To appear in *The British Journal for the Philosophy of Science*

The Quantum Doomsday Argument

Alastair Wilson[1]

December 23rd, 2014

Abstract: If the most familiar overlapping interpretation of Everettian quantum mechanics (EQM) is correct, then we are constantly splitting into multiple people. This consequence gives rise to the *quantum doomsday argument*, which threatens to draw crippling epistemic consequences from EQM. However, a diverging ('parallel universe') interpretation of EQM undermines the quantum doomsday argument completely. This appears to tell in favour of the diverging interpretation. But it is surprising that a metaphysical question that is apparently underdetermined by the physics should be settled by purely epistemological considerations; and I argue that the argument for divergence based on the quantum doomsday effect is ultimately unsuccessful. I discuss how some influential treatments of Everettian confirmation handle the quantum doomsday puzzle, and suggest that it can most satisfyingly be resolved via a naturalistic approach to the metaphysics of modality.

1. Introduction

Realists of various stripes about possible worlds face the question of whether to adopt a *diverging* or an *overlapping* form of realism. (Overlapping worlds have parts in common; diverging worlds do not.) A case of special interest is Everettian quantum mechanics (EQM), also known as the 'Many-Worlds Interpretation'[2]. Since it is a live empirical possibility that EQM is correct, it is important to understand its metaphysical implications.

---

[1] Philosophy, University of Birmingham, Birmingham B15 2TT, UK; Philosophy, Monash University, Victoria 3800, Australia.

[2] For an introduction to this approach to quantum mechanics, see Wallace (2012).

Elsewhere I have given arguments, based on metaphysical and semantic considerations, for a diverging understanding of EQM. In this paper I will explore - and ultimately reject - an alternative epistemological argument for divergence. In brief, the argument is that more familiar overlapping form of EQM gives rise to a quantum-mechanical version of the *doomsday effect*. This threatens overlapping EQM with crippling epistemic consequences: if we come to accept overlapping EQM, we should become almost certain that we will have no future experience whatsoever.

I won't assess the original doomsday argument directly in this paper, though in §2 I quickly sketch it, classifying responses into two categories. In §3 I present the quantum doomsday argument, and show how it threatens overlapping EQM with what Barrett (1999) has called 'empirical incoherence'. Strikingly, the diverging version of EQM defended by Saunders (2010) and by Wilson (2012) completely undermines the quantum doomsday argument (§4). If diverging EQM is correct, then *ceteris paribus* there are just as many early temporal parts of a given person as there are late temporal parts. This can be turned into a dilemma for overlap: either it leads to absurd doomsday consequences, or it is automatically disconfirmed. Either way, divergence is the preferable hypothesis.

At this point we might reasonably suspect that something has gone wrong. What looked like a purely metaphysical question has turned out to have radical epistemological consequences. In §5 I explain how some prominent treatments of probability in EQM handle the puzzle. The problem of avoiding a quantum doomsday effect is a special case of the so-called quantitative problem with probability in EQM: the problem of explaining why objective probabilities are to be given by the Born rule. A variety of proposals exist for solving the quantitative problem, and they all clash with the indifference principle that drives the quantum doomsday argument. If these solutions can be sustained, the quantum doomsday argument is solved *en passant*.

Nonetheless, the indifference principle used in the quantum doomsday argument remains *prima facie* plausible; and treatments of Everettian probability which conflict with it have some explaining to do. Why is indifference reasoning about self-location appropriate in non-Everettian contexts but inappropriate in

Everettian contexts? In §6 I describe and defend a resolution to the puzzle, which makes use of a naturalistic approach to the metaphysics of modality and points towards a quantum-mechanical version of modal realism.

2. The doomsday argument

The original doomsday argument is a puzzle about self-locating belief. A curiosity of probability theory, it has been independently identified several times over. Perhaps the most well-known sources are (Carter 1983), Gott (1993), and Leslie (1989)[3]. Driving the argument is the idea that the evidence which we have of our *birth rank* – of the number of humans born before us – confirms the hypothesis that the total population of humans will not be very much greater than the total to have lived up to the present moment[4].

By itself, the purported probabilistic connection between our birth rank and the end of the human species does not establish that doomsday will come soon. To secure that gloomy conclusion, we need also to take into account the rapid recent rise in human population. If the population remains around current levels for at least another few millennia, then you and I will turn out to have lived surprisingly early in the history of humanity. Familiar Bayesian reasoning takes the unlikelihood of some evidence on a hypothesis to tell against the hypothesis. Accordingly, our now[5] being alive comprises evidence against the hypothesis that the population will remain around current levels for at least another few millennia. Informally: if doom lies in the distant future, then we're living

[3] There are a number of distinct versions of the Doomsday argument. The version given by Gott (1993) requires no assumptions about the distribution of humans throughout human history, and accordingly has less dramatic consequences then the version given by Leslie (1989), which takes into account the recent dramatic rise in human population. Sober (2002) refers to these arguments as 'Gott's line' and 'Leslie's wedge' respectively. The connection with EQM that is drawn in the present paper involves Leslie's version of the argument. (See also Bradley and Fitelson 2003 for further distinctions amongst versions of the argument.)

[4] It has sometimes been suggested (e.g. by Monton 2003) that the doomsday argument can be run without the need for knowledge of birth rank. In my view, Bradley (2005) convincingly argues that Monton's argument neglects observation selection effects, and that knowledge of birth rank is required for a doomsday effect after all.

[5] Throughout this paper, 'now' is intended to refer rigidly; substitute a date if you prefer.

unusually early, but if doom will be within the next few hundred years, then we're not living unusually early.

