An Overview of Probability in the Everett Interpretation

Simon Ellersgaard Nielsen

July 31, 2010

Abstract

The Everett interpretation faces no challenge more pertinent than the problem of how to square manifest determinism with the probability familiar from the conventional quantum algorithm. In this paper I review recent attempts at solving this problem at the conceptual and quantitative (Born) level. I conclude that momentous progress has been made, but certain aspects (subjective uncertainty, decision-theoretic axioms) still require further development.

Contents

1 Introduction: Everett's Many Worlds Proposal

1.1 The Measurement Problem

The inherent incompleteness of the conventional quantum mechanical algorithm (often referred to as the Copenhagen interpretation) is well-known to scholars of foundational physics. Recall that the generic quantum state vector $|\psi\rangle$ is postulated to be governed by two manifestly incompatible dynamical laws; on the one hand, if a system is disturbed by a measurement of some physical observable (formally represented by a Hermitian operator, $\boldsymbol{X}$), the state is posited to undergo an instantaneous non-linear collapse into one of the eigenstates of $\boldsymbol{X}$ viz. $\{|x_i\rangle\}|_{i=1}^{n}$ with 'Born' probability $p(x_i) = |\langle x_i|\psi(t)\rangle|^2 \in [0,1]$. In honor of John von Neumann, [17], we designate this fundamentally indeterministic law '**process I**'. On the other hand, if left to itself, the state will evolve in a deterministic and linear manner in accordance with $|\psi(t_1)\rangle = \boldsymbol{U}(t_0,t_1)|\psi(t_0)\rangle$, where $\boldsymbol{U}(t_0,t_1)$ is an energy-dependent unitary operator on the n-dimensional Hilbert space $\mathcal{H}_S$ - call this '**process II**'. Symptomatic for the latter law is - of course - Schödinger evolution, in which case

$$|\psi(t_1)\rangle = \exp\left(-\frac{i}{\hbar}\int_{t_0}^{t_1} \boldsymbol{H} dt\right)|\psi(t_0)\rangle \tag{1}$$

where $\boldsymbol{H}$ is the systemic Hamiltonian.

The *problem* with this picture is that the line of demarcation between these laws (i.e. the measurement process) is notoriously nebulous. *Prima facie*, one might conjecture that the interaction between a macroscopic apparatus and a quantum system is tantamount to process I kicking into effect, yet a moment's thought should reveal that there is no non-arbitrary definition of what constitutes the macro-realm. Indeed, there is seemingly nothing to bar us from treating the measurement apparatus *per se* as a hyper-complex quantum system (it is, after all, supervenient upon a multitude of molecules) whence the deterministic process II will give rise to a result contradicting the former (a so-called ”macroscopic superposition”).

To remedy this impasse it is sometimes argued that a stringent definition of the measurement process requires reference to an altogether different ontological category viz. *consciousness*. Nevertheless, it should be fairly obvious that this ”regression” into metaphysics is very controversial: not only does the very fabric of mentality elude us, but we must also enquire at which point in our evolutionary history consciousness reached a sufficient level of complexity for the universal wave function (state-vector) to collapse? These are excruciatingly difficult questions, which reek of empirical vacuity, wherefore a more physically level-headed approach is desirable.

1.2 Everett's Worlds

In this paper we shall concern ourselves with a radical solution to the measurement problem which eradicates the stochastic process I *in toto* from the axioms of quantum mechanics (i.e. which rejects the very notion that the ”act of measuring” is an ontologically significant concept). Explicitly, we shall contemplate an interpretation attributed to the late Hugh Everett III, [8], which during the past decade has been extensively developed by the philosophy of physics community at the University of Oxford (cf. [5], [10], [16], [19], etc.).

The central idea of the Everett interpretation is that the universal wave function Ψ exhausts our lower level (fundamental) ontology of the material universe, where Ψ *at all times* is governed by the deterministic process II. To appreciate the implications of this, consider the measurement of spin $\boldsymbol{S}_z$ on an electron in the positive $\boldsymbol{S}_x$ eigenstate $|\uparrow_{x^+}\rangle$ by an observer O with a reliable apparatus M. This will result in the evolution:

$$|O_0\rangle \otimes |M_0\rangle \otimes |\uparrow_{x^+}\rangle \xrightarrow{\boldsymbol{U}} \frac{1}{\sqrt{2}}\left(|O_{z^+}\rangle \otimes |M_{z^+}\rangle \otimes |\uparrow_{z^+}\rangle + |O_{z^-}\rangle \otimes |M_{z^-}\rangle \otimes |\uparrow_{z^-}\rangle\right) \tag{2}$$

Here, $|O_0\rangle$ is the initial (0) quantum state vector representing the observer, whereas the subscripts z^+ and z^- represent the observer with a particular belief as to the outcome of the spin-z measurement (analogously, $|M_0\rangle$ is the initial state vector of the apparatus etc.).

Now, on the conventional interpretation, the superposition in (2) is a state where there is no matter of fact concerning the outcome of the measurement.

But we know from direct introspection that this is erroneous: measurements *do* have definite outcomes. Concordantly, if we are to make any sense of Everett's conjecture, it seems that we literally must construe evolution into *entangled* states as the *branching of worlds* (observe that this principle makes no reference to the terms 'observer', 'measurement' or 'macro-'). Thus, (2) should be seen as a case where a single universe fissions into two distinct worlds: one in which the observer records a positive eigenvalue $+\frac{\hbar}{2}$ for his spin measurement along the z-axis, and one in which he records a negative eigenvalue $-\frac{\hbar}{2}$. Since the universal state vector by assumption has proliferated in this manner since the beginning of time, it follows that the totality of facts is a *multiverse* comprised of a near-infinite number of worlds in one-to-one correspondence with the terms in the superposition. (As an aside, observe that this many-worlds scenario is not tantamount to an affirmation of David Lewis' *modal realism* postulate; the Everettian worlds are naturally restricted by the requirement of compatibility with unitary evolution from a particular set of *initial conditions*).

1.3 The Basis Problem

Alas, there is a sense in which this scheme falls short of being well-defined: as it happens there is nothing in the quantum formalism *per se* which stipulates which basis in the Hilbert space one should employ to identify the nature of the physical worlds. More concretely, the superposition in (2) is mathematically equivalent to

$$\frac{1}{\sqrt{2}}\left(|\zeta^{+}\rangle \otimes |\uparrow_{x^{+}}\rangle + |\zeta^{-}\rangle \otimes |\uparrow_{x^{-}}\rangle\right) \tag{3}$$

where $|\zeta^{\pm}\rangle = \frac{1}{\sqrt{2}}(|O_{z^{+}}\rangle \otimes |M_{z^{+}}\rangle \pm |O_{z^{-}}\rangle \otimes |M_{z^{-}}\rangle)$, which looks like two worlds in each of which the electron has a definite value for spin along the x-axis. Since $\boldsymbol{S}_x$ does not commute with $\boldsymbol{S}_z$ a flagrant contradiction now seems to undermine the Everettian project in its current formulation, [2].

Proposed solutions to this problem generally make the assumption that the relevant worlds must somehow be added *explicitly* to the quantum formalism. However, it has recently been argued that this is unnecessary: in particular, it appears that the *decoherence program* can provide a gateway out of this conundrum insofar as we are willing to embrace a certain measure of vagueness in our macro-ontology, [21], [22]. The idea is that the kind of Hamiltonians which actually obtain, encode structures which behave dynamically essentially as classical worlds and which *for all practical purposes* are causally isolated from one another. Provided that we adopt a Dennettian approach ([6]) according to which macro-structures ultimately must be seen as emergent (fuzzy) patterns in the micro ontology (a step which seems to harmonize with the reductionism inherent to the scientific enterprise), our near-classical, near-causally isolated worlds may be treated as the physical structures in which we are imbedded. In other words, it would be superfluous to add an auxiliary determination of the relevant worlds to the quantum formalism - the very nature by which quantum mechanics operates essentially selects a *preferred basis*[1].

1.4 The Probabilistic Predicament

As I see it, a more pressing objection to the Everett program is the bipartite problem of probability. Firstly, there is a challenge of downright **incoherence**: since the many worlds interpretation has a manifestly deterministic dynamics, how can it even make sense to assign probabilities to outcomes? Secondly, even if Everettian stochasticity somehow does make sense, surely there is also a **quantitative** problem to be solved: for why should the probabilities be given by the Born rule as in the conventional quantum algorithm? These concerns will be the topic of consideration for the remaining part of this paper; in particular we shall scrutinize recent attempts by Deutsch ([7]), Wallace ([19], [20]) and Greaves ([10]) to remedy this deplorable situation[2].

[1] Here's how decoherence works: let $|\epsilon_{z+}\rangle$ be the environmental state which has adjusted itself to the $|\uparrow_{z+}\rangle$ state, and let $|\epsilon_{z-}\rangle$ have an analogous interpretation, then the orthogonality relation $\langle\epsilon_{zi}|\epsilon_{zj}\rangle \approx \delta_{ij}$, $i, j \in \{+, -\}$ is an excellent approximation. This means that if we write down the reduced density matrix on the electron system ($\mathcal{H}_S$) of the entangled state $|\chi\rangle = (|\epsilon_{z+}\rangle \otimes |\uparrow_{z+}\rangle + |\epsilon_{z-}\rangle \otimes |\uparrow_{z-}\rangle)/\sqrt{2}$ viz.

$$\rho_{\text{red}} = \mathrm{Tr}_{\mathcal{H}_\epsilon}(|\chi\rangle\langle\chi|)$$

then $\rho_{\text{red}} \approx 0.5|\uparrow_{z+}\rangle\langle\uparrow_{z+}| + 0.5|\uparrow_{z-}\rangle\langle\uparrow_{z-}|$. The latter equation may be interpreted as "with probability 0.5 the particle definitely has spin-up and with probability 0.5 the particle definitely has spin-down". I.e. the wave will *appear* to have collapsed and classicality will appear to have been recovered.

[2] Initial reactions to the probability problem tend to be skeptical: can't we simply assign a probability-weight to an Everettian branch in accordance with how likely the mind of the observer is to find itself in that branch post-measurement? No; that would be to commit ourselves to an outdated and dubious *Cartesian dualism*: recall that the body with which *O* is identified will split into two copies when the world branches. Hence, if we make the common assumption that the mind is an emergent property of the neurological structures in the brain, both post-branching observers will be conscious and claim to be the original observer.

