Three kinds of branching universe (UNUBLISHED paper kept for archival reasons - don't assume I still endorse anything written here)

David Wallace[*]

July 2005

Abstract

In the light of recent work suggesting that the quantum probability rule can be derived in the Everett interpretation via decision theory, I consider what physical features of quantum mechanics make this possible. I analyse the status of the probabiliy rule in three different models of branching universes, each somewhat more complicated than the last, and conclude that only in the last model — in which the branching structure, as in quantum mechanics, emerges in a somewhat imprecise way from the underlying physical reality — is it possible to derive a probability rule, or indeed to behave in any rational way at all.

1 Introduction

The 'probability problem' of the Everett interpretation of quantum mechanics has long been its bugbear, attracting more hostile attention even than the 'preferred basis problem' (now generally, though not universally, agreed to have been solved via considerations of decoherence). If the universe splits after a measurement, with every possible measurement outcome realised in some branch, then how can it even make sense to talk about the probabilities of each outcome? And if it does make sense, shouldn't the probability of each branch be equal — after all, even if a two-outcome experiment leads to branches of unequal weights, there are still two branches, one for each result.

The problem has recently been revitalised by a seminal proposal from David Deutsch (Deutsch 1999). Deutsch considers the probability problem from the perspective of decision theory: he asks "given that I am about to undergo branching, what preferences should I have between different courses of action?"

---

[*]Magdalen College, Oxford OX1 4AU

That is: suppose that I know that I will have multiple descendants who see different measurement results, and suppose that my actions now may have different consequences for different descendants (for instance, if I take a bet on what the outcome of the measurement is going to be). Then: how am I to balance the interests of my various descendants?

This provides a framework to make probability talk, as applied to branching, meaningful: we can say that two groups of my descendants are equiprobable if I am indifferent between actions that give some reward to the first set and actions which give the same reward to the other set. (Whether this definition picks out *probability* or just some facsimile is another matter; see Saunders (1998, 2005) and Wallace (2005a, 2005b) for arguments that it is, and Greaves (2004) for a dissenting view.)

Rather more surprisingly, Deutsch was able to prove that not only must rational agents in a quantum universe make decisions by assigning probabilities to the branches, they must assign probabilities *equal to the quantum weights*. That is: the Born rule is derivable, within unitary (i. e., Everettian) quantum mechanics, from decision-theoretic principles.

Deutsch's proof was heavily criticised when it first appeared (e. g. by Barnum, Caves, Finkelstein, Fuchs, and Schack (2000) and (Gill 2005)); I have argued that such criticisms are mistaken in Wallace (2003b) and presented my own hopefully-improved version of Deutsch's proof in Wallace (2003c). But both Deutsch's proof and mine are quite heavily entangled with technical details of quantum mechanics, such that it can be rather difficult to see in virtue of what, exactly, Deutsch's rather startling result actually holds.

This paper attempts to remedy this. It is not a mathematically rigorous derivation of Deutsch's result (see the above papers for that); rather, it is a presentation of the 'bare bones' of the result. To clarify exactly what features of quantum mechanics make the proof possible, I consider not just Everettian quantum mechanics, but more generic models of ' branching universes: three in all, each more similar than the last to full quantum mechanics.

The result of this investigation is, I think, interesting: quite specific and subtle features of quantum mechanics are required for the decision-theoretic derivation to be fully satisfactory. If we want probabilities in branching universes, not just any branching universe will do.

2   Life in branching universes

By a 'branching universe', I do not mean an indeterministic universe where the past does not determine the future; I mean a universe which literally, physically, and fairly frequently splits into multiple copies.

I shall present three different sorts of branching universe in the course of this paper, but to make clear the sort of thing that I am considering, I shall describe the first sort — the *minimal branching universe*, or MBU — here. Such a universe is specified by the following:

1. A set $\mathcal{U}$ of possible instantaneous states — not of the entire universe, but of a branch. Elements of $\mathcal{U}$ play the role in branching universes that instantaneous states of the universe play in non-branching universes.

2. A relation $<$ defined between pairs of elements of $\mathcal{U}$, with the intended reading of $x < y$ being 'y is in the immediate future of x'.

In terms of $<$ we can specify a family of relations $<_n$ as follows: $x <_n y$ iff there exist elements $z_1, \ldots z_{n-1}$ of $\mathcal{U}$ such that $x < z_1 < \ldots z_{n-1} < y$ (to be read as 'x is n moments before y'. We can also define $<^*$ (to be read 'x is before y) by: $x <^* y$ iff for some n, $x <_n y$. (Formally, this makes $<^*$ the transitive closure of $<$). $<$ has the requisite structure if we require

1. For any x and y, there exists z such that $z <^* x$ and $z <^* y$ (all states are connected)

2. If $x <^* y$ then $\neg(y <^* x)$ (no closed loops)

3. If $x < y$ and $z < y$ then $x = z$ (no recombining of branches)

4. For every x there is some y such that $x < y$ (no last moment of time)

A pair $(\mathcal{U}, <)$ satisfying these constraints is intended to be a *complete* description of a MBU (hence 'minimal'; other branching universes will be given a richer structure). Time is discrete in the model but this is only for technical convenience; nothing hangs on it.

Topologically such a universe would look something like a tree, bifurcating in the future direction. an agent at any point in the tree would have a unique past but multiple futures. (See figure 1 for an example.)