It will be helpful to see how this plays out in a toy example. To simplify matters, let's assume that only two scenarios are epistemic possibilities for us (numbers are chosen arbitrarily):

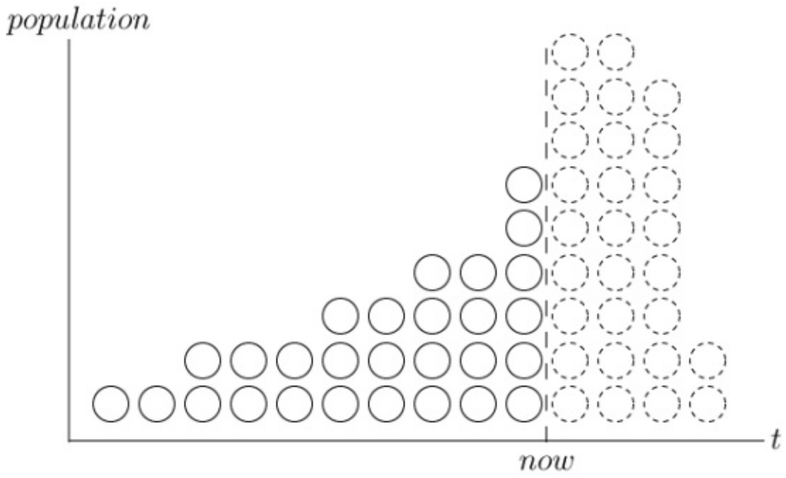

Fig 1. Doom Soon: Total population will be twice the total to now.

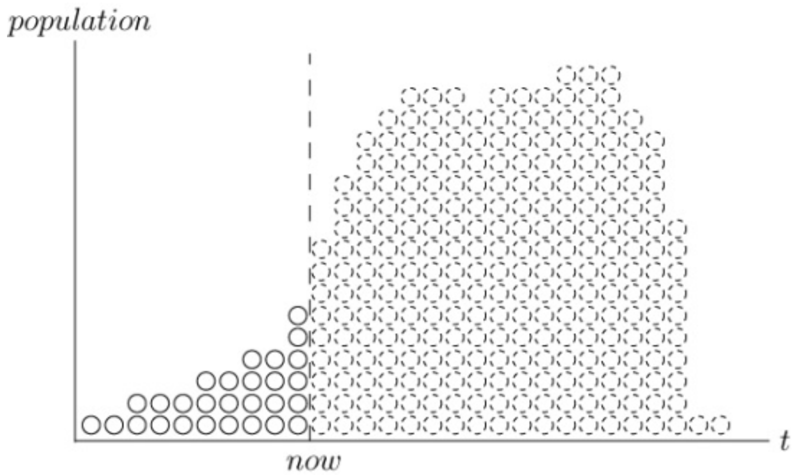

Fig 2. Doom Later: Total population will be ten times the total to now.

Evidence about birth rank is evidence about the number of people who have have lived before you. To make sense of what it would be to obtain such evidence, we can consider a scenario (perhaps involving amnesia) in which someone is completely certain that the DoomSoon scenario (timelessly) obtains but is completely uncertain of the current date. They have the information provided by

Figure 1, but without the *now* marker. For all they know, they could be the very first person to live, or the very last[6]. How should they distribute their credences amongst the various locations they might occupy?

A popular answer to this question (defended by Adam Elga, amongst others) appeals to a restricted principle of indifference:

> ***Restricted principle of indifference:*** credence should be evenly distributed amongst *similar* centred worlds: subjectively indistinguishable locations within a single possible world. (Elga 2000, 2004.)

The arguments that follow don't in fact need the full strength of this indifference principle. Any remotely plausible rule for distributing credence amongst locations will lead to doomsday effect of the sort that I'll describe[7]. For ease of presentation, though, I'll work with Elga's principle for the next few sections.

Granting the restricted principle of indifference, an agent who knows only that either DoomSoon or DoomLater obtains, and is ignorant of their own birth rank, should have the following conditional credences:

$$\mathrm{P(LiveByNow|DoomSoon)} = 1/2$$

$$\mathrm{P(LiveByNow|DoomLater)} = 1/10$$

$$\mathrm{P(LiveByNow|DoomSoon)} > \mathrm{P(LiveByNow|DoomLater)}$$

Since we're assuming that either DoomSoon or DoomLater is true, if LiveByNow confirms DoomSoon over DoomLater then it confirms DoomSoon over its negation. So conditionalizing on LiveByNow increases credence in DoomSoon:

$$\mathrm{P(DoomSoon|LiveByNow)} > \mathrm{P(DoomSoon)}$$

Thus, in this simplified setup, a doomsday shift arises merely from the discovery that we are alive now.

There are two alternative ways of escaping the doomsday conclusion: rejecting the existence of a doomsday shift, or adjusting priors to compensate for

---

[6] This assumption makes the maths simplest; but of course it isn't necessary for the effect. All that's required for some degree of doomsday-related change in credence is that the agent shouldn't start off certain of their birth rank.

[7] See Bradley & Fitelson (2003) for a defence of this claim.

it. The former strategy involves some fairly radical moves: for example, giving up some aspect of probabilism, or denying that birth rank is contingent. I find responses of this sort implausible, and I won't discuss them further[8]. A more interesting strategy involves the 'Self-Indication Assumption':

> ***Self-Indication Assumption (SIA):*** The prior probability of a world *w* should be weighted by the number of agents who exist at *w*. (Bostrom 2002)

SIA expresses the thought that our own existence is more likely in a populous world. Heuristically: such worlds contain more 'slots' that we could have been born into. Applying SIA to our doomsday scenario:

$$P(\text{LiveAtAll}|\text{DoomLater}) = 5 * P(\text{LiveAtAll}|\text{DoomSoon})$$

This effect exactly cancels the doomsday shift (see e.g. Bartha & Hitchcock 1999). But SIA is controversial, and most epistemologists reject it. Bostrom's 'presumptious philosopher' objection (Bostrom 2002) elicits strong intuitions that SIA ought not to decide between cosmological theories that differ only in their predictions for the size of the universe. I will return to SIA in §4.