2 Elusive Probability

2.1 Credence & Objective Probability

At this point it is incumbent on us to take a detour into the philosophical foundations of probability: as a starting point it is worthwhile illuminating the dual semantics of the term. *For one*, there is subjective probability (coined '**credence**') which we employ as a measure of our personal degree of belief in a given hypothesis. This phenomenon is in itself fairly well defined: from Savage's *decision theoretic* considerations, we can show that the agent must quantify his uncertainty in terms of probability (or alternatively contravene some plausible set of principles connecting rational behavior with personal preferences) - this will be explicated below. Disregarding eliminativism, we may safely assert that credence falls short of capturing robust, observer independent probability (the so-called '**objective probability**', or **OP**, [20]) which is found in the physical sciences. Perhaps surprisingly, we have yet to uncover a truly satisfactory analysis of OP: after all, can't we simply take the pedestrian view that probability is the *relative frequency* of events in a boundless number of trials, i.e. $\frac{n_x}{n_\infty} = p(x)$, where n_x is the number of trials with outcome x and n_∞ is the total number of trials tending to infinity? Despite an incontestable intuitive appeal, the problem with this view is that we are epistemically barred from accessing anything like n_∞. The best we can do is metaphysically slippery extrapolations of our limited sequence observations, which very well could deviate wildly from the results in the infinite trial limit (for example, upon flipping a fair coin, it is in principle possible to get 100 'heads' in a sequence of 100 trials). And although this does not altogether undermine the frequency interpretation of OP, it certainly does establish that the program is in dire need of more careful scrutiny.

Whilst our understanding of the nature of objective probability borders on the non-existent, we do however have a fairly good idea of how it integrates itself into our general conceptual scheme, viz. via Lewis' *principal principle* (**PP**):

Definition 1. (PP) If an agent knows that the objective probability of an event ξ is p, then that agent is rationally compelled to set his or her personal credence in ξ equal to p. More formally, let X_p be the proposition that $OP(\xi) = p$ and A be any admissible proposition compatible with X_p then

$$Cr(\xi|A \wedge X_p) = p. \tag{4}$$

This, in turn, enables us to offer at least a *functional definition* of OP (see [20]):

Definition 2. (Functional OP) If some physical theory T defines some magnitude c for events, then c is OP just if any agent believing T is compelled to set his or her credences equal to c. That is, c is OP iff: $\forall\xi$ if $[T \wedge B \rightarrow c(\xi) = p]$, where B is any admissible background information, then

$$Cr(\xi|B \wedge T) = p. \tag{5}$$

Although this definition delicately evades an explication of what OP actually is, the definition is not without merit: for consider the case where B is the proposition P_c='c satisfies the functional definition' (or more loosely the *principal principle*). Suppose that we accept both T and P_c (which *inter alia* implies that we accept the existence of OP), then we should set our credence in the event ξ equal to $c(\xi)$. Provided that $c(\xi)$ computed from T significantly exceeds our preliminary credence in ξ, $Cr(\xi)$, we should accordingly treat $[T \wedge P_c]$ as the explanans for ξ (and thence take ξ as evidence for T and P_c). This is vividly demonstrated using Bayes' Theorem:

$$Cr(T \wedge P_c|\xi) \stackrel{\text{def}}{=} \frac{Cr(\xi|T \wedge P_c)Cr(T \wedge P_c)}{Cr(\xi)}; \tag{6}$$

hence, using the PP:

$$\frac{Cr(T \wedge P_c|\xi)}{Cr(T \wedge P_c)} = \frac{c(\xi)}{Cr(\xi)}. \tag{7}$$

wherefore *we can acquire powerful evidence for OP (functionally defined), in spite of being left in the dark about its true nature*[3].

[3] There is an interesting question as to whether the functional definition of OP is *complete*. Here's what I mean: in deriving the functional definition from the *principal principle*, can we really be sure that we haven't omitted some property Q of OP, such that something satisfying definition 2, but not possessing Q, does *not* qualify as OP? Semantically this objection certainly seems defensible (people do refer to 'probability' ambiguously). However, if we look at the OP from a strictly scientific perspective (i.e. as the aspect apparently called for in experimental situations) it is less clear what kind of evidence we could possibly have for Q. For as Wallace points out ([20], p. 662), consider the case where we have some theory T_1 which assigns high *genuine* probabilities to experimental data which we in fact observe (we say that T_1 explains the data well). Furthermore, suppose we have some other theory T_2 which involve only Q-lacking OP-satisfying quantities (*'quasi-probabilities'*) s.t. T_2 assigns quasi-probabilities to the data *exactly* equal to the genuine probabilities assigned by T_1. Under these circumstances, we may reasonably question the relevance of Q: for in the process in which we continue to test our theories T_1 and T_2, 'probability' (quasi or genuine) is tied to our observations solely through the PP, which - mind you - is oblivious to whether property Q is satisfied.

2.2 Cautious Functionalism

Of course, as philosophers we would like to know what actually lies behind the concept of objective probability. One option is to jump on the functionalist bandwagon ([20], p. 659), according to which OP is a set of physically definable properties which can be defined independently of PP (but which can be shown to satisfy the functional definition). Alas, projects within this field remain at a rather rudimentary stage of development: for instance, for the aforementioned frequentist program, it is notoriously difficult to establish that the functional definition of probability is satisfied. However, it is profoundly questionable whether this "failure of imagination" on our part is enough for us to abandon functionalism altogether. For consider the alternatives, viz. *primitivism* (the functional OP definition is taken as a fundamental natural law[4], [20], p. 660) or *eliminativism* (the nihility of OP, [20], p. 660) - both of which patently are remarkably radical strategies.

In case of primitivism, we are opting for a *rationality principle* (which, I think, is what functional OP amounts to) as a natural law on a par with the dynamical laws of space-time and field theories. Clearly, this would have some rather counterintuitive modal implications: we could, for example, consider a possible world which is completely identical to our universe from a physical perspective, but which nonetheless has entirely different implications for what it means to be rational (i.e. where '$\neg$Functional OP' is true). But surely, if we have our doubts about whether there are such physically identical rationally different possible worlds, then eliminativism should strike us as absurd: it will, for instance, have us believe that the half-life of a decaying nucleus is a subjective phenomenon. Thus, it seems reasonable to proceed along the lines of a *cautious* mode of functionalism.

Cautious functionalism operates as follows: given a theory T which defines some property c (seemingly playing the role of probability), collect evidence for the joint hypothesis $[T \wedge P_c]$ as above. However, bear in mind that T comes with an attached promissory note: eventually we will need to account for (i) how to define c independently of the PP, and (ii) given this characterization of c, how P_c can be derived. And although T might be construed as *phenomenological* until (i) and (ii) have been answered satisfactorily, T can still be highly explanatory (as is the case with thermodynamics and special relativity[5]).

[4] Thus the functional OP definition is *not* to be shown to hold for an independently definable property c. Rather, it is simply postulated to be true of c and it defines c via its role in the law.

[5] It is generally recognized that physical theories fall within two different classes. Firstly, there are those theories which claim to be *constructive* in the sense that they have a rigid dynamics which describes the behavior of purportedly ontologically significant entities. Classical

3 Subjective Uncertainty in the Everett Interpretation

3.1 Saunder's Argument

Let us contemplate the problem of *incoherence* raised in subsection 1.4: employing the vernacular above, the quantum algorithm assigns objective probabilities to the possible outcomes of (quantum) experiments. A rational agent is concordantly compelled to match his or her credences with these objective chances (if known), and thence the agent harbors a quantitative measure of uncertainty with respect to the experimental outcome(s). The Everettian's task is thus to account for how *subjective uncertainty* (**SU**) can be squared with the manifest determinism of the interpretation, where every possible outcome is realized.

For this reason, Simon Saunders ([16]) has offered us this intuition pump: consider the case of ″classical″ fission, where an individual O with a brain of perfect functional bilateral symmetry, has his *corpus callosum* severed in a Sperryian surgical procedure. The left cerebral hemisphere is subsequently placed in the (brainless) body O', whilst the right cerebral hemisphere is place in the (brainless) body O'' and the required neural connections are established. Now in ordinary, non-fission situations, the fact that O expects to become his future self, is brought about by the fact that his present and future selves are connected by the right causal and structural relations. What, then, should O expect post-fission i.e. when he has more than one future self? Saunders sees three possibilities:

1. O should expect abnormality: e.g. to somehow end up as *both* O' and O''.

2. O should expect to become O' or O'', but not both.

3. O should expect oblivion i.e. to cease to exist upon having his brain severed.

kinetic theory is suggested to be such a theory (of course, nobody would take the ideas of hard-spheres and Newtonian motion as more than an approximation today). Secondly, there are those kind of theories (coined *phenomenological* or *principle*) which we know do not reflect physical reality at its deepest level: theories, which FAPP describe the behavior of macro-ontological entities and concepts well, but which we hope ultimately will be replaced by more fundamental theories. Thermodynamics, e.g., is utterly oblivious to what gasses actually consist of, yet it is remarkably good at describing how they behave. But surely, we would still like to know what a measurable thing like *pressure* actually is! In a similar vein Einstein hoped [*The Times*, 1919] that the kinematics of the acclaimed special theory of relativity might find a more constructive derivation, rather than its current principle form.

Although we certainly can imagine the minds of O' and O'' to be interconnected via some instantaneous, all-incompassing telepathic link, (1) should strike the reader as highly implausible. For under the reasonable assumption that consciousness should be treated as an emergent property from the appropriate physical structures (the brain), we should abandon (1) immediately, since no such thing exists between O' and O''. Likewise, (3) is preposterous: intuitively, if only one cerebral hemisphere was transplanted (and the other annihilated) we would be inclined towards saying that the agent O has survived (after all, the post-surgery individual would be mentally just like O and there is a direct causal link between the person stages (elements of physical continuity)). Why, then, should O expect oblivion just because his other cerebral hemisphere also is given a "new home"? By the process of elimination, (2) is the one to opt for, and in the absence of some strong criterion as to *which* of O' and O'' the original O will become, he will have to treat that question as subjectively indeterministic. And this is relevant to the topic, because the classicality of the fission by no means is integral to the argument: we could equally well consider the kind of person splitting occurring in quantum branching i.e. *agents in an Everettian universe should treat their own branching as subjectively indeterministic*.

