\widehat{X}_m$ with eigenstates $x_{m,n}$ all distinct, so that for each n $\{x_{m,n}\}$ is an increasing sequence tending to x_n. By **Dominance** this forces $\mathcal{V}(|\psi\rangle, \widehat{X}) \geq EU(|\psi\rangle)$; repeating with a decreasing sequence proves the result.

Stage 2 *If* $|\psi\rangle = \sum_i a_i |\lambda_i\rangle$*, where the* a_i *are all rational, then* $\mathcal{V}(|\psi\rangle) = EU(|\psi\rangle)$*.*

Any such state may be written

$$|\psi\rangle = (1/\sqrt{N}) \sum_i \sqrt{m_i} |\lambda_i\rangle , \tag{39}$$

where the m_i are integers satisfying $\sum_i m_i = n$. Such a game associates a weight m_i/N to payoff x_i.

But now consider an equally-weighted superposition $|\psi'\rangle$ of N eigenstates of $\widehat{X}$ where a payoff of x_1 is given for any of the first m_1 eigenstates, x_2 for the next m_2, and so forth. Such a game is known (from stage 1) to have value $(1/N)(m_1 x_1 + \cdots + m_N x_n) \equiv EU(|\psi\rangle)$. But such a **game** also associates a weight m_i/N to payoffs of value x_i, so by GE we have $\langle|\psi\rangle\rangle \simeq \langle|\psi'\rangle\rangle$ and the result follows.

Stage 3 *For all states* $|\psi\rangle$ *which are superpositions of finitely many eigenstates of* $\widehat{X}$*,* $\mathcal{V}(|\psi\rangle) = EU(|\psi\rangle)$*.*

By GE, it is sufficient to consider only states

$$|\psi\rangle = \sum_i \alpha_i |\lambda_i\rangle \tag{40}$$

with positive real α_1. Let $|\mu_i\rangle$, $(1 \leq i \leq N)$, be a further set of eigenstates of $\widehat{X}$, orthogonal to each other and to the $|\lambda_i\rangle$ and with eigenstates y_i distinct from each other and all strictly less than all of the x_i (that we can always find such a set of states, or reformulate the problem so that we can, is a consequence of GE). For each i, $1 \leq i \leq N$, let a_i^n be an increasing series of rational numbers converging on α_i^2, and define

$$|\psi_n\rangle = \sum_i \sqrt{a_i^n} |\lambda_i\rangle + \sum_i \sqrt{\alpha_i^2 - a_i^n} |\mu_i\rangle . \tag{41}$$

It follows from stage 2 that $\mathcal{V}(|\psi_n\rangle) = EU(|\psi_n\rangle)$, and from **Dominance** that for all n, $\mathcal{V}(|\psi\rangle) \geq \mathcal{V}(|\psi_n\rangle)$. Trivially $lim_{n\to\infty} EU(|\psi_n\rangle) = EU(|\psi\rangle)$, so $\mathcal{V}(|\psi\rangle) \geq EU(|\psi\rangle)$. Repeating the construction with all the y_i strictly greater than all the x_i gives $\mathcal{V}(|\psi\rangle) \leq EU(|\psi\rangle)$, and the result follows. □

6 Critique

Barnum, Caves, Finkelstein, Fuchs and Schank, in their critique of Deutsch's paper (Barnum, Caves, Finkelstein, Fuchs, and Schack 2000), make three objections:

1. Deutsch claims to derive probability from the non-probabilistic parts of quantum mechanics and decision theory. But the non-probabilistic part of decision theory already entails probability.

2. Deutsch's proof is technically flawed and contains a *non sequitur*.

3. Gleason's Theorem renders Deutsch's proof redundant.

Responding on Deutsch's behalf to these objections provides a useful analysis of the concepts and methods of his proof, and will be the topic of this section.