The doomsday argument comes in both 'personal' and 'social' versions. Instead of considering the birth rank of some person - how far through human history they're located - we can consider the birth distance of some person-stage - how far the stage is through the life of the person of whom it's a stage. Then instead of raising the probability for the imminent end of the human race, the doomsday effect raises the probability for the imminent death of the person concerned. The personal version of the argument lacks the full force of the social version because (so we normally think) person-stages aren't sparser at the beginning of a life in the way that people were sparser early in human history.

---

[8] Responses of these sort are given by Benétreau-Dupin (forthcoming) and Greenberg (1999) respectively.

3. The quantum doomsday argument

Darren Bradley has recently remarked[9] that there is no analogue of the doomsday argument in the context of Everettian quantum mechanics. This is incorrect. On the standard *overlapping* interpretation of EQM, worlds are constantly branching: that is, literally splitting into multiple copies of themselves. The corresponding re-fusion is suppressed by the temporally asymmetric process of decoherence, which produces branching of worlds towards the future but no branching towards the past[10]. This temporal asymmetry of decoherence leads to a disparity in the number of *branches*, or *Everett worlds*[11], between early times and later times. Most world-stages are closer to the end of the world of which they're a part than to its beginning. There is a corresponding disparity in the number of inhabitants of worlds between early times and later times. Most agents are closer to the end of the population of which they're a part than to its beginning.

The quantum doomsday argument proceeds exactly on the model of the original doomsday argument. We can consider analogous population scenarios:

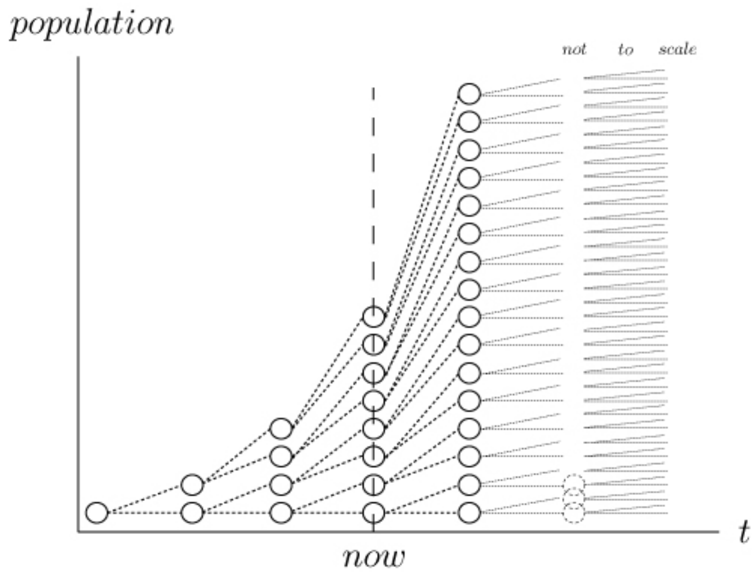

Fig 3. Branching Doom Soon: Total population will be just over twice the total to now.

[9] Bradley (2011), p.15.

[10] The ultimate source of the temporal asymmetry of decoherence presumably lies in a low-entanglement initial quantum state of the universe. This will not matter for our purposes.

[11] I use these terms interchangeably, in the spirit of the 'many worlds interpretation'.

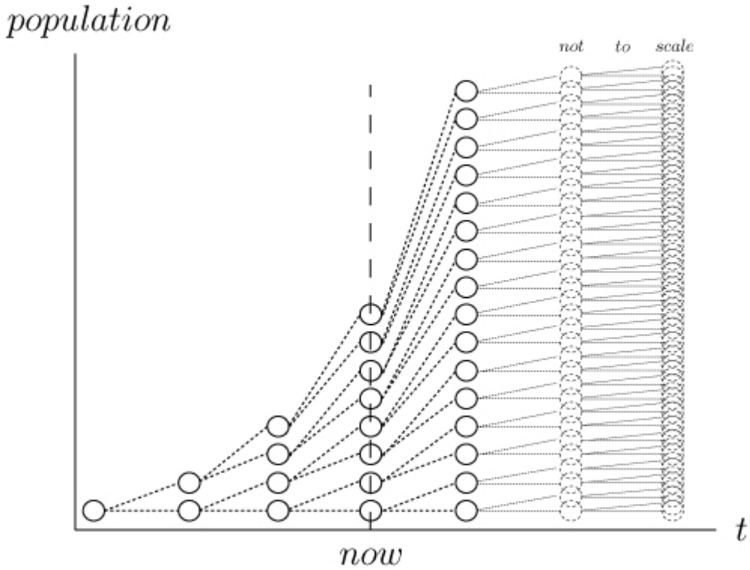

Fig 4. Branching Doom Later: Total population will be just over seven times the total to now.