Figure 1: A branching universe

Time

What would it be like to be such an agent? In particular, what attitude should he take towards an event which will lead to branching, and thus to

multiple futures $f_1 \ldots f_n$ for him? One possible attitude might be uncertainty: such an agent should be in the same sort of cognitive state that agents in non-branching universes adopt when they know that one of $f_1 \ldots f_n$ will obtain but are ignorant as to which. Call this possibility *Subjective Uncertainty* (SU), where the 'subjective' acknowledges that from a God's-eye view all is determinate.[1] It was first proposed, so far as I know, by Saunders (1998), and I defend it in Wallace (2005a).

An alternative, and in some ways more radical, attitude might be called, by contrast, *objective determinism* (OD). Branching leads, deterministically, to the agent having multiple future descendants. Rationally speaking, he should act to benefit his future descendants, for exactly the same reason that people in non-branching possible worlds would act to benefit their single descendant. Situations of conflict may arise between the interests of his descendants (such as when he takes a bet which pays off only in one branch), in which case he will have to weigh up how much he wishes to prioritise each descendant's interests.

It is not the purpose of this paper to resolve which of these perspectives is more appropriate; I present them as background. My concern is slightly different: what strategy should the agent adopt towards *rational action*? One natural strategy is given by the

> **Probabilistic Hypothesis:** In a branching universe, a rational agent faced with a branching situation should assign a probability measure across the branches, and should act in such a way as to maximise expected utility with respect to that probability measure.
>
> In particular, if he is offered the choice between a number of games which lead to various of his future descendants receiving some fixed reward, he should rationally prefer that game in which the reward is given to the set of descendants with the highest total probability.[2]

My purpose here is to investigate to what extent the Probabilistic Hypothesis can be defended in various sorts of branching universe, and to ask further: what constraints on the probability measure are imposed by the physical structure of the universe.

For future use, let us identify two ways in which the Probabilistic Hypothesis might be threatened:

Threats of probabilistic failure: Probabilistic failure would occur if there existed some circumstance in a branching universe in which agents were either rationally required, or at any rate rationally permitted, to violate the Probabilistic Hypothesis. Probabilistic failure does not in any way make a branching universe hostile to rational agents; it simply means that they will have to use something other than the Probabilistic Hypothesis as a guide to action.

---

[1] But it need not be linked to first-person expectations: 'there will be a sea battle tomorrow' might be as uncertain as 'I will see spin up'.

[2] In fact, most decision theorists accept that this 'special case' implies the general result, given certain technical assumptions: it does not seem possible to give quantitative meaning to the utility of an action other than via the odds we would accept for it.

Threats of incoherence: Incoherence would occur if there existed some circumstance in a branching universe in which no rational action at all was possible — that is, some circumstance in which any course of action violated some intuitively reasonable rationality principle. Widespread incoherence would not indicate that a branching universe was metaphysically impossible, but it would mean that intelligent life (at least of our kind) could not really function in such a universe.

We can also distinguish between widespread and localised failure of the Probabilistic Hypothesis. If either incoherence or probabilistic failure occurs in reasonably generic circumstances then the Probabilistic Hypothesis must be abandoned wholesale; conversely, if the Probabilistic Hypothesis fails only in extremely specialised and contrived situations then it remains a useful guide to action emphceteris paribus.

3   Decision-theoretic assumptions

I will be concerned (at least to begin with) with a somewhat simplified decision-making situation. An agent will be asked to choose between a variety of possible futures (involving variable amounts of branching, and variable events on each of the branches). Crucially, notice that here 'futures' means future states of the entire universe, not just specific future branches.

I shall assume that the rewards given to the agent's future selves are in the forms of tokens of some description, to be cashed for items of genuine value at some later time t_{cash}. The problem may be further simplified by assuming that there is only *one* sort of token. (The token may be thought of as a large-denomination banknote which may be spent only after t_{cash}.)

I shall require the agent to conform to the following principles:

Dominance: If two possible futures A and B have the same branching structure and differ only in that branch A leads to a reward being given and B does not, then A is preferable to B.

Preparation Indifference: If two possible futures are *physically identical* after some time t which is prior to the token-cashing time t_{cash}, then the agent is indifferent between them.

Constancy: If two possible futures each lead to the token being given out on either all branches or on none, the agent is indifferent between them.

How are these axioms to be justified? Preparation Indifference is a consequence of the stylised nature of the problem as I have formulated it: the only outcomes on which the agents preferences depend are possession or non-possession of tokens and the tokens are of no use to the agent prior to t_{cash}.

Dominance at first sight appears to follow from the assumption that I care about the interests of my future selves. Suppose I am offered one of two bets. The first bet I know will certainly benefit all of my descendants at least as

much as (and possibly more than) the second; therefore I should take the first in preference to the second.

Though this seems persuasive (to me at any rate) it has been challenged by Adam Elga[3], who points out that it is in conflict with another intuitively plausible idea: that we might seek to maximise diversity in our successors. For instance, suppose that I might have a successful career either as a musician or a physicist. I have some preference for becoming a physicist, but I'd rather become a physicist in some branches and a musician in others, than a physicist in all branches.

You either get this intuition or you don't. I don't, as it happens, but plenty of people do, and so it deserves a reply. Mine comes in two stages:

1. Something similar actually happens in ordinary decision-making, even without the assumption of branching. I might have a last square of chocolate which I can give to one of two friends. If I actually had to choose I'd give it to Alice rather than Bob ... but I'd much rather toss a coin and decide at random — apparently in violation of **dominance**.

   This is not particularly mysterious, nor particularly difficult to incorporate into decision-theory without contradicting **dominance**. We should just acknowledge that the options are underdescribed: the coin-tossing has as its two outcomes Alice and Bob getting the chocolate *as a result of a fair process*, and I prefer either of these outcomes to 'Alice gets the chocolate because I chose to favour her'. The preference can just be incorporated in a richer description of the set of consequences.