The problem with this argument is the presupposition of blatantly archaic notions about the relevance of classical personal identity (which *per definitionem* is an equivalence relation[6]), despite a metaphysically shaky foundation. I see no justifiable reason why one should regard the expectation of becoming both as the anomalous state of affairs in which an essentially single mind supervenes upon two spatially separated brains. Rather, upon being confronted with the prospect of fission, O should expect to end up as *each* of the branching copies, where O' and O'' harbor distinct minds (and thus experiences), yet both claim to be the direct temporal descendant of O (from psychological continuity). From a classical *identity* point of view this is paradoxical because the situation violates transitivity: however, as has been persuasively argued by Parfit ([14]), what truly seems to matter in the identity stakes is *personal survival* (the continuation of mentality) - and surely, O' and O'' will both purport to have survived the surgical procedure as O. Importantly, personal survival is *not* an equivalence relation, wherefore multiple copies of an agent coherently can claim concurrently that they are survivors of one and the same person[7].

[6] I.e. the identity relation, $\Im$, satisfies *reflexivity* ($\forall x : x\Im x$), *symmetry* ($\forall x\forall y : [x\Im y \leftrightarrow y\Im x]$) and *transitivity* ($\forall x\forall y\forall z : [x\Im y \wedge y\Im z \rightarrow x\Im z]$).

[7] Let me attempt to sharpen this critique some more: in the same way that Dennett and Wallace view macro-objects as fuzzy patterns in the micro-ontology, I see the *self* as an evolutionarily convenient construct brought about by a bundle of abstract categories (memories, thoughts,

Hence, on Parfit's analysis, Saunders' argument fails to demonstrate why we are required to consider branching as subjectively indeterministic: the incoherence problem remains problematic.

3.2 Wallace's Semantics

Let us contemplate an alternative (semantic) solution due to David Wallace ([20]): by considering how to interpret the language of inhabitants of a branching universe, he argues that the SU-problem can be solved. To this end, imagine an intelligent race of beings inhabiting a branching (Everettian) universe, *although they are oblivious of this fact*. As it happens, upon being confronted with what (unbeknownst to them) really are branching events, they are disposed to say "I am uncertain$_L$ what is going to occur", where 'uncertain$_L$' is a word in their language L. Furthermore, they are normally disposed to utter "x will$_L$ happen" iff x happens in *every* branch post-assertation. When asking their philosophers of language (likewise ignorant of the branching) to account for 'uncertain$_L$' and 'will$_L$ happen', the philosophers will based on a false (yet considered[8]) metaphysics say that one should be uncertain$_L$ of something just if there is an objective matter about which to be uncertain$_L$, and something will$_L$ happen just if it happens in the single determinate future.

What do the phrases 'uncertain$_L$ about' and 'will$_L$ happen' really mean? It is clear that if we take the *elite view* and accept the philosophers' semantic analysis, then almost all beings make wildly inaccurate or downright meaningless claims. E.g. on the elite view, the proposition P='the current president will$_L$ not be reelected' is understood as 'in the single determinate future, the current president will lose the election'. But as none of the beings know, there are in fact multiple futures - some in which she might be reelected. The ques-

...). And just as we do not notice the inherent "fuzziness" of macro-patterns (such as molecular diffusion), we fail to recognize that the *self* might not be an altogether well-defined coherent entity, but rather some slightly fluctuating pattern of consciousness. Of course, Descartes famously argued that the only thing of which we can be certain is the *I*: but is it really? It seems to me that the aforementioned certainty really ought to pertain the *existence of sensations* (which jointly "conspire" to create the notion of the self). The upshot is that even in non-fission cases the question of "personal identity between two temporal slices of the same person" is meaningless: precisely because there strictly is no I-substance (just distinct mental patterns which are suitably related to create an overall sense of coherence). Thus, it is completely sensible to talk about personal survival in the face of fuzzy consciousness. (As an aside, it just might be possible to argue from an oscillating physical macro-pattern such as the brain (upon which we hold mentality is supervenient) to an oscillating abstract pattern such as the self. However, there is certainly no a priori reason why this should follow, so the argument requires a lot of work).

[8] Which is of course more than one can say of the non-philosophers, who simply use words conventionally.

tion is whether this makes it true that the president will be reelected (whence the beings' assertion is false) - or merely ill-posed (in which case P is gibberish). Alternatively, we might take the (non-elite) view that *the man on the street* uses 'uncertain$_L$' and 'will$_L$' entirely correctly (and that the philosophers are wrong), but this in turn faces the problem that although it makes the beings' discourse almost entirely accurate, this fact is purely *incidental*. E.g. statement P comes out true on the pedestrian view since 'will$_L$ not' negates that 'the president is reelected in every future branch' (again: the man on the street is no metaphysician; he simply uses terms in a given context because he was told that that was appropriate when he first accumulated an understanding of L[9]).

Following the plausible views of semantics advocated by Davidson, Lewis, Quine et al. there are no further facts about meaning beyond fit to usage and, accordingly, the best interpretation is that which optimizes the truth of the community's utterances. Thus, we would have to conclude that the non-elite view is superior, and *this* has profound implications for the way we use *our* language, assuming our world is in fact described by Everettian quantum theory. Specifically, it would seem that we are completely justified in our characterization of our attitude towards quantum measurements as 'uncertain' (the fact that it is an erroneous metaphysics which initially led to this term is irrelevant). As Wallace stresses, observe that the elite/non-elite view is not a mere linguistic dispute: the point is that our existing uncertainty locutions (and associated conceptual framework) for all we know *already* refer to quantum branching. In this way, we can with full justification employ our existing machinery for corroborating physical theories to the Everett interpretation (wherefore any evidence for the quantum algorithm can be regarded as evidence for the Everett interpretation). And this would seem to be the solution to the SU-problem: we *should* be genuinely uncertain about which outcome of branching occurs, and credences can coherently be assigned in spite of perfect objective knowledge ([20]).

3.3 Quantitative Preliminary

Since each experimentally possibly outcome is assigned a quantum *weight* (here, simply understood as modulus squared of the amplitude), we might conjecture that these quantities are something which *could* fit the function definition of OP from subsection 2.1. Now we could simply take 'weights satisfy the functional definition' as an axiom, in which case the Everett interpretation

[9] An analogy: most people can distinguish objects which are red, although they have little idea of what 'red' and 'color' really mean.

implies the quantum algorithm (and any evidence for the algorithm may be taken as supporting the many worlds interpretation). In itself this would be quite impressive as it would offer us a solution to the measurement problem: moreover, by adopting the cautious functionalist approach espoused above, we can just sit back and hope that some future argument ultimately will account for why weights fit the functional definition. However, recent work by Deutsch ([7]) and Wallace ([19]) indicates that the Everettian's situation is brighter than this: in fact, using principles of decision theory *it appears that the quantitative problem can be solved completely*: in other words, *assuming the correctness of the Everett interpretation*, it can be *shown* that rational agents set probability = weight. The finer details of this argument will be scrutinized in the following section.

4 The Deutsch-Wallace Theorem

4.1 Quantum Games

Before the Deutsch-Wallace (**DW**) theorem is presented, it is necessary to understand (a) the notion of a ”quantum game” and (b) the axioms of decision theory (see subsection 4.2), both of which are integral to the proof thereof.

Quite informally, a quantum game is a bet placed on the outcome of a measurement: evidently, this involves a system being prepared in some particular quantum state, the measurement of some physical observable and a reward which depends on the result of the measurement. More technically,

Definition 3. A quantum **game** (boldface) is an ordered triple $\langle |\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$, where

- $|\psi\rangle$ is a state vector in some Hilbert space $\mathcal{H}_S$.
- $\boldsymbol{X}$ is a self-adjoint operator acting on $\mathcal{H}_S$, which is assumed throughout to have a discrete spectrum (written as $\sigma(\boldsymbol{X})$).
- $\mathcal{P}$ is a function from the spectrum of $\boldsymbol{X}$ into the reals: $\mathcal{P} : \sigma(\boldsymbol{X}) \mapsto \mathbb{R}$.

In itself, this is really a mathematical entity, and of course our interest is the physical processes which are described by that object. To this end, let us define a game (not in boldface) as the physical process which instantiates some **game**, where:

Definition 4. A given process *instantiates* some **game** $\langle|\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$ iff that process consists of:

1. The preparation of some quantum system, S, in the state represented by $|\psi\rangle \in \mathcal{H}_S$, where $\mathcal{H}_S$ is the state space of the system.
2. The measurement[10] on S of the physical observable represented by $\boldsymbol{X}$.
3. The monetary reward of $\mathcal{P}(x)$ [units of currency] in each branch in which result 'x' was recorded.

(In a similar vein, a **compound game** of rank n is defined as a triple $\langle|\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$, where $\mathcal{P}$ is a map from $\sigma(\boldsymbol{X})$ to simple **games** (definition 3) and **compound games** of rank $n-1$. Unsurprisingly, a compound game is any physical process instantiating a **compound game**). The reason why we bother with differentiating between quantum **games** and games is as follows: they do not stand in one-to-one correspondence (a fact exploited in the DW theorem). More precisely, the instantiation map from games to **games** is one-to-many (as is the inverse map from **games** to games) - there are for instance many ways to construct a measuring device! For this reason, let us define an equivalence relation $\sim$ between **games**: $\mathfrak{g} \sim \mathfrak{g}'$ if **games** $\mathfrak{g}$ and $\mathfrak{g}'$ are instantiated by the same game.