We begin with Barnum *et al*'s claim that the non-probabilistic part of decision theory already entails probabilities. They are referring results like the 'Probability representation theorem' quoted in section 3, by which we deduce that a rational agent confronted with uncertainty will always quantify that uncertainty by means of probabilities.[5] Since this theorem can be proved with no reference to quantum theory (in particular, with no use of the **Physicality** assumption), it certainly is not the case that Deutsch can claim to have derived the very *concept* of probability. (Of course, the representation theorem certainly makes no mention of the Born rule; Deutsch *can* still claim to have derived the specific probability rule in question.) In fact, Barnum *et al*'s criticism can be sharpened: Deutsch cannot claim, either, to have deduced the existence of *uncertainty* from his starting-point, for the decision-theoretic assumptions he makes apply only to a situation where uncertainty is already present. (In the language of section 1 this is to say that Deutsch's work arguably solves the Quantitative Problem but not the Incoherence problem.)

What of Barnum *et al*'s second criticism, of Deutsch's proof itself? Translating their objections into my notation, their concern is basically that Deutsch assumes, without justification, the rule $\mathcal{V}(|\psi\rangle, \widehat{X}) = \mathcal{V}(\widehat{U}|\psi\rangle, \widehat{U}\widehat{X}\widehat{U}^\dagger)$ (this is, in effect, their equation (13), which they believe Deutsch requires as an additional assumption).

Of course, (13) is a direct consequence of **Physicality** (via ME). The reason that this argument is unavailable to Barnum *et al* is that they treat the measurement process as *primitive*: to them (in this paper at any rate) a measurement is *axiomatically* specified by the operator being measured, and consideration of the physical process by which it is measured is irrelevant.

This brings up an interesting ambiguity in the phrase "non-probabilistic part of quantum mechanics", used in both papers. Barnum *et al* regard quantum mechanics in essentially the Dirac-von Neumann paradigm: there are periods where the dynamics are unitary and deterministic, followed by periods of stochastic evolution, corresponding to measurements and where the probabilities are given by the Born rule. In this framework, the "non-probabilistic part" naturally means the unitary, deterministic part, and the resulting theory is *physically incomplete*, in the sense that it does not describe even physically what happens during a measurement. This is the context in which they are able to offer what is effectively an alternative collapse rule (their equation (14)) which contradicts the Born rule.

To Deutsch, though, "quantum mechanics" means Everettian quantum mechanics, which is (at least from a God's-eye view) a deterministic theory. As

[5] In fact, they quote a related, but stronger result due to L. J. Savage, which may be found in Savage (1972) and is discussed in my (2002).

such, to Deutsch the "non-probabilistic part of quantum mechanics" means the *whole* of quantum mechanics, and there is no space for additional collapse rules — but there is also no axiomatic concept of measurement, hence the need for measurement neutrality to be either assumed or argued for.

Understanding the difference between Deutsch's conception of QM, and that of Barnum *et al*, is also central to seeing why the latter regard Gleason's Theorem as so central here.[6] For if we are looking for a probabilistic rule to describe what happens at collapse, then Gleason's Theorem tells us that this rule must be the Born Rule, provided only that its probabilities are non-contextual. As an added bonus, it proves that the physical state must be a (pure or mixed) Hilbert-space state, which in principle allows the state to be regarded simply as an epistemic notion (summarizing an agent's ignorance).

The situation is rather different in the Everett interpretation. Here it is the physical state that is our starting point, and the structure of a measurement is derived rather than postulated. As such, there is no logical space for a deduction of the state from the observables.

Nonetheless, might Gleason's Theorem provide us with the Born rule in the Everett interpretation also? It could be used in the following three-step proof of (the variant form of) Deutsch's theorem:

1. We begin by proving:

   Non-contextuality: If $\mathcal{V}$ is a value function satisfying **Dominance**, **Additivity** and **Physicality**, then

   $$\mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}) = \sum_{x \in \sigma(X)} \mathcal{V}(|\psi\rangle, \widehat{P}_X(x), \mathrm{id}_{P_X(x)})\mathcal{P}(x). \tag{42}$$

   (See the Appendix for the proof.)