2. We might get around this by stipulating that the agent forgets that the decision was made in the first place. But in that case, he will just reverse it — after all, if I'd rather be a physicist than a musician and momentarily decide to be a musician, I'd be irrational not to change my mind back again.

It is possible to come up with examples to which neither of these responses seem to apply (in the discussion of Elga's proposal in Oxford, one participant suggested an agent who chooses to start one of two different new civilisations on a newly terraformed planet) but these rapidly seem to become extremely contrived, and don't seem to apply to everyday decision-making situations.

Another fairly contrived example is 'quantum suicide', or 'quantum russian roulette', for many years a standard after-hours discussion topic for foundationally inclined physicists and recently discussed in print by Lewis (2001), Lewis (2000), Papineau (2003), and Tappenden (2004). Suppose that I am utterly selfish, caring nothing of others, and that I am offered a large bet which I receive only on a few branches: in all other branches I am instantly and painlessly annihilated. Since I know that all of my successors will receive the reward (no matter that in many branches I will have no successors at all), I should accept the bet. If this argument is correct then it violates Dominance: the action where

---

[3]In a seminar in Oxford, not yet published so far as I know.

I get the reward in all branches dominates the one where I am annihilated in some branches, but the quantum roulette argument suggests that I should be indifferent between the two.

I am myself sceptical about quantum roulette, but for our purposes the important point is that whether or not it is rational, it occurs only in highly specialised contexts: those where I am instantly annihilated if I lose the bet, and where in addition I care about nothing except my own feelings.

Quantum roulette, and the Elga variety-driven counter-example, offer what section 2 called 'localised threats of probabilistic failure'. That is, the threat they pose to the probabilistic hypothesis applies only in very specialised circumstances (and, crucially, in circumstances which are not at all connected to the experiments which have been used to test quantum mechanics!) From here on I shall assume that Dominance remains justified in 'generic' circumstances, even if we can occasionally contrive situations where it fails.

The justification of Constancy presents rather more serious problems. In a branching universe, Constancy is equivalent to *Branch indifference*, the principle that an agent is indifferent to the occurrence of branching provided that each of his successors receives no reward or punishment subsequent to the branching. And this is far from obvious. It does not seem at all unreasonable to form the view that one wishes to have (for instance) as many successors as possible.[4]

Is it *in fact* unreasonable? It depends upon the details of our branching universe, but there are two strong objections available in any branching universe where branching is occurring all the time:

1. The epistemic objection: to take decisions in such a universe, an agent who was not branch indifferent would have to be keeping microscopically detailed track of all manner of branch-inducing events (such as quantum decays, in an Everettian universe) despite the fact that none of these events have any detectable effect on him. This is beyond the plausible capabilities of any agent.

2. The small-world objection: it has long been recognised (see, e. g., Savage 1972) that decision-making will be impossibly complicated unless it is possible to identify (in at least a rough-and-ready manner) a point after which the dust has settled and the value to an agent of consequences can actually be assessed. But if an agent is not branch indifferent, then such a point will never occur, and he will be faced with the impossible task of calculating how much branching will occur across the entire lifetime of the Universe (contingent on his choice of action) in order to weigh up the value, now, to him of carrying out a certain act.

Both of these objections rely on the assumption that a rational strategy must actually be realisable in at least some idealised sense. In previous work (Wallace 2003c) I took this as self-evident, but to my surprise this has not

---

[4]If we are justified in adopting the SU view described in section 2, Constancy is actually compulsory: branch indifference just becomes indifference between two actions each of which is certain to give us the same fixed reward.

7

generally been accepted (mostly this has emerged in conversation; however, see also Lewis 2003). I therefore offer a brief defence:

1. Decision theory is something of a hybrid. It is to some extent normative (that is, it tells us what we *should* do, and exposes us to rational criticism if we violate its precepts); it is to some extent descriptive (that is, it provides an idealised account of actual decision-making). Both of these are impossible if decision theory instructs us to do something wildly beyond even our idealised abilities.

2. If we are prepared to be even slightly instrumentalist in our criteria for belief ascription, it may not even *make sense* to suppose that an agent genuinely wants to do something that is ridiculously beyond even their idealised capabilities. To adapt an example from Wallace (2003c), suppose I say that I desire (*ceteris paribus*) to date someone with a prime number of atoms in their body. It is not even remotely possible for me to take any action which even slightly moves me towards that goal. In practice my actual dating strategy will have to fall back on "secondary" principles which have no connection at all to my "primary" goal — and since those secondary principles are actually what underwrites my entire dating behaviour, arguably it makes more sense to say that *they* are my actual desires, and that my 'primary' desire is at best an impossible dream, at worst an empty utterance.

Unconvinced? Have patience: we shall see in section 8 that in quantum mechanics (and universes with a similar sort of branching) there is a much more powerful defence of **constancy** available, which does not rest on these practical issues. For now, note only that although Constancy is not *prima facie* obvious, it is extremely natural, at least in a universe where branching is ubiquitous.

4 Decision making in minimal branching universes

To begin with, consider the simple model of a branching universe described above: the 'minimal branching universe', or MBU, in which the branching structure is a complete description of the universe.

In particular, consider the following example of an MBU. There is exactly one splitting event, in which the world splits in two, and it is about to occur. In the first branch (say) some subatomic particle has decayed; in the second, the particle remains undecayed.