Before stating a general *equivalence theorem*, it will be convenient to enunciate some general conventions for **games** which henceforth will be adopted. Specifically, we shall in general assume that the operator $\boldsymbol{X}$ being measured is *degenerate*, viz. a given eigenvalue x_i of $\boldsymbol{X}$ might be associated with multiple mutually orthogonal eigenkets $|\lambda_i^{(j)}\rangle$. Thus, if $\boldsymbol{P}_{\boldsymbol{X}}(x_i)$ is the *projector* onto the eigensubspace of $\boldsymbol{X}$ with eigenvalue x_i:

$$\boldsymbol{X} = \sum_{x_i \in \sigma(\boldsymbol{X})} x_i \boldsymbol{P}_{\boldsymbol{X}}(x_i) \quad \text{where} \quad \boldsymbol{P}_{\boldsymbol{X}}(x_i) = \sum_{j=1}^{d_i} |\lambda_i^{(j)}\rangle\langle\lambda_i^{(j)}| \tag{8}$$

[10] As it was hinted in the introduction, we define the measurement procedure for the generic observable $\boldsymbol{X}$ on $|\psi\rangle$ as follows: let $|M_0\rangle \in \mathcal{H}_M$ be the state representing the pre-measurement apparatus s.t. $|M_0\rangle$ is an element in the decoherence-induced preferred basis. Furthermore, let $\{|\lambda_i\rangle\}$ be the set of eigenstates of $\boldsymbol{X}$ (i.e. $\boldsymbol{X}|\lambda_i\rangle = x_i|\lambda_i\rangle$) and let $\{|M; x_i; \alpha\rangle\}$ be the set of orthogonal readout states of $\mathcal{H}_S \otimes \mathcal{H}_M$, where each vector represents the apparatus displaying x_i as the measurement outcome. Upon expanding $|\psi\rangle$ in terms of the $|\lambda_i\rangle$s, a (generally "disturbing") *measurement* of $\boldsymbol{X}$ is then defined as the dynamical evolution, which takes $|\lambda_i\rangle \otimes |M_0\rangle$ into the state $\sum_\alpha \mu(\lambda_i; \alpha)|M; x_i; \alpha\rangle$, where $\mu(\lambda_i; \alpha) \in \mathbb{C}$ and $\sum_\alpha |\mu(\lambda_i; \alpha)|^2 = 1$. (The α label allows for the possibility that several different output states of the apparatus may correspond to the same measurement outcome).

where d_i is the degeneracy of eigenvalue x_i. Moreover, for a given **game**, $\mathfrak{g} = \langle |\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$, we define the *weight map* $W_{\mathfrak{g}} : \mathbb{R} \mapsto \mathbb{R}$ by

$$W_{\mathfrak{g}}(c) = \sum_{x \in \mathcal{P}^{-1}(c)} \langle \psi | \boldsymbol{P}_{\boldsymbol{X}}(x) | \psi \rangle. \tag{9}$$

This equation requires some elucidation: the summation ranges over that subset of eigenvalues in the spectrum, $\varrho(\boldsymbol{X}) \subseteq \sigma((\boldsymbol{X}))$, such that $\forall x \in \varrho(\boldsymbol{X}) : \mathcal{P}(x) = c$, where c is some fixed real number. Less technically, we sum over all branches in which payoff c is given. E.g. suppose we perform a measurement of spin $\boldsymbol{S}_z$ on the state $| \uparrow_{x^+} \rangle$ and that we are rewarded with \$10 if the outcome is $+\frac{\hbar}{2}$ and nothing if the outcome is $-\frac{\hbar}{2}$. It follows that $W_{\mathfrak{g}}(\$10) = \langle \uparrow_{x^+} \mid \uparrow_{z^+} \rangle \langle \uparrow_{z^+} \mid \uparrow_{x^+} \rangle = |1/\sqrt{2}|^2 = 1/2$ (likewise, $W_{\mathfrak{g}}(\$0) = 1/2$). This highlights the general point that the weight map of c simply is the sum of *weights* of those branches with payoff c. (Here we call the modulus squared of the amplitudes for 'weights' rather than 'probability' in order to avoid illegitimate presuppositions).

We can now state the *equivalence theorem* of **games**.

Theorem 1. Equivalence Theorem

1. *Payoff equivalence* (**PE**):

$$\langle |\psi\rangle, \boldsymbol{X}, \mathcal{P} \circ f \rangle \sim \langle |\psi\rangle, f(\boldsymbol{X}), \mathcal{P} \rangle \tag{10}$$

   where $f : \sigma(\boldsymbol{X}) \mapsto \mathbb{R}$.

2. *Measurement Equivalence* (**ME**):

$$\langle |\psi\rangle, \boldsymbol{X}, \mathcal{P} \rangle \sim \langle \boldsymbol{U} |\psi\rangle, \boldsymbol{X}', \mathcal{P}' \rangle \tag{11}$$

   where (i) $\boldsymbol{U}$ is a unitary transformation, (ii) $\boldsymbol{X}$ and $\boldsymbol{X}'$ satisfy $\boldsymbol{U}\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{U}$ (so their spectra are identical) and (iii) $\mathcal{P}$ and $\mathcal{P}'$ agree on $\sigma(\boldsymbol{X}) \equiv \sigma(\boldsymbol{X}')$. Observe that we here allow $\boldsymbol{U}$ to connect *different* Hilbert spaces; had we restricted ourselves to a fixed Hilbert space, then we obtain the result

$$\langle |\psi\rangle, \boldsymbol{X}, \mathcal{P} \rangle \sim \langle \boldsymbol{U} |\psi\rangle, \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^{\dagger}, \mathcal{P} \rangle \tag{12}$$

3. *General equivalence* (**GE**):

$$\mathfrak{g} \sim \mathfrak{g}' \text{ iff } W_{\mathfrak{g}} = W_{\mathfrak{g}'} \tag{13}$$

In this paper we shall just consider the Deutschian proof of the DW theorem, which makes use of PE and ME, but not GE. Thus, only the former two will be proven (the proof of the latter can be found in [19], p. 423):

Proof. **PE:** Let $|M; x_i\rangle$ be a read-out state of $\mathcal{H}_S \otimes \mathcal{H}_M$ (system + apparatus) i.e. the state which physically displays x_i in some way measurable by an observer. Now the rule associating a particular eigenvalue with a read-out state is purely conventional. Thus, change that convention: regard $|M; x_i\rangle$ as displaying $f(x_i)$ and instead of getting payoff $\mathcal{P} \circ f(x)$ from x, change the payoff to $\mathcal{P}(x)$. These two changes replaces the **game** $\langle|\psi\rangle, \boldsymbol{X}, \mathcal{P} \circ f\rangle$ with $\langle|\psi\rangle, f(\boldsymbol{X}), \mathcal{P}\rangle$, and all we have done is to change our labeling convention (in particular no physical change has been made). Thus, the **games** are equivalent. □

Proof. **ME:** Let us for the sake of simplicity assume that $\boldsymbol{X}$ and $\boldsymbol{X}'$ act on two different Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$. Since $\boldsymbol{U}\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{U}$ it is possible to label the eigenstates of $\boldsymbol{X}'$, viz. $\{|\mu_i\rangle\}_{i=1}^{n'}$, such that for $a \leqslant n$, $\boldsymbol{U}|\lambda_i\rangle = |\mu_i\rangle$ and $\boldsymbol{X}'|\mu_i\rangle = x_i|\mu_i\rangle$. Expanding $|\psi\rangle$ as a superposition of $\boldsymbol{X}$ eigenstates i.e. $\sum_{i=1}^{n} \alpha_i|\lambda_i\rangle$, let us consider the following physical process:

1. Prepare the physical system represented by the state space $\mathcal{H}$ in the state $|\psi\rangle$ and the system represented by $\mathcal{H}'$ in the state $|0'\rangle$, such that the overall quantum state is $|\psi\rangle \otimes |0'\rangle \otimes |M_0\rangle$, where $|M_0\rangle$ is the initial state for the measurement device acting on the $\mathcal{H}'$-system.

2. Let some unitary transformation $\boldsymbol{U}$ act on $\mathcal{H} \otimes \mathcal{H}'$ such as to realize the evolution $|\phi\rangle \otimes |0'\rangle \longrightarrow |0\rangle \otimes \boldsymbol{U}|\phi\rangle$, where $|\phi\rangle$ is an arbitrary $\mathcal{H}$ state and $|0\rangle$ is some fixed $\mathcal{H}$ state.

3. For notational convenience, scrap the system represented by $\mathcal{H}$, wherefore the joint system is now in the state $\boldsymbol{U}|\psi\rangle \otimes |M_0\rangle$.

4. Measure $\boldsymbol{X}'$ using the dynamics $|\mu_i\rangle \otimes |M_0\rangle \longrightarrow |M; x_i\rangle$, where $|M; x_i\rangle$ is the readout state associated with x_i (if $\boldsymbol{X}'$ is degenerate there will be several such mutually orthogonal states).

5. Clearly, the final state is now $\sum_{i=1}^{n} \alpha_i|M;x_i\rangle$ (in the degeneracy case this is insufficient as the sum only contains one ket per eigenvalue - the generalization is straightforward though). In those branches where x_i is recorded, give the payoff $\mathcal{P}'(x_i)$.

One way in which we might construe this process is as follows: steps (1)-(3) essentially prepare the state $\boldsymbol{U}|\psi\rangle \in \mathcal{H}'$, using an auxiliary system represented by $\mathcal{H}$. Step (4) is the case where the operator $\boldsymbol{X}$ is measured on the prepared state, and step (5) provides the payout $\mathcal{P}'$. Hence, the **game** $\langle \boldsymbol{U}|\psi\rangle, \boldsymbol{X}', \mathcal{P}'\rangle$ is instantiated.

An alternative way to view the process is this: upon viewing steps (2)-(4) as a "black box" process, where input and output states are all we know about, then the following transformation is realized:

$$\left(\sum_{i=1}^{n} \alpha_i|\lambda_i\rangle\right) \otimes |0'\rangle \otimes |M_0\rangle \longrightarrow \sum_{i=1}^{n} \alpha_i|M;x_i\rangle \tag{14}$$

which as a matter of definition amounts to a measurement of $\boldsymbol{X}$ on the state $|\psi\rangle$ using a measurement device with initial state $|0'\rangle \otimes |M_0\rangle$. Hence, process (1)-(5) can *also* be understood as the state of affair in which we prepare the state $|\psi\rangle \in \mathcal{H}$ (1), measure the operator $\boldsymbol{X}$ on $|\psi\rangle$ (2)-(4), and provide a payout $\mathcal{P}$ (5). Since there is no physical difference between providing payoff $\mathcal{P}$ and $\mathcal{P}'$, it follows that the process is an instantiation of the **game** $\langle|\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$.

Obviously there is no physical difference between these two descriptions of (1)-(5); therefore $\langle|\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle \sim \langle \boldsymbol{U}|\psi\rangle, \boldsymbol{X}', \mathcal{P}'\rangle$. □

4.2 Decision Theory

Following Deutsch ([7]) let us now introduce some decision theoretic assumptions about rational agents' preferences between games. To this end, let us define a *value function*, $\mathfrak{V} : \{\textbf{games}\} \mapsto \mathbb{R}$, such that if some **game**'s payoff function is constant ($= c$), then the value of that **game** is c (as a matter of notational convenience, let us write the value of $\langle|\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$ as $\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P})$). The governing idea behind introducing this quantity is that a rational agent prefers a **game** $\mathfrak{g}$ to another $\mathfrak{g}'$ if and only if $\mathfrak{V}(\mathfrak{g}) > \mathfrak{V}(\mathfrak{g}')$ (think of $\mathfrak{V}(\mathfrak{g})$ as the "cash value" of $\mathfrak{g}$ to the agent, who will be indifferent between playing $\mathfrak{g}$ and receiving a reward of value $\mathfrak{V}(\mathfrak{g})$).