The evidence that you have lived before the present confirms Branching Doom Soon over Branching Doom Later. That is, conditionalizing on the truth of EQM under a branching interpretation greatly increases the size of the doomsday shift. I will refer to this magnified doomsday shift as the 'quantum doomsday shift'. In our simplified setup, a quantum doomsday shift arises merely from the discovery that we are alive now:

P(LiveByNow|BranchingDoomSoon) ≈ 1/2

P(LiveByNow|BranchingDoomLater) ≈ 1/7

The more branching, the larger the quantum doomsday shift. Since branch number grows exponentially, the size of the shift is gigantic. For example, if every branch splits into two once per second, the number of branches is 2^n after n seconds. If the branching is into more than two branches, the disparity becomes even more striking, since branch number grows exponentially. In the bifurcating case, after n seconds the number of branch-temporal-parts in existence, 2^n, is greater than the total number to have existed up to that point (2^{n-1}). And if t_2 is 10 branching-events after t_1, then there are 2^{10} more branches at t_2 than there are at t_1.

Moreover, branching is utterly ubiquitous. Decoherence occurs on timescales of 10^{-20} seconds: and the branching is into countless branches. (Even if branch number is in some sense or other indeterminate, it's determinately gigantic; see §5 for further discussion.) So the quantum doomsday shift is going to be overwhelmingly larger in size than the ordinary doomsday shift.

The original doomsday argument runs as follows. Traditional estimates of the likely end of human civilization neglect the doomsday shift; but the doomsday shift is an unavoidable consequence of probability theory combined with empirical facts concerning population distribution; so traditional estimates of the likely end of human civilization are far too optimistic. We can construct a parallel quantum doomsday argument as follows. Previous estimates of the likely end of human civilization – even those which take into account the doomsday shift – neglect the quantum doomsday shift; but the quantum doomsday shift is an unavoidable consequence of probability theory combined with the hypothesis that EQM is correct; so, if EQM is correct, then previous estimates of the likely end of human civilization – even those which take into account the doomsday shift – are far too optimistic.

As with the original doomsday argument, the quantum doomsday argument comes in a personal version as well as a social version. The social version of the quantum doomsday argument considers birth rank of persons: the personal version considers 'life rank' of person-stages. Taking into account the rapid timescale of decoherence, the social version of the argument concludes that we should expect society to end almost immediately and the personal version concludes that we should expect our own lives to end almost immediately. Needless to say, neither of these conclusions is palatable.

A few quick clarifications and caveats:

i) Contemporary Everettians often deny that branch number is meaningful or well-defined. We will return to this point in §5, but for the sake of the argument I will assume that we can make sense of branch number.

ii) I am assuming that there is no 'quantum suicide' effect leading a subject to expect only survival when facing a risky upcoming quantum event. In my view, Papineau (2003) persuasively undermines the argument for this effect; see also Wallace (2012) §10.2.2.

iii) Since EQM is deterministic at the fundamental level, an agent who knows the whole quantum state can in principle know all the facts about the future evolution of the entire multiverse. The quantum doomsday argument does not apply to such agents, just as the regular doomsday argument does not apply to agents who know when doom will come.

The quantum doomsday argument shows that the hypothesis that EQM is correct has radical and deeply unpalatable epistemic consequences. So, how should a defender of overlapping EQM respond to the quantum doomsday argument? One obvious option is to adopt the self-indication assumption. As described in §2 above, prior credences that are in accordance with SIA can be used to cancel the doomsday effect; and this applies to the quantum doomsday effect just as it applies to the original doomsday effect. More populous multiverses are *a priori* more likely than less populous ones, according to their relative populations; allowing both for SIA and for a quantum doomsday shift leaves credences back where they started.

However, there are a number of reasons to be unhappy with the resulting theoretical position. Firstly, SIA suffers from objections like Bostrom's 'presumptious philosopher': the number of observers predicted just doesn't seem epistemically relevant to theory-choice in the way that SIA specifies. Secondly, the position threatens to lead to problematic 'automatic' confirmation of many-world theories such as EQM over single-world theories such as Bohmian mechanics. (After all, multiverses are vastly more populous than individual universes.) Thirdly, the position leads to disaster in light of the availability of a non-overlapping interpretation of EQM. That is the topic of the next section.

4. Doomsday and divergence

Recent work on the metaphysics of EQM (Saunders 2010, Wilson 2012) has applied David Lewis' distinction (Lewis 1986) between divergence and overlap. The usual interpretation of the decoherent histories formalism has it representing an ontology according to which histories have initial segments in common. But the formalism can equally be interpreted as representing a set of histories that are mereologically distinct from one another. Pictorially:

Fig 5. Overlap & divergence

For Saunders and Wilson, making sense of future-directed probabilities is the main motivation for a diverging understanding of EQM. Divergence allows for a simple bivalent semantics that can apply to both past-directed and future-directed thought and talk. It does so by providing a self-locating, or *indexical*, content for all contingent propositions, including those about the future. (See Saunders 2010, Wilson 2012, 2013.) David Deutsch also endorses an approach ('fungible worlds') that is closely related to divergence (Deutsch 2011); and influential recent work by Anthony Aguirre and Max Tegmark (Aguirre & Tegmark 2011) presupposes a diverging picture[12].

What I think is striking is that the diverging version of EQM completely undermines the quantum doomsday argument. (I will focus on the personal version of the argument; the same point applies to the social version.) If diverging EQM is correct, then there are just as many early temporal parts of a given person as there are late temporal parts:

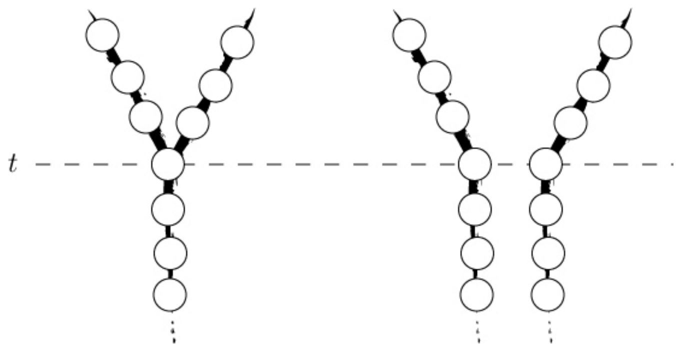

Fig 6. Populations given overlap & given divergence.