An agent must choose between two actions. If he takes action A, he will be rewarded by receiving a thousand dollars (i.e., the token) iff the particle decays. Action B has the opposite consequence: the thousand-dollar reward is given iff the particle does not decay. The agent has no interest at all in whether or not the particle decayed except insofar as it affects his reward, and the bet will be paid within a millisecond of the decay taking place.

The twist is this: the physical process by which the bet is paid has been designed to erase all knowledge of whether or not the particle decayed, and also all knowledge of which bet was taken. This means that according to whether the agent chooses A or B, the possible states of the Universe are given by Figures 2 and 3:

Figure 2: Choice A

Agent with cash   Agent with nothing

Payoff   No payoff

Decay   No decay

        Agent chooses A

Time t_0

Observe that:

1. Choices A and B differ only between t_0 and t_1: after t_1 they are completely identical.

2. The agent is completely indifferent to anything that happens between t_0 and t_1. (This may seem counter-intuitive, in that he cares about getting the thousand-dollar bill, and that happens before t_1. However, he only values the cheque insofar as he gets to keep and spend it after t_1; the value to him of possessing it just in the one millisecond between the decay and t_1 is nil. In the formal 'token' talk of the previous section, t_{cash} is clearly greater than t_1.)

3. Therefore, by Preparation Indifference, the agent is indifferent between choice A and choice B: that is, at least in this circumstance the agent is indifferent between a bet on the decay and a bet against the decay.

4. Therefore, the agent's preferences fit the Probabilistic Hypothesis, with a probability of 0.5 for each branch. That is, *the agent regards the decay and its absence as equiprobable.*

Furthermore, the knowledge erasure used in the above argument plays no essential role. For the agent doesn't care, *per se*, which bet he placed once he

Figure 3: Choice B

Agent with nothing   Agent with cash

No payoff   Payoff

Decay   No decay

Agent chooses B

T**t**$_0$me

gets his reward, nor is he interested in whether or not the decay occurred. That is, each post-branching agent is indifferent to the erasure. If (banknote-plus-erasure) has the same value as (banknote) and (no-banknote-plus-erasure) has the same value as (no-banknote), it follows that removing the erasure from the specification of A and B does not change the agent's indifference between them.

It isn't even important whether the erasure is *physically possible*. (In realistic physical theories, I suspect, it wouldn't be: the information would be dispersed so thoroughly that no-one could recover it, but in principle it would still be there). For suppose $x_1, \ldots, x_N$ are all the possible states where the agent has received his cheque and where the information about the bet is sufficiently thoroughly dispersed that the agent has no access to it; suppose $y_1, \ldots y_M$, similarly, are all the possible states where the agent got no cheque and where all the information has been dispersed. Then the dispersal process in choice A is one of the MN processes

(thousand dollars+decay) $\longrightarrow (x_n)$;
(thousand-dollars+no decay) $\longrightarrow (y_m)$

and (since the agent is indifferent to dispersal) he is indifferent as to which process it is. (If he knows sufficient physics he may know that only one is actually possible, but this doesn't affect his ignorance about, and indifference to, which one is the possible one. It follows that he is indifferent between the *actual* process which occurs when he chooses A, and the (probably physically impossible) process which leads to two branches, one in state x_1 and one in state y_1.

Similarly, the agent is indifferent as to which dispersal process is used in case B, and (applying the same logic) is indifferent between choice B and a process

which again leads to two branches one in state x_1 and one in state y_1. It follows again that he is indifferent between A and B.

It might even be the case that erasure must itself cause branching. No matter: by **constancy** an agent is indifferent between (a) a branching event each of whose post-branch states is one of x_1 through x_n, and (b) the occurrence of x_1 with certainty.

This is a rather striking result. In an MBU, an agent is *rationally compelled* to assign equal probability to each branch in a two-way branching event. No actual assumption of indifference between the branches was made: rather, the result was derived from existing principles of decision-theoretic rationality which don't presume a branching Universe at all. Furthermore, the result easily generalises to arbitrarily many branches: any bet on one branch can be made physically identical to a bet on another branch via the erasure mechanism, and therefore any two branches must be regarded as equally likely. (If the number of branches is finite (equal to N, say), this entails that the agent's beliefs are uniquely represented by assigning probability $1/N$ to each branch.)

To achieve some grasp on how this can have occurred, recall that historically symmetry has always rivalled frequency as a constraint on probabilities: we assign each face of a die probability $1/6$, for instance, because (somehow) they are correctly regarded as being related by a somehow-relevant symmetry. But in the absence of branching, this argument falls foul of the observation that ultimately one event or the other must actually occur, breaking the symmetry. In modern physics we normally get around this by requiring either that the theory is stochastic (which reintroduces objective chance as primitive) or that the probability distribution over initial conditions obeys the symmetry (which begs the question).

In a branching universe, on the other hand, no single outcome occurs. It is therefore possible that the symmetry constrains the probability without further ado. Of course, the problem is that the two-way splitting event described earlier isn't really completely symmetric: tiny details, irrelevant to the agent (like: did the atom actually decay) break the symmetry. It would have been tempting to resolve this, rather than via an excursion into erasure and dispersal, via some principle like

> *Irrelevancy neglect:* In a branching universe, any details of the future which are irrelevant to all of my successors should be irrelevant to me right now.

However, it is at least interesting, and will be crucial later, that we can establish the equiprobability result without adopting *irrelevancy neglect*.

5 MBU theories are generally incoherent

Consider the following scenario in an MBU, contemplated by some agent at t_0. At t_1 some coin is flipped (or some other chance process occurs), inducing a two-way branching process whose branches can be labelled 'H' and 'T'. In the

T branch nothing further occurs; in the H branch a second coin is flipped at t_2, leading to two further branches labelled HH and HT. The time between t_0 and t_2 is measured in seconds, and the agent receives no relevant rewards or punishments in this time. (See Figure 4).