We now make certain key assumptions about how rational agents must behave, by imposing the following restrictions on $\mathfrak{V}$:

- **Dominance:** If $\forall x : \mathcal{P}(x) \geqslant \mathcal{P}'(x)$ then

  $$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}) \geqslant \mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}'). \tag{15}$$

  I.e. if one game invariably leads to a better reward than another, rationally you must prefer the former.

- **Substitutivity:** If $\mathfrak{g}_{\mathrm{comp}}$ is a compound **game** formed from some **game** $\mathfrak{g}$ by substituting for its payoffs $\{c_i\}|_{i=1}^n$ **games** $\{\mathfrak{g}_i\}|_{i=1}^n$ such that $\mathfrak{V}(\mathfrak{g}_i) = c_i$, then $\mathfrak{V}(\mathfrak{g}_{\mathrm{comp}}) = \mathfrak{V}(\mathfrak{g})$. *In words*: if you are indifferent between receiving a reward of value c and playing some game, it is rational to be indifferent between a *chance* of getting c and the same chance of playing that game.

- **Weak additivity:** If k is any real number, then

  $$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P} + k) = \mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}) + k. \tag{16}$$

  To motivate this, consider a physical process which instantiates the game $\mathfrak{g} = \langle|\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$, after which a reward of value k is delivered with certainty. Physically, this is tantamount to performing a measurement on $|\psi\rangle$ and thence receiving, sequentially, two rewards upon getting result x_i: *first* one of value $\mathcal{P}(x_i)$ and *then* one of value k. This, of course, amounts to the single reward $\mathcal{P}(x_i) + k$ and thus the physical process instantiates $\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P} + k)$. Very well, suppose instead that k is received prior to playing $\mathfrak{g}$: by **substitutivity** it is rational to be indifferent between (a) receiving k then playing $\mathfrak{g}$ and (b) receiving k then receiving $\mathfrak{V}(\mathfrak{g})$. And clearly, the latter amounts to the "lump-sum" payment of $\mathfrak{V}(\mathfrak{g}) + k$ as desired.

- **Zero-sum:** For a given payoff function $\mathcal{P}$, let $-\mathcal{P}$ be defined by $(-\mathcal{P})(x) = -(\mathcal{P}(x))$, then:

  $$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, -\mathcal{P}) = -\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}). \tag{17}$$

  Motivating this, consider the case where two agents (I & II), with equivalent preferences, play a game where a gain to either one of them is balanced by a loss to the other. It seems reasonable to assume that if one actively wants to play (expects to benefit), then the other actively wants *not* to play (expects to lose out). Suppose $\mathfrak{g} = \langle|\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle$ and agent I plays $\mathfrak{g}' = \langle|\psi\rangle, \boldsymbol{X}, \mathcal{P} - \mathfrak{V}(\mathfrak{g})\rangle$, with agent II acting as a banker (i.e. playing $-\mathfrak{g}' = \langle|\psi\rangle, \boldsymbol{X}, \mathfrak{V}(\mathfrak{g}) - \mathcal{P}\rangle$). From **weak additivity** $\mathfrak{V}(\mathfrak{g}') = 0$, so agent I is indifferent to playing $\mathfrak{g}'$. Thus, agent II is indifferent to playing $-\mathfrak{g}'$ and (applying the lemma again) **zero-sum** must hold.

For now, let us simply take these (rather strong![11]) assumptions at face value and state the final assumption we need to make viz. **physicality**[12].:

- **Physicality:** Two **games** instantiated by the same physical process have the same value:

  $$\mathfrak{g} \sim \mathfrak{g}' \rightarrow \mathfrak{V}(\mathfrak{g}) = \mathfrak{V}(\mathfrak{g}'). \tag{18}$$

  ”Obviously” real agents have preferences between games, not **games** (but see subsection 4.4 for subtleties).

.

4.3 The DW Proof

Using the formalism developed in subsections 4.1 and 4.2 we are finally able to state and prove the **Deutsch-Wallace theorem**. Recall the sheer significance of this result: if we accept the proof we will have shown that Everettians, based on pure principles of rationality, should set their credences in the various branches equal to those given by the Born rule. And if we trust the work developed previously in this paper, it turns out to be perfectly coherent that agents harbor this kind of uncertainty (although the many-worlds interpretation is deterministic) - hence, *the problem of probability will be solved*.

[11] As Wallace points out ([19], p. 426), they entail that ”it is rational to bet the mortgage on a 1-in-1.000.000 chance of winning the GNP of Europe”.

[12] As an aside, observe that **weak additivity** and **zero-sum** are special cases of the general principle of **additivity**, viz. $\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P} + \mathcal{P}') = \mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}) + \mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}')$. This principle is in fact *not* assumed by Deutsch, but this is unlikely to be due to any particular non-triviality vis-a-vis the other decision-theoretic axioms. However, the conjunction of **additivity** and **dominance** does allow us to prove a **probability representation theorem** (see Appendix) the content of which is far from trivial. Explicitly, the representation theorem states that if $\mathfrak{V}$ is a value function satisfying the axioms, then $\mathfrak{V}$ is given by the formula $\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}) = \sum_{x\in\sigma(\boldsymbol{X})} Pr_{\psi,\boldsymbol{X}}(x)\mathcal{P}(x)$, where the numbers $\{Pr_{\psi,\boldsymbol{X}}(x) \in \mathbb{R}|0 \leqslant x \leqslant 1\}$ depend on $|\psi\rangle$ and $\boldsymbol{X}$, but not on $\mathcal{P}$. Also $\sum_x Pr_{\psi,\boldsymbol{X}}(x) = 1$. More on this in section 5.

Theorem 2. (DW theorem): if $\mathfrak{V}$ is a value function which satisfies **physicality**, **weak additivity**, **substitutivity**, **dominance**, and **zero-sum**, then $\mathfrak{V}$ is given *uniquely* by the Born rule:

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}) = \sum_{x \in \sigma(\boldsymbol{X})} \langle\psi|\boldsymbol{P_X}(x)|\psi\rangle\mathcal{P}(x) \equiv \sum_{c \in \mathcal{P}[\sigma(\boldsymbol{X})]} cW_{\mathfrak{g}}(c) \tag{19}$$

More prosaically: the value of a quantum game is the sum of the payoffs given in the various Everettian branches, where each branch is weighted by its Born probability (recall that each possible eigenvalue outcome of a measurement corresponds one-to-one with a branch).

Now the proof of this theorem is somewhat intricate: following DW ([19]) we break it down into six steps (lemmas) of increasing order of complexity (such that lemma $n+1$ assumes the validity of lemma n, where $\{n \in \mathbb{N}|1 < n < 5\}$). Once these steps have been established, the DW theorem readily follows.

Lemma 1. Let $|\psi\rangle = \frac{1}{\sqrt{2}}(|\lambda_1\rangle + |\lambda_2\rangle)$, then $\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, id_{\boldsymbol{X}}) \equiv \mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = \frac{1}{2}(x_1 + x_2)$, where the function $id_{\boldsymbol{X}}$ is a restriction of the identity map $id(x) = x$ to $\sigma(\boldsymbol{X})$.

Proof. Consider $\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, id_{\boldsymbol{X}}) + k$. Applying **weak additivity** and then PE we obtain:

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, id_{\boldsymbol{X}}) + k = \mathfrak{V}(|\psi\rangle, \boldsymbol{X}, id_{\boldsymbol{X}} + k) = \mathfrak{V}(|\psi\rangle, \boldsymbol{X} + k, id_{\boldsymbol{X}}) \tag{20}$$

Correspondingly, applying PE and **zero-sum** to $\mathfrak{V}(|\psi\rangle, -\boldsymbol{X}, id_{\boldsymbol{X}})$ we get:

$$\mathfrak{V}(|\psi\rangle, -\boldsymbol{X}, id_{\boldsymbol{X}}) = \mathfrak{V}(|\psi\rangle, \boldsymbol{X}, -id_{\boldsymbol{X}}) = -\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, id_{\boldsymbol{X}}) \tag{21}$$

and (20) and (21) jointly imply that:

$$\mathfrak{V}(|\psi\rangle, -\boldsymbol{X} + k) = -\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) + k \tag{22}$$

Suppose we let f be the function of reflection about the point $\frac{1}{2}(x_1 + x_2)$ i.e. $f(x) = -x + x_1 + x_2$. Then provided that $\boldsymbol{X}$ is *non-degenerate* and $\sigma(\boldsymbol{X})$ is invariant under the action of f, the operator $\boldsymbol{U}_f$ (defined by $\boldsymbol{U}_f \boldsymbol{X} \boldsymbol{U}_f^\dagger = f(\boldsymbol{X})$) will be well-defined and will leave $|\psi\rangle$ invariant. From ME we immediately obtain:

$$\mathfrak{V}(|\psi\rangle, -\boldsymbol{X} + x_1 + x_2) = \mathfrak{V}(|\psi\rangle, \boldsymbol{X}) \tag{23}$$

where the LHS from (22) can be written as $-\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) + x_1 + x_2$ and thus (23) can be rewritten as:

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = \tfrac{1}{2}(x_1 + x_2) \tag{24}$$

as desired[13]. □

Lemma 2. If $N = 2^n$ where $n \in \mathbb{Z}^+$, and if $|\psi\rangle = \frac{1}{\sqrt{N}}(|\lambda_1\rangle + ... + |\lambda_N\rangle)$, then

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = \tfrac{1}{N}(x_1 + ... + x_N) \tag{25}$$

Proof. The proof of this is recursive on n; let us start by defining the following quantities:

- $|\psi\rangle = \frac{1}{2}\sum_{i=1}^{4}|\lambda_i\rangle$
- $|\alpha\rangle = \frac{1}{\sqrt{2}}\sum_{i=1}^{2}|\lambda_i\rangle$ and $|\beta\rangle = \frac{1}{\sqrt{2}}\sum_{i=3}^{4}|\lambda_i\rangle$
- $y_\alpha = \frac{1}{2}(x_1 + x_2)$ and $y_\beta = \frac{1}{2}(x_3 + x_4)$
- $\boldsymbol{Y} = y_\alpha|\alpha\rangle\langle\alpha| + y_\beta|\beta\rangle\langle\beta|$