[12] Anthony Aguirre has confirmed (p.c.) that the proposal in Aguirre & Tegmark (2011) is intended to be understood on the model of divergence.

The quantum doomsday argument applies to overlapping EQM, but not to diverging EQM. Hence a differential overall doomsday effect arises: we ought to expect to live much longer under the diverging version of EQM than we would expect to live under the overlapping version of EQM.

This is all rather puzzling; and the puzzle can be sharpened into a troubling dilemma for the friend of overlap. According to overlapping EQM, there are vastly fewer people in existence than there are according to diverging EQM. SIA instructs us to weight the prior probability of a hypothesis by the number of people whose existence it entails. Accordingly:

> **Horn 1:** If SIA is correct, diverging EQM has a much greater prior probability than overlapping EQM.

However, avoiding this horn by rejecting SIA leaves us facing the quantum doomsday effect. Overlapping EQM leads to a runaway increase in the number of worlds over time, while diverging EQM does not. This increase in the number of worlds gives rise to the quantum doomsday effect.

> **Horn 2:** If SIA is incorrect, overlapping EQM leads to an overwhelming quantum doomsday effect, while diverging EQM does not.

It seems safe to assume that generating an overwhelming quantum doomsday effect makes an overlapping version of EQM untenable (horn 2). But avoiding the quantum doomsday effect by appeal to SIA leads us to give divergence much greater prior probability than we give overlap (horn 1). So overlapping EQM faces a serious dilemma. Whether or not SIA is correct, overlap is untenable and divergence is the only viable metaphysic for EQM.

The contrast between overlap and divergence looks paradigmatically metaphysical; indeed, Wallace is sceptical of the coherence of the distinction on this very basis. It is therefore rather surprising that it should have the dramatic epistemic consequences highlighted by the dilemma just presented. These consequences cast doubt on Wallace's claim that the distinction is 'fairly contentless' (Wallace 2012 p.287): if the reasoning above can be sustained, then divergence and overlap appear to give very different verdicts about the rational epistemic state of an agent who comes to believe EQM.

The relation between metaphysics and epistemology is subtle and complex, and we should expect there to be some epistemological constraints on physical theories. But the dilemma for overlap arising from the quantum doomsday effect really does look suspiciously quick: the falsity of a metaphysical doctrine like overlap should not be derivable simply from structural features of our confirmation theory. And in fact, the reasoning *is* too quick; there is no easy route to divergence from consideration of the quantum doomsday effect. Understanding why will require a more detailed look at Everettian probability.

Authors including David Deutsch, David Wallace, Simon Saunders and Hilary Greaves have recently developed complex and sophisticated treatments of probability in EQM. In the next section (§5), I'll explain how these authors handle problems like the quantum doomsday argument independently of the question of divergence vs. overlap by providing an Everettian-friendly decision-theoretic framework. Such treatments do not permit indifference reasoning to be applied across Everett worlds in such a way as to generate a quantum doomsday effect, so the argument for divergence based on the quantum doomsday effect lapses. However, the justification for restricting indifference reasoning in this way remains controversial. In §6, I'll describe a novel approach that provides a clear rationale for the needed restriction on the principle of indifference. The key move is to adopt a naturalistic metaphysics of modality, construing EQM as a form of modal realism.

5. Solutions to the quantitative problem

Probability has frequently been targeted by critics of EQM. The 'probability problem' for Everettians can be divided into three sub-problems:

- ***The Incoherence Problem:*** How does probability even make sense in a multiverse theory like EQM? – what are the probabilities probabilities *of* ?
- ***The Quantitative Problem:*** Given EQM, why should probabilities of outcomes be given by the Born Rule?
- ***The Epistemic Problem:*** Given EQM, how does our ordinary statistical evidence count as evidence in favour of EQM?

The problems are not independent; for example, a solution to the incoherence problem is arguably a precondition for a solution to the quantitative problem and a solution to the quantitative problem is arguably a precondition for a solution to the epistemic problem. But for present purposes we can focus on the quantitative problem. A solution to the quantitative problem would suffice to dissolve the argument for divergence based on the quantum doomsday effect, since the Born Rule probabilities – or *branch weights* – incorporate no quantum doomsday effect.

A number of solutions to the quantitative problem have been proposed. Everett himself gave frequentist-style arguments for the Born Rule, and appealed to symmetry properties of the quantum state; other solutions have proceeded via invoking additional physical structure (Deutsch 1985; Albert & Loewer 1988), via a direct theoretical identification of branch weights with objective probabilities (Saunders 1998), or via decision theory. I will focus on the decision-theoretic approaches, as they have been the subject of vigorous recent discussion.

Decision-theoretic solutions to the quantitative problem derive originally from pioneering work by Deutsch (1999). Over the next decade these arguments were refined and strengthened, resulting in the systematic decision-theoretic treatments of Everettian probability given by Wallace (2003, 2006, 2007, 2010, 2012), Saunders (2005), Greaves (2004, 2007) and Greaves and Myrvold (2010). These authors (henceforth: WSGM) offer and defend sets of decision-theoretic axioms which give rise to decision theories appropriate to Everettian agents. In these decision theories, branch weights play the functional role of objective chances. The WSGM decision-theoretic treatments entail that rational epistemic agents should set their prior credences (conditional on the truth of EQM) in being located in a given Everett world equal to the branch weight of that world – an Everettian form of what David Lewis called the Principal Principle[13].