Figure 4: Iterated 2-way branching

HH   HT   T

    H   T

    2nd flip

    H   T

    1st flip

    Agent contemplates

Time t_0
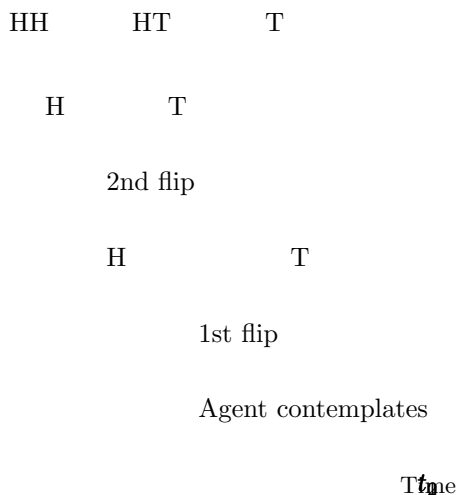
The question is: what probability should the agent, at t_0, ascribe to the first coin landing heads? The obvious answer is $1/2$: the coin-flip leads to two-way branching, and by the arguments of the previous section this requires probability $1/2$ for each outcome. Whether or not the second coin is flipped does not affect the first coin, so this result is unchanged by the second flip.

But now consider Figure 5. Here, three branches again arise, and are again labelled HH, HT and T; but they arise from a single, three-way branching event. By the same arguments each must have probability $1/3$, so here the chance of the first coin being heads is $2/3$.

But now suppose the agent is *betting* on whether the coin lands heads, and suppose the reward is given long after t_2. Then the post-t_2 states in figures 4 and 5 are identical up to the erasure of irrelevant data. By the arguments of the previous section, and via the **identical-futures principle**, it follows that the agent must be indifferent between betting on heads in the first scenario, and betting on it in the second. But this is incompatible with the assignation of *different* probabilities to the two cases.

I conclude that in general, MBUs in which iterated branching is possible are not compatible with any consistent allocation of probabilities to the branches without violating one or more decision-theoretic axioms. In fact, the situation is worse than this: since the argument of section 4 actually compels us to accept the contradiction above, there is no rational strategy at all — probabilistic or otherwise — compatible with the decision-theoretic axioms. In the terminology

Figure 5: 3-way branching

HH   HT   T

3-way flip

Agent contemplates

Time

of section 2, this is a threat of widespread incoherence: it shows that rational
life is impossible in any MBU in which branching occurs at all frequently.[5]

6 Weighted branching

If all branching universes were MBUs, then this paper would simply be *reductio*
of the idea that branching can be interpreted as uncertainty. But in fact there
are other sorts of branching theories. Consider next the *weighted* branching
universes, or WBUs. In theories of this kind, the universe is not completely
defined by specifying the branching structure of nodes. Rather, to each node n
we must also assign a *weight* $w(n)$ (to be regarded as a new categorical property
of that node). Weights should be positive elements of some ordered commutative
group (on first reading, just regard weights as real numbers between 0 and 1)
and — crucially — they are required to satisfy the following principle:

Weight additivity If x is a node and $y_1, \ldots y_n$ are nodes satisfying

1. $x <_* y_j$ for each j (each y_j is futurewards of x);
2. not $y_i <_* y_j$ for any i, j (no two y_i are in the same branch)
3. if $\mathcal{B}$ is a maximal $<_*$-ordered subset of $\mathcal{S}$ (i.e., a branch) then $\mathcal{B}$ contains some y_i

then $w(x) = w(y_1) + \ldots w(y_n)$.

---

[5]There is one *prima facie* loophole: I made use of Constancy at one point in the proof of
section 4, and it is perhaps open to an inhabitant of an MBU to reject it. That this loophole
can be closed, I leave as an exercise for the reader.

Now, suppose we consider again 'Choice A' and 'Choice B' from section 4: that is, we consider bets on the decay or otherwise of an atom whose decay induces 2-way branching. This time, however, suppose that we are in a WBU and that the decay and no-decay branches have weights x and y respectively. Now the post-t_2 states of the Universe, after the information about the bet has been erased, are

Choice A: One weight-x branch in which the agent receives a thousand dollars; one weight-y branch in which the agent receives nothing.

Choice B: One weight-y branch in which the agent receives a thousand dollars; one weight-x branch in which the agent receives nothing.

(Note that as a consequence of **Weight additivity**, the erasure process cannot affect the weight of the branch(es) in which a reward is given.)

It follows that in a weighted branching universe, an agent is rationally compelled to regard two branches as equally likely only *provided that they have equal weight.* This dissolves the paradox of the previous section: there is no allocation of weights whereby the H and T branches have equal weight after t_1 and yet the HH, HT and T branches have equal weight after t_2.

This takes us a good way towards establishing the Probabilistic Hypothesis for WBUs. In fact, it takes us a good way towards establishing a much stronger principle, which we might call the *Weight Principle*: in a WBU, agents are rationally compelled to set their probabilities proportional to the weights. (Certain assumptions about the richness of the branching structure would also be required to prove the result in full generality.)

7   What are weights?

By replacing an MBU by a WBU, we appear to have defeated the threat of incoherence identified in section 5. But one might still feel uneasy. What, after all, *are* these weights, and why on Earth should we care about them in decision-making?

One reason to be suspicious of them is that two WBUs might differ only by a different assignation of weights to the various nodes. This being the case, how could an agent possibly know that he was in one WBU rather than the other?