From lemma 1, the value of the game $\mathfrak{g} = \langle|\psi\rangle, \boldsymbol{Y}\rangle$ is $\frac{1}{2}(y_\alpha + y_\beta) = \frac{1}{4}(x_1 + x_2 + x_3 + x_4)$. Consider the y_α branch, in which a reward of $\frac{1}{2}(x_1 + x_2)$ is given. By **substitutivity** a rational agent will be indifferent between receiving that reward and playing the game $\mathfrak{g}_\alpha = \langle|\psi\rangle, \boldsymbol{X}\rangle$ (as it is of corresponding value). Analogous considerations apply to y_β; hence, the value to an agent measuring $\boldsymbol{Y}$ on $|\psi\rangle$ and then playing $\mathfrak{g}_\alpha$ or $\mathfrak{g}_\beta$ according to the result of the measurement is $\frac{1}{4}(x_1 + x_2 + x_3 + x_4)$. But the physical process instantiating this sequence of **games** is simply

$$\left(\tfrac{1}{2}\sum_{i=1}^{4}|\lambda_i\rangle\right) \otimes |M_0\rangle \longrightarrow \tfrac{1}{2}\sum_{i=1}^{4}|M; x_i\rangle \tag{26}$$

which also happens to be an instantiation of the **game** $\langle|\psi\rangle, \boldsymbol{X}\rangle$. So the result holds for $N = 4$. But for $N \in \{8, 16, 32, ...\}$ we simply follow a prescription analogous to the one above (the generalization is straightforward). Hence, the lemma follows. □

[13] The non-degeneracy/non-invariance assumption is not crucial: if either does not obtain let $\mathcal{Q}$ be the span of $\{|\lambda_1\rangle, |\lambda_2\rangle\}$ and let $\mathfrak{V} : \mathcal{Q} \mapsto \mathcal{H}$ be the embedding map, then ME gives us: $\langle|\psi\rangle, \boldsymbol{X}|_{\mathcal{Q}}\rangle \sim \langle|\psi\rangle, \boldsymbol{X}\rangle$ which allows us to generalize.

Lemma 3. Let $N = 2^n$ as before, and a_1, $a_2 \in \mathbb{Z}^+$ such that $a_1 + a_2 = N$. Let $|\psi\rangle = \frac{1}{\sqrt{N}}(\sqrt{a_1}|\lambda_1\rangle + \sqrt{a_2}|\lambda_2\rangle)$ then

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = \tfrac{1}{N}(a_1 x_1 + a_2 x_2) \tag{27}$$

Proof. From ME we may without loss of generality assume that $\mathcal{H}$ is spanned by $|\lambda_1\rangle$, $|\lambda_2\rangle$. Suppose $\mathcal{H}'$ is an N-dimensional Hilbert space which is spanned by $|\mu_1\rangle, ..., |\mu_N\rangle$, then let us define

- $\boldsymbol{Y} = \sum_{i=1}^{N} i|\mu_i\rangle\langle\mu_i|$
- $f(i) = x_1$ for i between 1 and a_1, otherwise $f(i) = x_2$.
- $\boldsymbol{V} : \mathcal{H} \mapsto \mathcal{H}'$ by

$$\boldsymbol{V}|\lambda_1\rangle = \tfrac{1}{\sqrt{a_1}}\sum_{i=1}^{a_1}|\mu_i\rangle \;\text{ and }\; \boldsymbol{V}|\lambda_2\rangle = \tfrac{1}{\sqrt{a_2}}\sum_{i=a_1+1}^{N}|\mu_i\rangle \tag{28}$$

Since $f(\boldsymbol{Y})\boldsymbol{V} = \boldsymbol{V}\boldsymbol{X}$ we have from ME and PE:

$$\langle|\psi\rangle, \boldsymbol{X}\rangle \sim \langle\boldsymbol{V}|\psi\rangle, f(\boldsymbol{Y}), id_{f(\boldsymbol{Y})}\rangle \sim \langle\boldsymbol{V}|\psi\rangle, \boldsymbol{Y}, f \circ id_{\boldsymbol{Y}}\rangle \tag{29}$$

But $\boldsymbol{V}|\psi\rangle = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}|\mu_i\rangle$ so the result follows from lemma 2:

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = \tfrac{1}{N}\left(\sum_{i=1}^{a_1} f(i) + \sum_{i=a_1+1}^{a_2} f(i)\right) = \tfrac{1}{N}(a_1 x_1 + a_2 x_2) \tag{30}$$

where the second bullet point has been used for the last equality. □

Lemma 4. Let $\{a \in \mathbb{R} | 0 < a < 1\}$ and let $|\psi\rangle = \sqrt{a}|\lambda_1\rangle + \sqrt{1-a}|\lambda_2\rangle$. Then

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = a x_1 + (1-a) x_2 \tag{31}$$

Proof. Without loss of generality, suppose $x_1 \leqslant x_2$. Let us make the following definitions:

- $\mathfrak{g} = \langle|\psi\rangle\rangle$, where $\boldsymbol{X}$ is implicit.
- $\{a_n\}$ is a *decreasing* sequence of numbers of the form $a_n = A_n/2^n$ where $A_n \in \mathbb{Z}^+$ s.t. $a_n \to a$ as $n \to \infty$[14].

[14] This is always possible as numbers of this form are *dense* in $\mathbb{R}^+$.

- $|\psi_n\rangle = \sqrt{a_n}|\lambda_1\rangle + \sqrt{1-a_n}|\lambda_2\rangle$.
- $|\phi_n\rangle = \frac{1}{\sqrt{a_n}}(\sqrt{a}|\lambda_1\rangle + \sqrt{a_n - a}|\lambda_2\rangle)$
- $\mathfrak{g}_n = \langle|\psi_n\rangle\rangle$
- $\mathfrak{g}'_n = \langle|\phi_n\rangle\rangle$

From lemma 3 it follows that $\mathfrak{V}(\mathfrak{g}_n) = a_n x_1 + (1-a_n)x_2$. Although we do not know the value of $\mathfrak{V}(\mathfrak{g}'_n)$, it follows from **dominance** that it is at least x_1. So from **substitutivity** the value to a rational agent of measuring $|\psi_n\rangle$ and then receiving x_2 [units of currency] if the result is x_2 and playing $\mathfrak{g}'_n$ if the result is x_1, is at least $\mathfrak{V}(\mathfrak{g}_n)$. But if we think about it, this sequence of games is just an instantiation of $\mathfrak{g}$. To see this, notice that the end state is one in which a reward of x_1 [units of currency] is given with amplitude

$$\sqrt{a_n \cdot \tfrac{a}{a_n}} = \sqrt{a} \tag{32}$$

whilst a reward of x_2 [units of currency] is given with amplitude

$$\sqrt{a_n \cdot \tfrac{a_n - a}{a_n} + (1-a_n)} = \sqrt{1-a} \tag{33}$$

It follows that $\forall n : \mathfrak{V}(\mathfrak{g}) \geqslant \mathfrak{V}(\mathfrak{g}_n)$ and thence $\mathfrak{V}(\mathfrak{g}) \geqslant a x_1 + (1-a)x_2$.

Upon considering a similar argument with an *increasing* sequence, we can likewise establish that $\mathfrak{V}(\mathfrak{g}) \leqslant a x_1 + (1-a)x_2$ wherefore lemma 4 follows. □

Lemma 5. Let $\alpha_1, \alpha_2 \in \mathbb{C}$ such that $|\alpha_1|^2 + |\alpha_2|^2 = 1$, and let $|\psi\rangle = \alpha_1|\lambda_1\rangle + \alpha_2|\lambda_2\rangle$, then

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = |\alpha_1|^2 x_1 + |\alpha_2|^2 x_2 \tag{34}$$

Proof. Let us define $\boldsymbol{U} = \sum_i \exp(i\theta_i)|\lambda_i\rangle\langle\lambda_i|$, then we see that $\boldsymbol{U}$ leaves $\boldsymbol{X}$ invariant, and so ME gives us $\langle \boldsymbol{U}|\psi\rangle, \boldsymbol{X}\rangle \sim \langle|\psi\rangle, \boldsymbol{X}\rangle$. But the eigenstate $\boldsymbol{U}|\psi\rangle$ has only positive real coefficients, and so the value is given by lemma 4. □

Lemma 6. Let $|\psi\rangle = \sum_i \alpha_i |\lambda_i\rangle$ then

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}) = \sum_i |\alpha_i|^2 x_i \tag{35}$$

Proof. In exactly the same way that lemma 2 generalized lemma 1, lemma 6 generalizes lemma 5 and the proof runs analogously. I.e. using **substitutivity** any n-term measurement can be construed as successive 2-term measurements. □

Now lemma 6 if of course just the DW theorem subject to the special condition that the payoff function is the identity ($\mathcal{P} = id_{\boldsymbol{X}}$). The generalization to arbitrary $\mathcal{P}$ is utterly trivial:

Proof. Because of PE we have

$$\langle |\psi\rangle, \boldsymbol{X}, \mathcal{P}\rangle \sim \langle |\psi\rangle, \mathcal{P}(\boldsymbol{X}), id_{\boldsymbol{X}}\rangle \tag{36}$$

so we were completely justified in using $id_{\boldsymbol{X}}$ as the default payoff function throughout these lemmas. □

This concludes the proof of the Deutsch Wallace theorem.

4.4 Measurement Neutrality

In the preceding subsections we were fairly meticulous in our explication of the assumptions required in the DW derivation of the Born rule. *Firstly*, we assumed the correctness of the Everett interpretation. (Whether this is a plausible move ultimately boils down to one's metaphysical preferences, but certainly Many Worlds is the one *realist* approach to the quantum algorithm, which doesn't "regress" into hidden variables (as the Pilot Wave theory), objectively stochastic dynamics (as Dynamical Collapse theories) or ontological significant observers (Many-Minds / variants of von Neumann)). *Secondly*, we assumed the validity of regarding Everettian branching as subjectively uncertain (the acceptability of this move is closely connected to whether the reader accepts Wallace's argument from semantics). In fact, our ability to import classical decision theory into quantum mechanics rests entirely on this assumption. *Thirdly*, we assumed a fairly strong set of decision theoretic postulates: without decision theory we have no license to convert uncertainty into probability (and none of the constraints imposed upon those probabilities which ultimately allow the DW theorem to be proven). The *fourth* and final assumption we made has hitherto remained hidden in the shadows by our notation: it is known as **measurement neutrality** and it is what ties the value function $\mathfrak{V}$ together with real decision making. More precisely, it is the claim that "a rational agent is indifferent between two physical games whenever they instantiate the same **game**" (cf. the assumption of *physicality*) - i.e. as long as a process matches the definition of a measurement $\boldsymbol{X}$ on $|\psi\rangle$, the finer details of how that measurement is executed is irrelevant for decision-making purposes.