2. Gleason's Theorem now tells us that

   $$\mathcal{V}(|\psi\rangle, \widehat{P}_X(x), \mathrm{id}_{P_X(x)}) = \mathsf{Tr}(\rho \widehat{P_X}(x)), \tag{43}$$

   where ρ is dependent on $|\psi\rangle$ but not on $\widehat{X}$.

3. Let $\mathcal{S}$ be the one-dimensional Hilbert space spanned by $|\psi\rangle$, and let $\widehat{U}$ be the embedding map of $\mathcal{S}$ into $\mathcal{H}$. Then $|\psi\rangle\langle\psi|\widehat{U} = \widehat{U}\mathsf{1}$, and so

   $$\langle |\psi\rangle, \mathsf{1}, \mathrm{id}_1\rangle \simeq \langle |\psi\rangle, |\psi\rangle\langle\psi|, \mathrm{id}_{|\psi\rangle\langle\psi|}\rangle, \tag{44}$$

   by ME. But since id_1 is a constant, the LHS game has value 1, hence so does the RHS one. In turn this forces $\Pr(1) = 1$, which is given only by $\rho = |\psi\rangle\langle\psi|$.)

[6] Recall that Gleason's Theorem tells us that for any map f from the projectors on a Hilbert space of 3+ dimensions to $[0, 1]$, such that if $\{\widehat{P}_i\}$ is an complete orthonormal set of projectors then $\sum_i f(\widehat{P}_i) = 1$, there exists some density operator ρ_f such that $f(\widehat{P}) = \mathsf{Tr}(\widehat{P}\rho_f)$.

Although this proof is valid,[7] it certainly does not obviate the importance of Deutsch's proof:

1. Even applying probabilistic notions to branching requires decision theory, to justify quantifying uncertainty by means of probability. (To be sure, the use of decision theory to justify probabilities long predates Deutsch.)

2. The central insight in Deutsch's work (other than the observation that decision theory allows us to get clear exactly how probabilities apply to the Everett interpretation) is that 'games' do not correspond one-to-one with physical situations — in my exegesis this is represented by the Equivalence Theorem, of course. Steps 1 and 3 in the proof above rely heavily on that theorem, without which Gleason's Theorem falls short of establishing the Born rule.

3. More prosaically, the proof in this section is vastly more complex than Deutsch's own proof. Proving **Non-contextuality** from the decision-theoretic axioms is scarcely simpler than proving Deutsch's theorem itself (since most of the hard work goes into proving the Equivalence Theorem, which both utilize) and that is before deploying Gleason's Theorem, the proof of which is far from trivial.

Gleason's Theorem, then, seems to offer little or no illumination of, or improvement to, Deutsch's proof.

7 Measurement Neutrality

We have seen that Deutsch's proof rests upon the observation that many **games** — i. e., triples $\langle |\psi\rangle, \widehat{X}, \mathcal{P}\rangle$ — correspond to a single physical game. This is possible because we are treating measurement, not as primitive, but as a physical process.

But this being so, there is a converse issue to address. Two *different* physical games can instantiate the same **game**; what of an agent who prefers one to the other? Such an agent's preferences would not be represented effectively by a function $\mathcal{V}$ on games.

Ruling this out is equivalent to assuming:

Measurement neutrality: A rational agent is indifferent between two physical games whenever they instantiate the same **game**.

The measurement neutrality assumption is hidden by Deutsch's (and my) notation. In effect it is the assumption that, provided that a given physical process fits the definition of a measurement of $\widehat{X}$ on $|\psi\rangle$, the details of how that measurement is done don't matter for decision-making purposes. I will give two examples to show why — despite appearances — it is not an altogether trivial assumption.