This objection requires us to engage with the complicated question of how we might test a branching-universe theory. A full answer to this would lie beyond the scope of this thesis (see Wallace (2005a) for more detailed discussion) but in short, we require that the theory predicts that it is rational (*ceteris paribus*) to expect to find oneself in a high-weight world.

So: suppose one is about to carry out a long sequence of experiments whose results are uncertain (and so which, on the assumption that the universe is a WBU, will lead to branching). For convenience, we suppose that weights are real numbers and that we have proved the Weight Principle in full generality (i. e., not just for equal-weighted branches).

Suppose that some possible sequence of outcomes x is given high weight on the assumption that we live in WBU1, and low weight on the assumption that we live in WBU2. Then an agent who accepts WBU1 should expect (with very high subjective probability) that he will experience x and one who accepts WBU2 should expect (again with very high subjective probability) that he will not experience x. A standard application of Bayes' Theorem tells us that, unless he'd previously given extremely high subjective probability to being in WBU2, if he experiences x then he ought tentatively to conclude that he is in WBU1.

To be sure, in reaching this conclusion he implicitly accepts that there are many other low-weight copies of him who have not seen x, and who may therefore very well have reached the wrong conclusion. But why should he care about these poor unfortunates? After all, they have low weights, and we've proved that rational agents don't care about low-weight situations.

So I do not think it is correct to regard weights as experimentally inaccessible. No experiment can determine *for certain* what the weights are, but this is already the case in stochastic theories, and they are little the worse for it.

But there is another, more decision-theoretic problem with the WBUs. However elegant the proofs, one is still left wondering *why* rational agents should give a damn about this 'new categorical property'.

To make this vivid, suppose (never mind how) that it is possible to build a 'weight detector', which registers the weight of a node on some sort of dial. Why on earth should an agent care about the reading on that dial? It doesn't impact at all on his hopes and fears, his loves and hatreds, his fame and fortune, or anything else which people normally value. And if he doesn't care, why should he assign different probabilities to branches with different readings on the dial?

This conundrum cannot be turned into an outright *proof* of the irrelevance of weights by erasure/dispersal arguments, as we did before: it's an essential consequence of weight additivity that weight cannot be changed except by inducing branching. But it is still troubling: granted that it isn't *in fact* physically possible to change the weight of a branch, the fact that an agent *would* (were it possible) be indifferent to the choice of doing it is still worrying.

The worry can be given precise form by noting that caring about the weights in deciding how to bet on branching is in violation of the *future indifference principle*: since each of my future descendants is indifferent to the weight of his node, then I should *right now* be indifferent to each of these weights.

The robust response here would be to say: very well, so much the worse for the *future-indifference principle*. If accepting it would force our decision-making process into incoherence, we'd better not accept it! But one might equally well argue that the principle is as plausible as any other decision-theoretic arguments, so that we have another widespread Incoherence Threat of the kind discussed in section 2 (albeit perhaps a weaker one than in the MBU case). Having reached this rather unsatisfactory conclusion, we leave the WBU theories.

8   Emergent branching

Our third class of branching universe, the *emergent branching universes* (EBUs) are structurally quite dissimilar to MBUs and WBUs. At the most fundamental level of description, EBUs do not involve branching trees of classical-world-like states at all: they involve the deterministic and unbranching evolution of a system which does not look in the slightest like our ordinary world.

The branching in EBUs, as their name suggests, emerges from the underlying system as a higher-level, approximate description — rather as higher-level ontology in general emerges from lower-level descriptions (see Dennett (1991) for more on this in general, and Wallace (2003a) for the particular case of quantum theory). A description of an EBU in terms of branching classical worlds is perfectly valid, provided it is understood that it is only an approximate description, leaving out some details and glossing others.

What does this 'approximate description' look like? It looks much like a WBU (a branching tree of classical states with a weight attached to each node) with one crucial distinction: any such description *must* be understood as a coarse-graining of a finer description which is itself emergent from the 'fundamental' EBU.

By a coarse-graining: I mean that each of the nodes of the emergent WBU must be regarded not necessarily as describing a single state, but as corresponding to an imprecise description fitting several. A more precise description would break the nodes down into many nodes, leading to a more detailed and complicated tree structure. (See figure 6 for an illustration of this; the nodes of the coarser-grained description are shown as circles around the finer-grained nodes to which they correspond.

Figure 6: A coarse-graining

Coarse-grain

To further explicate the notion of a coarse-graining, we can define an *event* as a set of states of $\mathcal{S}$ (the terminology is taken from decision theory). An event might be characterised by some description (e. g. , "there is a cat on the mat", or "X wins at least a thousand dollars") or it might be indescribable in any

simple way. In any case, a coarse-graining of a WBU can now be described as a WBU-like branching structure (complete with weights at each node) but with events rather than states for nodes.

In 'genuine' WBUs, the notions of 'coarse-graining' and 'event' are at most a technical convenience: for foundational work we might as well work with the maximally fine-grained description. But in EBUs, crucially (and by definition) *there is no fine-grained description.* Any emergent WBU description may be replaced by a finer-grained description, the whole process terminating only at the (inherently ill-defined) point at which the usefulness of the emergent WBU description begins to fail and we are forced to resort to the 'fundamental' description. As such, the emergent branching description of an EBU is always a tree of weighted events, and the 'states' drop entirely out of the picture.

How should a rational agent in an EBU choose to act? Such an agent's best description of his world will be in terms of some emergent WBU description, so the decision-theoretic results of the last section apply in large part. There are, however two crucial differences.