Prima facie, this certainly comes across as a reasonable assumption. Upon being confronted with two different measurement devices M_1 and M_2 for some observable $\boldsymbol{X}$, one certainly tends to engage in the kind of counterfactual reasoning[15] that "whichever eigenvalue x_i is obtained using device M_i, would equally well have been the case if M_j had been employed instead (where $i \neq j$ and $i, j = 1, 2$)". However, the postulate *does* encode some non-trivial implications: for example, it is measurement neutrality which is responsible for the fact that the DW theorem does not propagate ramifications into the Pilot Wave theory. For recall that this interpretation comes with a dual ontology viz. the universal state vector and corpuscles (the positions of which are the *hidden variables*). Accordingly, it is possible for two physical processes to agree as to the measurement carried out, the payoff given, and the Hilbert space state, but to disagree on the hidden variables (a rational agent might thus prefer one to the other).

Furthermore, even in the context of the Everett interpretation, measurement neutrality is incompatible with the strategy wherein all branches are regarded as equiprobable. To see this, suppose that we measure $\boldsymbol{S}_z$ on a spin-half particle and gain some monetary reward if the result is "spin-up" but lose money otherwise. Now measurement device M_1 (improbably) results in one branch for the spin-up result and one branch for the spin-down result, whilst device M_2 incorporates a quantum random-number generator which is triggered by a spin-up result, so that there are, say, 10^{12} spin-up branches and only one spin-down branch. Adopting the equiprobability strategy, we should be as likely to win as to lose if we use device M_1, but virtually certain to win if we use device M_2 - however, measurement neutrality instructs us that each is as good as the other (cementing the point that the principle has non-trivial implications).

5 The Fission Program

5.1 Half-baked SU

Of those aspects which go into Wallace's solution to the bipartite probability problem, 'subjective uncertainty' seems to be the most contentious one (see e.g. [1] and [15]). For this reason, Hilary Greaves ([9], [10]) has worked extensively on a solution known as the *Fission Program*, which eradicates the concept of SU altogether. Indeed, such a program must be construed as the nihility of objective chances: an agent who knows quantum mechanics (and the governing state vector) is not in any way uncertain about the outcomes of measurement. Rather, it will be the case that the agent knows that he has a plenitude of successors, wherefore he, upon being faced with Everettian branching, must

[15] Maybe at a mere subconscious level.

consider the interests of these post-branching successors by taking the course of action which best serves them.

More precisely, the fission program should to be understood as offering a *reinterpretation* of the decision-theoretic axioms such that they no longer apply to an agent's ignorance of his single future (which was why we needed SU in the first place), but to the agent's predilection vis-a-vis his branching successors. To exemplify this, consider the aforementioned principle of **dominance**, which, to repeat, states that a rational agent must prefer A to B insofar as A rewards *him* better than B (irrespective of how the future turns out). On the fission approach this translates as follows: a rational agent should regard A as preferable B if each of his future successors is rewarded more richly under A than under B. Applying analogous reinterpretations uniformly to our decision theory, we acquire a representation theorem (see footnote 12) which tells us that agents choose that action which maximizes *expected utility* (EU), where the weights in the EU formula are not credences in unknown outcomes, but a measure pertaining to how much that agent *cares* about each of his future branching descendants. We shall henceforth refer to this as the *caring measure*.

How does this caring measure relate to quantum mechanical weight of the respective branches? Fissionists propose the following rationality principle - designated the *quantum caring principle* (**QCP**):

Definition 5. (QCP): Rational agents are compelled to allocate caring measures to branches in proportion to their quantum mechanical weight, when they know the latter. Thus, if ξ is a given proposition, T is the Everett$_{\text{fission}}$ interpretation, and X is the proposition that "the weight, at the time in question, of all branches on which $|\xi| = \top$ is x (relative to the agent)", then QCP requires that

$$Cr(\xi | T \wedge X) = x \tag{37}$$

It follows that if QCP is true then rational agents (conscious of the fact that they are imbedded in an Everettian universe) will act just as they would have acted in an indeterministic universe where the conventional quantum algorithm was true: despite the non-existence of objective probabilities. But if we think about it, our justification for QCP is no worse off than our justification for 'credence = weight'. This follows from the fact that the proof in subsection 4.2 extends *mutatis mutandis* to the fission program, with the decision-theoretic axioms properly reinterpreted (so as to entail that 'caring measure = weight'). And even if we are not prepared to accept this proof, QCP does not obviously come across as a more contentious rationality principle than PP. In both cases it appears that we are prepared to continue using the principle, even if we do not know how to derive it.

5.2 The Fatness Objection

There is an interesting critique of QCP and the proof thereof, due to David Albert ([1]). Suppose that I self-consciouly decide that the degree to which it is reasonable for me to care about what transpires in a given one of my future branches, ought to be proportional to how corpulent I am in that branch (employing the rationale that those branches in which there is more of me deserve to attract more attention). The question is, whether it is any more irrational or incoherent to adopt one's *fatness* as a caring measure, γ, than it is to adopt the modulus squared of the quantum amplitude as a caring measure[16]?

One reason to be suspicious of γ might be that the coherence of a caring measure which depends *exclusively* on obesity, is going to depend on there being some perfectly definite matter of fact about exactly how many branches there are, which is in fact highly unlikely (Greaves[17]). However, this objection is easily avoided by modifying the example: suppose that I adopt the slightly more sophisticated caring measure, Γ, where I care about what transpires on a given branch in proportion to the product of my fatness in that branch and the associated branch weight.

A second worry is that adopting caring measure Γ is somehow *inconsistent* with or *irrational* in light of the claim that I am as a matter of fact entirely indifferent as to whether *I* am fat or thin. But there is nothing incoherent about me having no preference between two different *non-branching* deterministic future evolutions between which my level of obesity is markedly different, and at the same time be eager to arrange that things are to my liking on the branch in which I am fatter when faced with a genuinely branching event. For upon being faced with genuine branching we supposed rationale-wise that on those branches where there is more of me, there is more to be concerned about, whilst no such considerations can apply to non-branching cases, since the entirety of me, fat or thin, will be on the single branch to come. And there is nothing paradoxical about the fact that while I care a great deal about the relative level of obesity of my branching descendants, those same fatness values are going to be of no concern whatsoever to the descendants themselves. For surely it is perfectly consistent for me to harbor particular interests at $t = t_0$

[16] True, the latter is justified via the DW proof, but here we assume that the axioms underpinning this derivation stand in need of justification.

[17] See the paper by Albert [1], p. 11; also take a look at Greaves [9], pp. 11-14, for an interesting discussion of allowed caring measures.

with respect to my branching descendants at $t = t_1$, which do *not* correspond to the interests of a given one of those t_1-descendants vis-a-vis his circumstances at that time.

As Wallace points out ([1], p. 12), adopting a fatness caring measure might, admittedly, in practice prove onerous: it might for example involve me trying to anticipate - even control - how much food I ingest on some particular future branch. But surely, the prevalence of such difficulties have no direct bearing on the question whether it is *reasonable* for me to care about the matter. Analogously, although current technology renders a quadrupling of my (Spearman) g-factor impossible, that does imply that I find such an intelligence-boost any less desirable. Nor does the fact that we are computationally incapable of predicting how even relatively mundane actions might propagate chaotically through time and lead to the end of millions of innocent lives, mean that we simply don't care whether people die as a result of our actions. Obviously, any altruistic being does whatever is in his or her power to secure the continuation of his species.

Thus, we certainly have some ground for doubting whether it is altogether irrational for us to adopt a caring measure which does not comply with the branch weights.

5.3 The Wrong Question

Another relevant objection to the fission program is that it, as presented above, provides an answer to the wrong question (cf. [1], [20]): more precisely, we are told that 'supposing that we believed that the Everett interpretation was true, what rationality principles should we conform to in deciding how to live our lives?'. And whilst this surely is an interesting enquiry if we were diehard Everettians, our situation is rather that we want to know whether we should believe the Many Worlds interpretation in the first place: id est *is the Everett picture explanatory of our current epistemic situation*? And certainly, it is far from obvious how to answer this question if we are barred from our customary "probabilistic vernacular", as the fissionists would have us believe.

Faced with this objection, one might rightly question what would make us come to accept the fission program in the first place? A reasonable conjecture would be, that the Everett$_{\text{fission}}$ interpretation offers an explanation of observed phenomena just as good as the quantum algorithm (while at the same time solving the measurement problem). Observed phenomena which, recall, essentially are a vast array of experimental outcomes the frequencies of which harmonize very well with the probabilities defined by quantum mechanics. Nevertheless, the fact that the fission program predicts that there are branches where this indeed is true (and concordantly ascribes high weights to those branches), offers us no reason why it is rational to assume that we are in such a branch. What the fission program really provides is a prudential reason to care about successors in proportion to their branch weight, but as Wallace observes (Ibid, p. 667) that does not seem to be of epistemic import.

This point is perhaps better understood from a Bayesian perspective of theory confirmation. Recall that our credence in a given hypothesis traditionally (and plausibly!) is understood as being *updated* qua a conditionalizing procedure. I.e. if $Cr_A(B)$ is our credence in some proposition B subsequent to learning A, then $Cr_A(B) = Cr(B|A) \equiv Cr(A \wedge B)/Cr(A)$. Thus, using Bayes' theorem, if T is some physical theory, and ξ is some evidence, then our updated credence should be

$$Cr_\xi(T) = \frac{Cr_T(\xi)Cr(T)}{Cr(\xi)} \tag{38}$$

Thus, if we treat the quantity $Cr(\xi)$ as *low* (i.e. event ξ is unlikely *a priori*), but $Cr_T(\xi)$ is *high* (i.e. the theory ascribes high OP to the event, since the Principal Principle sets $c(\xi) = Cr_T(\xi)$), then equation (38) instructs us that our credence in T should rise upon observing ξ. The problem with the fission program is that outcomes simply *cannot* be regarded as being of high or low objective probability: since all outcomes occur it follows that $Cr_T(\xi) = 1$ *irrespective* of the weight of ξ. But if quantum weights do not appear in the Bayesian updating rule, then - if true - it cannot be the case that our observation of heigh weight events provides any evidental support for quantum mechanics whatsoever.