[7] It isn't valid in two-dimensional Hilbert spaces, of course — but it would be disingenuous to claim this as an advantage for Deutsch's proof. That proof (and my variant on it) makes extensive use of auxiliary systems, of arbitrarily high dimension.

Firstly, observe that is the explanation as to why Deutsch's theorem (which is, after all, a provable *theorem*) nonetheless has no implications for the probability problem in 'hidden-variable' theories, such as the de Broglie-Bohm theory (Bohm 1952; Holland 1993). For in such theories, the physical state of a system is represented not just by a Hilbert-space vector $|\psi\rangle$, but also by some set ω of hidden variables, so that the overall state is an ordered pair $\langle|\psi\rangle, \omega\rangle$. (In the de Broglie-Bohm theory, for instance, ω is the position of the corpuscles.) It is thus possible for two physical processes to agree as to the measurement carried out, the payoff given, and the Hilbert-space state, but to disagree as to ω — hence a rational agent might prefer one process to the other.

To see how this might happen in practice, specialize to the de Broglie-Bohm theory, and to position measurements. Suppose, in particular, that we consider a measurement of the spatial position of a particle in one dimension, and assume that the quantum state is $|\psi\rangle = (1/\sqrt{2})(|x\rangle + |-x\rangle)$, where $|x\rangle$ and $|-x\rangle$ are eigenvectors of position with eigenvalues x and $-x$ respectively, and that the payoff function is id_X. Stage 1 of Deutsch's proof (page 14) establishes that the value of this game is zero, relying in the process on the invariance of $|\psi\rangle$ under reflection about the origin; but unless the corpuscle state is also invariant about reflection, this argument cannot be expected to apply to the de Broglie-Bohm theory. And in fact, the corpuscle position *cannot* be invariant under reflection, except in conditions so extreme as to break the connection between outcomes and the Hilbert-space state entirely, for the possible outcomes of the measurement are $\pm x$ and so the corpuscle must have one of those two positions.[8]

Secondly, even in the context of the Everett interpretation measurement neutrality rules out the strategy of regarding all branches as equiprobable, independently of their amplitudes. For suppose I play a game where I measure a spin-half particle and gain money if the result is 'spin-up' but lose money otherwise. Measurement device #1 (improbably) results in one branch for the spin-up result and one branch for the spin-down result; device #2 incorporates a quantum random-number generator triggered by a spin-up result, so that there are a trillion spin-up branches and only one spin-down one. The equiprobability strategy tells me that I am as likely to gain as to lose if I use device #1, but almost certain to win if I use device #2 — yet measurement neutrality tells me that each is as good as the other.

(To be sure, this particular result is *already* implied if we adopt Saunders' subjective-uncertainty ('SU') viewpoint on quantum-mechanical branching (in which, recall, the correct attitude of an agent prior to branching is to expect that they will experience one of the outcomes, but to be uncertain as to which.) For device #2 is really just device #1, followed by the triggering of the randomizer, and so the SU description of its function is: "either spin-up will occur, or spin-down. If it's spin-up, some random process will occur in the innards of the

[8] We could, of course, try to get round this problem by considering a probability distribution over hidden variables and requiring the *distribution* to be symmetric. Fairly clearly, this forces a distribution assigning probability 0.5 to both $+x$ and $-x$. A Deutsch-style argument can now be applied, and yields the unedifying conclusion that *if* the particle is at position $+x$ 50 % of the time, it is rational to bet at even odds that it will be found there when measured.

measuring device (but it won't affect my winnings.)" Looked at this way, the equiprobability assumption is already in trouble.)

The instinctive response to measurement neutrality, nonetheless, is usually that it *is* trivial — who cares exactly *how* a measurement device works, provided that it works? What justifies this instinctive response is presumably something like this: let A and B be possible measurement devices for some observable $\widehat{X}$, and for each eigenvalue x of $\widehat{X}$ let the agent be indifferent between the x-readout states of A and those of B. Then if the agent is currently planning to use device A, he can reason, "Suppose I get an arbitrary result x. Had I used device B I would still have got result x, and would not care about the difference caused in the readout state by changing devices; therefore, I should be indifferent about swapping to device B."