The WSGM decision-theoretic treatments differ in a number of details, but they all incorporate restrictions on what a rational agent can in principle define their preferences over. To adapt terminology from Savage, the WSGM treatments identify the *decision-theoretic state* with individual Everett worlds

[13] See Lewis (1980/1986).

rather than with entire multiverses[14]. The decision-theoretic state is ordinarily thought of as representing the agent and the whole of their environment: every decision the agent makes is rationally assessable only with respect to the consequences the decision has for the agent's decision-theoretic state.

The kind of self-locating uncertainty which drives the quantum doomsday argument is uncertainty about where in a branching multiverse one is located. Uncertainty of this sort incorporates both uncertainty about which Everett world the agent is in and uncertainty about which centre an agent occupies within their Everett world. This distinction matters a great deal, because WSGM handle these latter two kinds of uncertainty very differently. Uncertainty about which Everett world an agent is in is uncertainty about what their decision-theoretic state is like. Uncertainty about which centre an agent occupies within an Everett world remains even when the agent's decision-theoretic state is fully specified. As they have been developed thus far, the WSGM treatments do not take account of the possibility of self-locating uncertainty within a single Everett world, but any extension of the treatments to account for this possibility is constrained by the need to assign credences to Everett worlds according to the Born rule.

Through their identification of decision-theoretic states with Everett worlds, the WSGM decision-theoretic treatments effectively factor uncertainty about self-location within a branching multiverse into two components: self-locating uncertainty within a single Everett world, and uncertainty about which Everett world one is in. The treatments impose the following constraint (which corresponds to the Born rule) on credences about which Everett world one is in: these credences, conditional on the truth of EQM, should match the weights assigned to each Everett world. However, they place no special constraints on self-locating uncertainty within a single Everett world. Consequently, the WSGM treatments leave open the possibility of applying something like Elga's restricted indifference principle within any specific Everett world, but they rule out the possibility of applying it across Everett worlds. While indifference reasoning turns

[14] This move is made, more or less explicitly, by Wallace (e.g. Wallace 2012 p.163-166), by Saunders (Saunders 2005 p. 229-230), by Greaves (Greaves 2007 p.128-9) and by Greaves & Myrvold (2010 p.296-8). See also (Greaves 2007 p.147).

on relative numbers of outcomes, branch weights are insensitive to fine-grainings of the history space, which means that altering the number of worlds on which an outcome occurs will not affect the probability of that outcome[15].

In the WSGM decision-theoretic frameworks the indifference principles driving the quantum doomsday argument cannot be applied across different branches of the multiverse, so the argument for the quantum doomsday shift breaks down. Since the WSGM decision theoretic treatments are neutral between divergence and overlap, this response to the quantum doomsday argument dissolves the dilemma of §4. Meanwhile, the original doomsday argument is unaffected. That argument, transposed to the Everettian setting, involves only self-locating uncertainty within a single Everett world, so the indifference principle that it involves does not clash with the Born rule. Nothing in the WSGM decision-theoretic treatments undermines Elga-style indifference reasoning when it is confined only to in-branch applications.

These results generalize beyond decision-theoretic solutions to the quantitative problem. Setting credences about which Everett world one is in according to the Born Rule is incompatible with respecting the indifference principle that drives the quantum doomsday argument. Accordingly, if *any* Everettian solution to the quantitative problem can be sustained, then the indifference principle driving the quantum doomsday effect is false and the argument for divergence based on the quantum doomsday shift is unsound. If no Everettian solution to the quantitative problem can be sustained, then EQM is in any case untenable. So the quantum doomsday effect cannot after all provide good grounds for adopting diverging EQM.

Two problems remain. How can Everettians explain away the apparent plausibility of the indifference reasoning that drives the quantum doomsday argument? And how can they justify continuing to apply indifference reasoning in the more familiar contexts that initially motivated Elga's restricted principle of indifference? These questions will be addressed in the next section.

---

[15] This holds true as long as the relevant decoherence conditions are satisfied; see Wallace (2012) §4.3.

6. Many worlds and modal realism

When doing decision theory in the classical one-world setting, in order to adequately account for self-locating uncertainty we need to treat the question of where in the world we are differently from the question of what the world is like. Likewise, if EQM is correct, in order to adequately account for self-locating uncertainty we need to treat the question of which Everett world we're in differently from the question of where we are in that Everett world. §5 showed that we can achieve this independently of the question of overlap vs. divergence, so long as we restrict indifference reasoning from applying across distinct Everett words. But it seems rather unsatisfying, and somewhat *ad hoc*, to simply stipulate such a restriction when setting out an Everettian decision theory. Do we have any independent reason to think that the resulting decision theory correctly models agents in an Everettian context?

I have a suggestion. A judicious modification to principles connecting the physics of EQM with the metaphysics of modality results in a metaphysical picture from which a WSGM-style decision theory follows naturally. Elsewhere I've proposed the following principle:

> ***Individualism***: Distinct Everett worlds comprise alternative metaphysical possibilities.