Anticipating this kind of objection, Greaves ([11]) has argued that on the *assumption* that we live in an Everettian branching universe, then we can construct an analogue for the Bayesian updating rule and prove its validity. Alas, this does not seem to remedy the situation in which we are concerned about evidence for the Everett interpretation, unless we make further assumptions. To see this, suppose her argument succeeds, and let T be the hypothesis that Many Worlds is true, and X_i be the hypothesis that the weight of branches in which evidence ξ occurs is x_i. Then Greave's transformed updating rule implies that

$$Cr_{\xi,T}(X_i) = \frac{x_i Cr_T(X_i)}{\sum_j x_j Cr_T(X_j)} \tag{39}$$

and although this formula surely allows us to update various credences all of which are conditional on T (i.e. the Everett interpretation), it does not allow us to make statements about rational credence in the Everett interpretation itself. In other words, the fission program is so perversely construed that it does not allow us to assess any evidence for the Everett proposal, because its very epistemic framework presumes the validity of the theory.

6 Conclusion

In this paper we have seen that a quantum dynamics consisting solely of the deterministic process II and a fundamental ontology consisting of nothing but the universal state vector will allow us to resolve the corrosive measurement problem faced by the conventional quantum algorithm. However, this Everettian approach entails the radical result that all terms in entangled quantum superpositions must be treated as branching worlds in a multiverse: worlds which we argued for all practical purposes are selected unambiguously in the process of decoherence. Nonetheless, the Everett interpretation faces a dual problem concerning the status of quantum probability viz. the worry that any talk of stochasticisity is downright incoherent (the subjective uncertainty problem), and the quantitative worry that the traditional Born probabilities are unjustifiable.

I argued that objective probability enters our scientific theories only through the principal principle and that the best approach to the phenomenon is cautious functionalism wherein we surmise that something eventually can be proved to satisfy PP. Whether the SU problem can in fact be solved remains contentious: we saw that Saunders' argument was insufficient, whilst Wallace's semantics might prove more promising. Insofar as it can be solved, the branch weights are candidates for OP in the sense that they satisfy the functional definition offered by PP (and thus the Everett interpretation might be no worse off than any other physical theory involving probability). However, we also saw that Deutsch and Wallace have offered us an ingenious decision-theoretic proof that ”probability = weight” - a proof which rests upon some fairly strong axioms of what constitutes rational behavior. Provided that we find reasons to doubt these postulates (such as the fatness objection) the DW theorem falls to the ground. Finally we saw that the SU-rejecting caring-measure alternative known as the fission program fails to provide an epistemically acceptable account of how we come to accept the Everett interpretation.

I conclude that the theory of Many Worlds has come a long way since it initially emerged with characteristic nebulosity in a paper by Hugh Everett III. While I maintain that the preferred basis problem has been completely solved, I still think that there are technicalities surrounding the solution(s) to the probability problem, which are worth looking at. More precisely, subjective uncertainty and the appeal to decision theory call for greater philosophical scrutiny. However, I do see the work by Deutsch, Wallace and Greaves as a profound step in the right direction, in the sense that even if the probability problem is insurmountable, now we have a much deeper understanding of what exactly goes wrong (which in itself is great philosophical insight).

In the case that the Everett interpretation ultimately proves untenable, I would not hesitate to recommend de Broglie Bohm's "Pilot Wave" theory as the primal alternative to understanding quantum mechanics. Being dynamically isomorphic to Many Worlds, the Pilot Wave solves the measurement problem. Also, by introducing corpuscles into its fundamental ontology, there exits an unambiguous preferred basis (since the actual world is taken to supervene upon these entities). Finally, the appearance of probability in the quantum algorithm is perhaps less mysterious: although the universal wave function encodes the blueprint for all physically possible worlds in a measurement scenario, only one of these is actually realized viz. the one into which the corpuscles end up. For a detailed discussion of this see [13].

Acknowledgements: I thank my former mentors, Wayne Myrvold and David Wallace, for teaching me to reason about foundational quantum mechanics. Special thanks to Jeremy Butterfield who made detailed and helpful comments on a draft of this paper.

7 Appendix

For the sake of completeness, let us prove (see [19], p. 438) the probability representation theorem, which was mentioned in sections 4 and 5 of this paper. Recall:

Theorem 3. Representation Theorem: If $\mathfrak{V}$ is a value function satisfying **additivity** and **dominance**, then $\mathfrak{V}$ is given by:

$$\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P}) = \sum_{x \in \sigma(\boldsymbol{X})} Pr_{\psi,\boldsymbol{X}}(x)\mathcal{P}(x) \tag{40}$$

where the numbers $\{Pr_{\psi,\boldsymbol{X}}(x) \in \mathbb{R} | 0 \leqslant x \leqslant 1\}$ depend on $|\psi\rangle$ and $\boldsymbol{X}$, but not on $\mathcal{P}$; and where

$$\sum_{x} Pr_{\psi,\boldsymbol{X}}(x) = 1. \tag{41}$$

To show this we need the lemma that $\mathfrak{V}$ is linear:

Lemma 7. Linearity: If $\mathfrak{V}$ satisfies **additivity** and **dominance**, then for any sets of real numbers $\{a_i\}|_{i=1}^{N}$ and payoffs $\{\mathcal{P}_i\}|_{i=1}^{N}$,

$$\mathfrak{V}\left(|\psi\rangle, \boldsymbol{X}, \sum_{i=1}^{N} a_i \mathcal{P}_i\right) = \sum_{i=1}^{N} a_i \mathfrak{V}\left(|\psi\rangle, \boldsymbol{X}, \mathcal{P}_i\right) \tag{42}$$

Proof. For notational convenience let us write $\mathfrak{V}(\mathcal{P})$ instead of $\mathfrak{V}(|\psi\rangle, \boldsymbol{X}, \mathcal{P})$. Let $a \in \mathbb{R}^+$ and let $\{k_n\}$ ans $\{m_n\}$ be sequences of integers s.t. $\{k_m/m_n\}$ is an *increasing* sequence which tends towards a. Using the principles of **dominance** and **additivity** we have $\forall n : m_n\mathfrak{V}(a\mathcal{P}) \geqslant k_n\mathfrak{V}(\mathcal{P})$ and thence $\mathfrak{V}(a\mathcal{P}) \geqslant a\mathfrak{V}(\mathcal{P})$. If we repeat this step with a *decreasing* sequence we find that the only common overlap is $\mathfrak{V}(a\mathcal{P}) = a\mathfrak{V}(\mathcal{P})$ for any $a \geqslant 0$. To extend this proof to negative a we simply use **zero-sum** and the full result follows from **additivity**. □

Thus, the probability representation theorem can be demonstrated:

Proof. For any $x \in \sigma(\boldsymbol{X})$ let us define $\delta_x(y)$ thus:

$$\delta_x(y) = \begin{cases} 1, & \text{if } y = x \\ 0, & \text{otherwise.} \end{cases}$$

Any payoff function $\mathcal{P}$ for $\sigma(\boldsymbol{X})$ can be expressed uniquely as

$$\mathcal{P} = \sum_{x \in \sigma(\boldsymbol{X})} \mathcal{P}(x)\delta_x$$

and applying the lemma of **linearity** we thus get

$$\mathfrak{V}(\mathcal{P}) = \sum_{x \in \sigma(\boldsymbol{X})} \mathcal{P}(x)\mathfrak{V}(\delta_x).$$

If we set $Pr(x) = \mathfrak{V}(\delta_x)$ then we establish equation (40), and putting $\forall x : \mathcal{P}(x) = 1$ gives equation (41) as a special case. □

References

[1] Albert, D. *Probability in the Everett Picture*, Everett at 50, Unpublished.

[2] Albert, D. *Quantum Mechanics and Experience*, Havard University Press, 1992.

[3] Barrett, J. *The Quantum Mechanics of Minds and Worlds*, Oxford University Press, 2003.

[4] Bell, J. *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, 1987.

[5] Brown, H. and Wallace, D. *Solving the measurement problem: de Broglie-Bohm loses out to Everett*, Foundations of Physics 35 (2005), pp. 517-540.

[6] Dennett, D. *Real Patterns*, Journal of Philosophy, 87, pp. 27-51.

[7] Deutsch, D. *Quantum Theory of Probability and Decisions*, Proceedings of the Royal Society of London, A455, 3129-3137.

[8] Everett III, H. *"Relative State" Formulation of Quantum Mechanics*, Reviews of Modern Physics, Vol. 29, No. 3, July, 1957.

[9] Greaves, H. *Probability in the Everett interpretation*, Philosophy Compass 2(1), Jan 2007, 109128.

[10] Greaves, H. *On the Everett epistemic Problem*, Studies in History and Philosophy of Modern Physics 38(1), March 2007, pp.120-152.

[11] Greaves, H. *Understanding Deutsch's probability in a deterministic multiverse*, Studies in History and Philosophy of Science, 35, pp. 423-456.

[12] Hughes, R. *The structure and Interpretation o Quantum Mechanics*, Harvard University Press, 2003.

[13] Nielsen, S. *On the quantum measurement problem and the de Broglie-Bohm interpretation*, Unpublished (available on request).

[14] Parfit, D. *Personal Identity*, Philosophical Review Vol. 80: 3-27, 1971.

[15] Price, H. *Decisions, Decisions, Decisions: Can Savage Salvage Everettian Probability?*, Available at the PhilSci Archive, ID 3886.

[16] Saunders, S. *Time, Quantum Mechanics, and Probability*, Synthese, 114, 373-404.

[17] von Neumann, J. *Mathematical Foundations of Quantum Mechanics*, Princeton University Press.

[18] Vaidman, L. *Many-Worlds Interpretation of Quantum Mechanics*, Stanford Encyclopedia of Philosophy, 2002.

[19] Wallace, D. *Everettian Rationality: defending Deutsch's approach to probability in the Everett Interpretation*, Studies in History and Philosophy of Modern Physics 34 (2003), 415-439.

[20] Wallace, D. *Epistemology Quantized*, Brit. J. Phil. Sci. 57 (2006), 655-689.

[21] Wallace, D. *Decoherence and Ontology: or, How I Learned to Stop Worrying and Love FAPP*, Forthcoming in Saunders, Barrett, Kent, and Wallace (ed.), *Many Worlds? Everett, Quantum Theory, and Reality* (OUP, forthcoming).

[22] Zurek, W. *Decoherence and the transition from quantum to classical*, Physics Today 43, 1991, 3644.