The only problem with this account is that it assumes that this sort of counterfactual reasoning is legitimate in the face of (subjective) uncertainty, and this is at best questionable (see, e. g., Redhead (1987) for a discussion, albeit not in the context of the Everett interpretation).

For a defence secure against this objection, consider how the traditional Dirac-von Neumann description of quantum mechanics treats measurement. In that account, a measurement device essentially does two things. When confronted with an eigenstate of the observable being measured, it reliably evolves into a state which displays the associated eigenvalue. In addition, though, when confronted with a superposition of eigenstates it causes wave-function collapse onto one of the eigenstates (after which the device can be seen as reliably evolving into a readout state, as above).

In the Dirac-von Neumann description, it is rather mysterious why a measurement device induces collapse of the wave-function. One has the impression that some mysterious power of the device, over and above its properties as a reliable detector of eigenstates, induces the collapse, and hence it is *prima facie* possible that this power might affect the probabilities of collapse (and thus that they might vary from device to device) — this would, of course, violate measurement neutrality. That this is not the case, and that the probabilities associated with the collapse are dependent only upon the state which collapses (and indeed are equal to those stipulated by the Born rule) is true by fiat in the Dirac-von Neumann description.

It is a strength of the Everett interpretation (at least as seen from the SU viewpoint) that it recovers the subjective validity of the Dirac-von Neumann description: once decoherence (and thus branching) occurs, subjectively there has been wave-function collapse. Furthermore there is no "mysterious power" of the measurement device involved: measurement devices by their nature amplify the superposition of eigenstates in the state to be measured up to macroscopic levels, causing decoherence, and this in turn leads to subjective collapse.

But this being the case, there is no rational justification for denying measurement neutrality. For the property of magnifying superpositions to macroscopic scales is one which all measurement devices possess equally, by definition — so if this is the only property of the devices relevant to collapse (after which the system is subjectively deterministic, and so differences between measurement devices are irrelevant) then no other properties can be relevant to a rational allocation of probabilities. The only relevant properties must be the state being measured, and the particular superposition which is magnified to macroscopic scales — that is, the state being measured, and the observable being measured on it.

8 Conclusion

I have shown that Deutsch's approach does indeed allow a derivation of the Born rule, from the following premises:

1. The correctness of the Everett interpretation.
2. The validity of regarding quantum branching, within the Everett interpretation, as uncertain (at least subjectively).
3. A fairly strong set of decision-theoretic axioms.
4. Measurement neutrality.

All four are needed. Without the Everett interpretation we cannot give a realist description of QM which eschews hidden variables of any sort, objectively stochastic dynamics, and an *a priori* privileged role for the observer. Without some license for agents to regard quantum branching as uncertain we cannot import classical decision theory into QM. Without decision theory itself we have no license to transform uncertainty into probability, and none of the constraints on those probabilities that allow Deutsch's Theorem to be proven. And without measurement neutrality we cannot draw any worthwhile conclusions from Deutsch's theorem, for it is the assumption that connects the value function $\mathcal{V}$ with real decision-making.

All four are reasonable, however. The Everett interpretation's various (non-probabilistic) foundational problems appear tractable; work by Saunders, Vaidman and others seems to justify the application of uncertainty-based concepts to branching; Deutsch's decision theory, though based on quite strong axioms, seems perfectly reasonable for small-scale betting;[9] measurement neutrality is at the least a plausible assumption, and may well be defensible by either of the routes sketched out in section 7.