According to Individualism, each Everett world is to be identified with a different metaphysically possible world. This identification fits naturally with an *indexical* conception of actuality for Everett worlds: the actual Everett world is the one in which we ourselves are located. The resulting picture is a form of modal realism, in the style of David Lewis (1986). While an Everettian modal realism would differ substantially from Lewisian modal realism – the possible worlds of Everettian modal realism would be emergent entities, would all share the same fundamental laws of nature, and might be neither spatio-temporally or causally isolated from one another[16] – the views offer similarly attractive reductive stories

[16] Whether Everett worlds are spatio-temporally isolated depends on whether we adopt a diverging or overlapping interpretation of the quantum formalism; both interpretations are compatible with Individualism and with indexical actuality. It is also to possible to interpret EQM so that Everett worlds are causally isolated from one another, regarding the

about the nature of modality. However, my proposed solution to the quantum doomsday problem does not force us to regard EQM as a fully general theory of modality. The solution involves treating individual Everett worlds as alternative metaphysical possibilities; we can (but need not) allow in addition for the metaphysical possibility of some non-Everettian scenarios[17].

As I have argued elsewhere, the combination of Individualism with an indexical conception of actuality helps to underwrite the WSGM decision-theoretic treatments. Everettians can retain Huw Price's compelling principle that "where goes ontology, there goes possible preference"[18] (Price 2010 p.380): they need only interpret the principle with an implicit quantifier-restriction to the ontology of the actual world. For Everettians who adopt an indexical view of actuality, such quantifier-restriction is endemic in ordinary talk and unmysterious. Likewise, the indexical conception of actuality allows us to retain the related principle that 'where goes actuality, there goes the epistemic possibility of self-location'. If 'actuality' refers indexically to an agent's own Everett world, then this principle is vindicated by the WSGM decision-theoretic frameworks. Uncertainty about one's location within an Everett world is treated as uncertainty about where in the actual world one is located, while uncertainty about which Everett world one is in is treated as uncertainty about ordinary contingent matters of fact.

Once we understand EQM along modal realist lines, Elga's indifference principle still gives rise to the original doomsday effect, but it gives rise to no

---

dynamical connections between worlds as giving rise to non-causal explanations. (Perhaps phenomena in each Everett world are explained non-causally by phenomena in nearby worlds, in the same sort of way that the shape of a jigsaw piece is explained by the shapes of the surrounding pieces.) Exploring these issues must await another occasion.

[17] Adopting Individualism provides some pressure towards a principle that I have elsewhere called Alignment: that to be a metaphysically possible world is to be an Everett world. Alignment involves some controversial commitments: in particular, it entails the metaphysical necessity of the fundamental laws of nature. Although I find this stronger view plausible, the arguments of this paper do not depend on it.

[18] To provide some context for this quote: Price is here questioning the restriction in the WGS decision theory of an agent's preferences to be defined only over states of the Everett world of which they are a part, rather than being defined over states of the whole Everett multiverse of which they are a part. His reasoning is that, in principle, we may assign value to anything at all – so long as it genuinely exists.

quantum doomsday effect[19]. If individual Everett worlds are possible worlds, then Elga's indifference principle, without any need for *ad hoc* modification, already tells an agent only about how to divide credence between various locations within a single Everett world. In other words, a modal realist reading of EQM makes it possible to solve the quantitative problem with Everettian probability without modifying a plausible principle of classical self-locating epistemology.

The Everettian modal realist rebuttal of the quantum doomsday argument applies equally to the overlapping picture and the diverging picture, and it thereby dissolves any methodological puzzlement elicited by the argument for divergence based on the quantum doomsday shift. The dialectical situation concerning the doomsday argument is rendered independent of the question of divergence vs. overlap, and confirmation theory by itself is no longer settling any metaphysical questions. I take this to be a very desirable result.

7. Conclusion

The quantum doomsday argument threatens to draw crazy confirmational consequences from Everettian quantum mechanics understood according to the overlap picture. These consequences can be avoided by switching to a diverging picture, but this gives rise to a further puzzle: how can a metaphysical dispute about the ontology of Everett worlds give rise to such a radical epistemological difference?

We can now see how to resolve these puzzles. Everettian solutions to the quantitative problem undermine the quantum doomsday argument and thereby dissolve the epistemological difference between divergence and overlap. In addition, Everettians who supplement their solution to the quantitative problem with a modal realist construal of the multiverse have a satisfying response to the argument for divergence based on the quantum doomsday shift: they can apply Elga's indifference principle in a way that avoids the quantum doomsday effect while recovering indifference reasoning in cases where it is intuitively plausible.

[19] The indexical conception of actuality also does not affect more familiar applications of Elga's indifference principle: the principle still applies in the usual way to the Sleeping Beauty puzzle (Elga 2000) and to Elga's duplication puzzle (Elga 2004).

The treatment of Everettian probability that I have proposed involves a naturalistic metaphysics of modality. This approach obviously makes modal metaphysics hostage to empirical fortune. If it turns out that EQM is incorrect (perhaps some dynamical-collapse alternative to quantum theory will be experimentally confirmed) then we will cease to have reason to think that Everett worlds play the possible-world theoretical role with respect to self-locating indifference reasoning. This will not appeal to those metaphysicians who think that demarcation of modal space is philosophers' work, and should be done *a priori*. But it ought to appeal to the growing number of more naturalistically-oriented metaphysicians who take empirical work to be relevant to – for example – the metaphysics of space, time, and substance[20].

[20] This research was supported by an Australian Research Council grant, "Neglected Problems of Time", hosted at Monash University, and it was completed during a visiting fellowship at the Centre for Time at the University of Sydney. I am grateful for conversations about this material with Anthony Aguirre, Adam Bales, Sam Baron, Yann Benétreau-Dupin, Rachael Briggs, Adam Caulton, David Deutsch, Rohan French, Toby Handfield, Nick Jones, Chris Meacham, Dave Ripley, Simon Saunders, and David Wallace, and to audiences at Monash, Birmingham, Leeds and the BSPS. My thanks also go to three anonymous referees for helpful feedback.