Firstly, Constancy is unavoidable for an agent in an EBU. For Constancy is threatened for a rational agent only by the possibility that such an agent cares about the raw number of his successors, and 'number of successors' is simply not defined in quantum mechanics or in any EBU. For although any given WBU description gives a definite value for the number of branches futurewards of any node, that value will in general fluctuate wildly as we move to successively finer-grained descriptions. (Put another way, all an agent can count up are the number of future chains of *events*, and an event can always be subdivided.)

On the reasonable assumption that an agent cares only about features of physical reality and not about artefacts of a particular approximate representation of that reality, we conclude that no such agent can care about the number of his successors. It is not that 'number of successors' is empirically inaccessible (as, I argued in section 3, is likely even in an MBU). It is that 'number of successors' is just meaningless.

Secondly, 'weight' is not a property of individual agents or their branches in an EBU. It is meaningful to talk about ascribing a weight to a node, but because of the indefinite possibility of fine-graining all we can do is say 'the weight of event E is w'. Statements like 'branch # 20445 has weight w' are meaningless because the idea of referring to a specific branch is itself meaningless.

Recall that at the end of the last section, I observed that the Probabilistic Hypothesis clashes with the intuition (formalised in the Future Indifference Prinicple) that if there is some property of an agent's branch to which he is indifferent, that property should not have featured in the agent's previous decision making. But 'weight' in an EBU is no longer any such property. It does not make sense for an agent to refer to 'the weight of my branch': all he can refer to are the (differing) weights of any number of events whose descriptions conform to his surroundings. As such, the threat from the Future Indifference Principle does not apply: an agent has no branch weight of his own to be indifferent to. (To repeat the thought experiment of a 'weight detector': in an EBU is it conceptually and not just practically impossible to build such a thing.)

17

As such, the decision-theoretic situation in an EBU is far more satisfactory than in a WBU. In the latter case we identified two potential problems: the lack of a pressing motivation for Constancy, and the feeling that weights, being properties of an agent about which he has no interest, ought to be decision-theoretically irrelevant. Neither problem applies to EBUs: the former because lack of any determinate notion of branch number forces Constancy on us; the latter because weights are no longer properties of particular branches. The Probabilistic Hypothesis is unproblematic in an EBU, as is section 6's proof that rational agents set their probabilities proportional to the weights.

9   Quantum mechanics

The (rather predictable) punchline, of course, is that Everettian quantum mechanics is an emergent branching universe. For the branching structure in quantum mechanics is picked out by the process of decoherence, and (as I discuss in more detail in Wallace (2003c)) the decoherence basis has the no-finest-grain property described above. We can always fine grain the decoherence basis by refining the accuracy with which it is described or by including more of the environment; the price is that the level of decoherence between basis elements may fall, eventually leading (at an inherently ill-defined point) to failure of decoherence and the requirement that we revert to the full, unitary description.

The correspondence with the structure of EBUs can be made explicit by using the consistent histories formalism. Working in the Schrödinger picture, let $\mathcal{H}$ be the space of projectors used to build up some particular decoherent and quasiclassical space of histories.[6] Elements of $\mathcal{H}$ will in general be infinite-dimensional projectors, and it will generically be possible to fine-grain them without losing decoherence; trivially, it will also be possible to coarse-grain them. (As a concrete example (due to Saunders (2005)) suppose that decoherence selects the configuration-space basis of some set of particles. The projectors will then project onto subsets of configuration space, and in general it will be possible to fine-grain these subsets without loss of decoherence.)

Now, consider the Boolean algebra $\mathcal{A}_{\mathcal{H}}$ generated from $\mathcal{H}$ by the admissible fine-grainings and the coarse-grainings. Any collection of elements $\widehat{P}_i$ from $\mathcal{A}_{\mathcal{H}}$ such that $\sum_i \widehat{P}_i = \widehat{1}$ will constitute the basis for a quasiclassical history space, and thus a possible branching-universe description of the quantum state.

In general this algebra, insofar as it is well-defined at all, will be nonatomic: there will be no 'finest' graining before decoherence fails. Mathematically we can perhaps idealise $\mathcal{A}_{\mathcal{H}}$ as infinitely fine-grainable, with the caveat that eventually (as we continue to look at finer and finer grainings) decoherence will fail and the projectors will no longer yield a valid branching-universes description.

(As a concrete example (due to Saunders (2005)) suppose that decoherence selects the configuration-space basis of some set of particles.[7] The algebra $\mathcal{A}_{\mathcal{H}}$

---

[6]I assume for technical convenience that the set $\mathcal{H}$ does not vary with time.

[7]Note that it is the configuration space of that set of particles that we are considering, not the configuration space of the universe as a whole: many particles are likely to be relegated

will then consist of all projectors onto open subsets of configuration space. When these projectors are taken as too fine-grained the decoherence condition will cease to apply and we will lose quasiclassicality, but there is no precisely-defined point at which this occurs.)

Quantum mechanics, then, defines three regimes of coarse-graining. The first, finest-grained, regime is one in which the 'branching-universe description' is purely mathematical: decoherence fails and so there is no sense in which the branching description captures an emergent branching reality. In the intermediate regime decoherence is applicable and the emergent branching structure is accurately identified; in the third and coarsest regime, the graining is too coarse, and unambiguously distinct chunks of macroscopic reality (such as states in which a cat is alive and states when it is dead) are lumped together. It is only the middle regime which really describes the emergent classical branches, but — crucially — that regime has boundaries which are inherently ill-defined.[8]

So: since quantum mechanics is an EBU, the Probabilistic Hypothesis holds for it, and rational agents in a quantum universe must maximise expected utility with respect to quantum-mechanical weights. In fact, it is even more apparent in QM than in a 'generic' EBU that the weight is a physically significant player and not just some function placed over an existing branching structure (as threatened to be the case in WBUs). For the branching structure of the quantum state, at whatever grain it is viewed, emerges from the underlying unitary dynamics on Hilbert space, and the latter arena cannot even be specified without the inner product, from which the weight derives.