Deutsch's own conclusion claims that "A decision maker who believes only the non-probabilistic part of the theory, and is 'rational' in the sense defined by a strictly non-probabilistic restriction of classical decision theory" will make decisions according to the Born rule." Tacit in Deutsch's paper is that 'the non-probabilistic part of the theory' means no-collapse quantum mechanics, Everett-interpreted but without prior assumptions about probability; it is less clear what the 'non-probabilistic restriction of classical decision theory' really means, but if it simply means classical decision theory, shorn of *explicit* assumptions

[9] In any case, Deutsch's decision theory can be very substantially weakened; see Wallace (2002).

about probabilities and applied to branching events as if they were uncertain-outcome events, then his claim seems essentially correct. The implications for a satisfactory resolution of the quantitative probability problem are then obvious — and profoundly important.

Acknowledgements

For valuable discussions, I am indebted to Hannah Barlow, Katherine Brading, Harvey Brown, Jeremy Butterfield, Adam Elga, Chris Fuchs, Hilary Greaves, Adrian Kent, Chris Timpson, Wojciech Zurek, to all those at the 2002 Oxford-Princeton philosophy of physics workshop, and especially to Simon Saunders and David Deutsch. Jeremy Butterfield and Simon Saunders also made detailed and helpful comments on the longer version of this paper.

Appendix: proofs of the Probability Representation Theorem and Non-Contextuality

> **Lemma (Linearity):** If $\mathcal{V}$ satisfies **Additivity** and **Dominance**, then for any sets of real numbers $\{a_i\}_{i=1}^N$ and payoffs $\{\mathcal{P}_i\}_{i=1}^N$,
>
> $$\mathcal{V}(|\psi\rangle, \widehat{X}, \sum_{i=1}^N a_i \mathcal{P}_i) = \sum_{i=1}^N a_i \mathcal{V}(|\psi\rangle, \widehat{X}, \mathcal{P}_i). \tag{45}$$

Proof of lemma: We will suppress $\widehat{X}$ and $|\psi\rangle$, writing just $\mathcal{V}(\mathcal{P})$ for $\mathcal{V}(|\psi\rangle,\widehat{X},\mathcal{P})$. Let a be any positive real number and let $\{k_n\}$ and $\{m_n\}$ be sequences of integers such that $\{k_m/m_n\}$ is an increasing sequence tending to a. By **Dominance** and **Additivity** we have $m_n\mathcal{V}(a\mathcal{P}) \geq k_n\mathcal{V}(\mathcal{P})$ for all n, and hence $\mathcal{V}(a\mathcal{P}) \geq a\mathcal{V}(\mathcal{P})$. Repeating this with a decreasing sequence, we get $\mathcal{V}(a\mathcal{P}) = a\mathcal{V}(\mathcal{P})$ for any $a \geq 0$; the extension to negative a is trivial (just use **Zero-sum**) and the full result follows from **Additivity**. □

Proof of representation theorem: For any $x \in \sigma(\widehat{X})$, define $\delta_x(y)$ as equal to 1 when $y = x$, and equal to 0 otherwise. Any payoff function $\mathcal{P}$ for $\sigma(\widehat{X})$ can be expressed uniquely as $\mathcal{P} = \sum_{x\in\sigma(X)} \mathcal{P}(x)\delta_x$, and so by **Linearity** we have $\mathcal{V}(\mathcal{P}) = \sum_{x\in\sigma(X)} \mathcal{P}(x)\mathcal{V}(\delta_x)$; setting $\Pr(x) = \mathcal{V}(\delta_x)$ establishes (23), and putting $\mathcal{P}(x) = 1$ for all x gives (24) as a special case. □

Proof of Non-contextuality: By PE and **Linearity** we have

$$\mathcal{V}(\mathcal{G}) = \sum_{x\in\sigma(X)} \mathcal{P}(x)\mathcal{V}(|\psi\rangle, \widehat{X}, \delta_x) = \sum_{x\in\sigma(X)} \mathcal{P}(x)\mathcal{V}(|\psi\rangle, \delta_x(\widehat{X}), \mathrm{id}_{\delta_x(X)}) \tag{46}$$

and the result is proved upon observing that $\delta_x(\widehat{X}) = \widehat{P}_X(x)$. □

References

Albert, D. and B. 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.