References

Aguirre, A. & Tegmark, M. (2011). 'Born in an Infinite Universe: a Cosmological Interpretation of Quantum Mechanics', *Physical Review D* 84(10): 105002.

Albert, D. Z. and Loewer, B., 1988, 'Interpreting the Many Worlds Interpretation', *Synthèse* 77: 195–213.

Barrett, J. A. (1999). *The Quantum Mechanics of Minds and Worlds.* New York: Oxford University Press.

Bartha, P. & Hitchcock, C. (1999). 'No-one knows the date or hour: an unorthodox application of Rev. Bayes' theorem', *Philosophy of Science (Procs.)* 66: S339-53.

Benétreau-Dupin, Y. (forthcoming). Blurring Out Cosmic Puzzles. *Philosophy of Science.*

Bostrom, N. (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy.* New York: Routledge.

Bradley, D. (2011). 'Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty', *The British Journal for the Philosophy of Science* 62(2): 323-342.

Bradley, D. and Fitelson, B. (2003). 'Monty Hall, Doomsday, and Confirmation', *Analysis* 63(1): 23-31.

Carter, B. (1983). 'The anthropic principle and its implications for biological evolution', *Phil. Trans. Roy. Soc. Lond.* A310: 347-363.

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

Deutsch, D (1999). 'Quantum theory of probability and decisions', *Proceedings of the Royal Society of London* A455: 3129-3137.

Deutsch, D. (2011). *The Beginning of Infinity.* Allen Lane.

Elga, A. (2000). 'Self-locating belief and the Sleeping Beauty problem', *Analysis* 60: 143-147.

Elga, A. (2004). 'Defeating Dr. Evil with self-locating belief', *Philosophy and Phenomenological Research* 69: 383–396.

Gott, J. R. (1993). 'Implications of the Copernican Principle for our Future Prospects', *Nature* 363: 315-9.

Graham, R. N. (1973). 'The Measurement of Relative Frequency', in B. DeWitt & N. Graham (eds.), *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton: Princeton University Press.

Greenberg, M. (1999) 'Apocalypse Not Just Now', *London Review of Books* (1 July): 19-22.

Greaves, H. (2004). 'Understanding Deutsch's probability in a deterministic multiverse'. *Studies in History and Philosophy of Modern Physics* 35, 423-56.

Greaves, H. (2007a). 'On the Everettian Epistemic Problem', *Studies in the History and Philosophy of Modern Physics* 38(1): 120-152.

Greaves, H. & Myrvold, W. (2010). 'Everett and Evidence', in S. Saunders, J. Barrett, A. Kent & D. Wallace (eds.), *Many Worlds? Everett, Quantum Theory, and Reality* (pp. 181-205). Oxford: Oxford University Press.

Leslie, J. (1989). *Universes*. London: Routledge.

Lewis, D. K. (1980/1986). 'A Subjectivist's Guide to Objective Chance', in ed. Jeffrey, *Studies in Inductive Logic and Probability*. Berkeley: University of California Press. Reprinted with postscripts in D.K. Lewis, *Philosophical Papers Vol. II*. New York: Oxford University Press.

Lewis, D. K. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Monton, B. (2003). 'The Doomsday Argument without Knowledge of Birth Rank', *The Philosophical Quarterly* 53(210): 79-82.

Papineau, D. (2003). 'Why you don't want to get in the box with Schrodinger's Cat', *Analysis* 63(277): 51-58.

Price, H. (2010). 'Decisions, decisions, decisions: Can Savage Salvage Everettian Probability?', in S. Saunders, J. Barrett, A. Kent & D. Wallace (eds.), *Many Worlds? Everett, Quantum Theory, and Reality* (pp. 181-205). Oxford: Oxford University Press.

Saunders, S. (1998). 'Time, Quantum Mechanics, and Probability', *Synthèse* 114: 373-404.

Saunders, S. (2010). 'Chance in the Everett Interpretation', in S. Saunders, J. Barrett, A. Kent & D. Wallace (eds.), *Many Worlds? Everett, Quantum Theory, and Reality* (pp. 181-205). Oxford: Oxford University Press.

Savage, L. J. (1972). *The Foundations of Statistics* (2nd ed.). Dover.

Sober, E. (2002). 'An Empirical Critique of Two Versions of the Doomsday Argument - Gott's Line and Leslie's Wedge', *Synthèse* 135: 415-430.

Wallace, D. (2003). 'Everettian Rationality: defending Deutsch's approach to probability in the Everett interpretation'. *Studies in the History and Philosophy of Modern Physics* 34: 415-438.

Wallace, D. (2007) 'Quantum Probability from Subjective Likelihood: improving on Deutsch's proof of the probability rule', *Studies in the History and Philosophy of Modern Physics*38 (2007), pp. 311--332.

Wallace, D. (2006). 'Epistemology Quantised: circumstances in which we should come to believe in the Everett interpretation', *British Journal for the Philosophy of Science* 57: 655-689.

Wallace, D. (2010). 'How to Prove the Born Rule', in eds. Saunders *et al.*, *Many Worlds? Everett, Quantum Theory, and Reality*. Oxford: Oxford University Press.

Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press.

Wilson, A. (2012). 'Everettian Quantum Mechanics Without Branching Time', *Synthèse* 188(1): 67-86.

Wilson, A. (2013). 'Objective Probability in Everettian Quantum Mechanics'. *The British Journal for Philosophy of Science* 64(4): 709-737.