10   Conclusion

In branching universes, probabilities emerge from symmetries. If an agent is to be split into n copies which are identical except for erasable irrelevant details, he must value each of these future descendants equally. This leads to intertemporal incoherence unless the 'minimal' branching structure is supplemented by a notion of 'weights' for each branch. If it is so supplemented, a rational agent in the resulting universe must value his successors in proportion to their weights. That is, in such a universe the weights play the same role in decision-making

---

to 'the environment' and traced over.

[8]The philosophical problem of how to understand 'inherently ill-defined' predicates has recently received considerable attention from philosophers of language (see (Keefe and Smith 1997) for a recent anthology). One currently-popular strategy, empiricism (recently defended by Williamson (1994)) poses a *prima facie* threat to the strategy I describe above: according to empiricism, the boundaries of application for 'vague' predicates like 'quasiclassical' are actually perfectly sharp, but language users do not and cannot know where they are. This would seem to suggest that, contrary to my argument, there really is a 'finest grain' description.

I am inclined to respond to this threat as follows: rational agents (as was argued in section 3 must choose their strategies so as to be implementable in at least idealised circumstances. Even if (contrary to my own view) we ought not reason to criticise someone for caring about branch number when there is not the slightest practical prospect of determining it, surely our criticisms have force if determining it is *in principle impossible*, as empiricism would have it? However, the matter probably deserves further study.

as 'probability' does in non-branching universes (whether or not the weights should just be identified *as* probabilities).

This leads to tension with the principle that since agents do not care about their own branch weights, they should ignore the branch weights of their successors — a principle which, if accepted, leads to intertemporal incoherence as before. The overall framework is also dependent on the principle of indifference to branching *per se*, without which the 'probabilities' of future branches would not add up to one.

However, suppose that the weighted branching structure is not fundamental, but emerges from some underlying physical reality in such a way that there is no 'finest-grained' structure of branches but only a vaguely-defined cutoff point below which 'branching' talk ceases to be useful (as is the case in quantum mechanics). Then neither of these problems arises: indifference to branching is forced upon us by the lack of any coherent notion of 'branch number', and the objection that we should not care about our own weights is removed when it is realised that weights do not attach to individuals at all.

The idea that our future is not determinate but consists of a myriad branching possibilities appears fanciful to some, appealing to others. It is interesting, though, that when we consider how probability may be fitted into such a scheme we are forced to place some very strong structural constraints on the nature of that branching future. It is more interesting still that our best current physical theory, interpreted realistically, satisfies those constraints.

References

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://arXiv.org/abs/quant-ph/9907024.

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. (1999). Quantum Theory of Probability and Decisions. *Proceedings of the Royal Society of London A455*, 3129–3137. Available online at http://arxiv.org/abs/quant-ph/9906015.

Gill, R. D. (2005). On an argument of David Deutsch. In M. Schürmann and U. Franz (Eds.), *Quantum Probability and Infinite Dimensional Analysis: from Foundations to Applications*, Volume 18, pp. 277–292. Singapore: World Scientific. Available online at http://arxiv.org/abs/quant-ph/0307188.

Greaves, H. (2004). Understanding Deutsch's probability in a deterministic multiverse. *Studies in the History and Philosophy of Modern Physics 35*, 423–456. Available online at http://arXiv.org/abs/quant-ph/0312136 or from http://philsci-archive.pitt.edu.

Keefe, R. and P. Smith (Eds.) (1997). *Vagueness: A Reader.* Cambridge, Mass.: MIT Press.

Lewis, D. (2001). How many lives has Schrödinger's cat? Unpublished manuscript.

Lewis, P. (2000). What is it like to be Schrödinger's cat? *Analysis 60*, 22–29.

Lewis, P. (2003). Deutsch on quantum decision theory. Available online from http://philsci-archive.pitt.edu.

Papineau, D. (2003). Why you don't want to get into the box with Schrödinger's cat. *Analysis 63*, 51–58.

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

Saunders, S. (2005). What is probability? In A. Elitzur, S. Dolev, and N. Kolenda (Eds.), *Quo Vadis, Quantum Mechanics?* Berlin: Springer-Verlag.

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

Tappenden, P. (2004). The ins and outs of Schrödinger's cat box: a response to Papineau. *Analysis 64*, 157–164.

Wallace, D. (2003a). Everett and Structure. *Studies in the History and Philosophy of Modern Physics 34*, 87–105. Available online at http://arXiv.org/abs/quant-ph/0107144 or from http://philsci-archive.pitt.edu.

Wallace, D. (2003b). Everettian rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studies in the History and Philosophy of Modern Physics 34*, 415–439. Available online at http://arXiv.org/abs/quant-ph/0303050 or from http://philsci-archive.pitt.edu.

Wallace, D. (2003c). Quantum probability from subjective likelihood: Improving on Deutsch's proof of the probability rule. Available online at http://arXiv.org/abs/quant-ph/0312157 or from http://philsci-archive.pitt.edu.

Wallace, D. (2005a). Epistemology quantized: circumstances in which we should come to believe in the Everett interpretation. Forthcoming.

Wallace, D. (2005b). Tensed talk in a branching universe. Forthcoming.

Williamson, T. (1994). *Vagueness.* London: Routledge.