Bohm, D. (1952). A Suggested Interpretation of Quantum Theory in Terms of "Hidden" Variables. *Physical Review 85*, 166–193.

Dennett, D. C. (1991). Real Patterns. *Journal of Philosophy 87*, 27–51. Reprinted in *Brainchildren*, D. Dennett, (London: Penguin 1998) pp. 95–120.

Deutsch, D. (1985). Quantum Theory as a Universal Physical Theory. *International Journal of Theoretical Physics 24*(1), 1–41.

Deutsch, D. (1996). Comment on Lockwood. *British Journal for the Philosophy of Science 47*, 222–228.

Deutsch, D. (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.

Deutsch, D. (2001). The Structure of the Multiverse. Available online at http://xxx.arXiv.org/abs/quant-ph/0104033.

DeWitt, B. (1970). Quantum Mechanics and Reality. *Physics Today 23*(9), 30–35. Reprinted in (DeWitt and Graham 1973).

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

Donald, M. (1997). On Many-Minds Interpretations of Quantum Theory. Available online at http://www.arxiv.org/abs/quant-ph/9703008.

Everett III, H. (1957). Relative State Formulation of Quantum Mechanics. *Review of Modern Physics 29*, 454–462. Reprinted in DeWitt and Graham (1973).

Holland, P. (1993). *The Quantum Theory of Motion*. Cambridge: Cambridge University Press.

Lockwood, M. (1989). *Mind, Brain and the Quantum: the compound 'I'*. Oxford: Blackwell Publishers.

Lockwood, M. (1996). 'Many Minds' Interpretations of Quantum Mechanics. *British Journal for the Philosophy of Science 47*, 159–188.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Redhead, M. (1987). *Incompleteness, Nonlocality and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*. Oxford: Oxford University Press.

Saunders, S. (1995). Time, Decoherence and Quantum Mechanics. *Synthese 102*, 235–266.

Saunders, S. (1997). Naturalizing Metaphysics. *The Monist 80*(1), 44–69.

Saunders, S. (1998). Time, Quantum Mechanics, and Probability. *Synthese 114*, 373–404.

Savage, L. J. (1972). *The foundations of statistics* (2nd ed.). New York: Dover.

Sudbery, A. (2000). Why am I me? and why is my world so classical? Available online at http://www.arxiv.org/abs/quant-ph/0011084.

Vaidman, L. (1998). On Schizophrenic Experiences of the Neutron or Why We Should Believe in the Many-Worlds Interpretation of Quantum Theory. *International Studies in Philosophy of Science 12*, 245–261. Available online at http://www/arxiv.org/abs/quant-ph/9609006.

Vaidman, L. (2001). The Many-Worlds Interpretation of Quantum Mechanics. To appear in the Stanford Encyclopedia of Philosophy; temporarily available online at http://www.tau.ac.il/ vaidman/mwi/mwst1.html.

Wallace, D. (2001a). Everett and Structure. Forthcoming in *Studies in the History and Philosophy of Modern Physics*; available online at http://xxx.arXiv.org/abs/quant-ph/0107144 or from http://philsci-archive.pitt.edu.

Wallace, D. (2001b). Worlds in the Everett Interpretation. Forthcoming in *Studies in the History and Philosophy of Modern Physics*; available online at http://www.arxiv.org/abs/quant-ph/0103092 or from http://philsci-archive.pitt.edu.

Wallace, D. (2002). Quantum Probability and Decision Theory, Revisited. Longer version of the present paper; available online from http://www.arXiv.org or from http://www.philsci-archive.pitt.edu.

Zurek, W. H. (1998). Decoherence, Einselection and the Existential Interpretation (The Rough Guide). *Philosophical Transactions of the Royal Society of London A356*, 1793–1820.