Killer Collapse
Empirically Probing the Philosophically
Unsatisfactory Region of GRW

Charles T. Sebens
University of Michigan, Department of Philosophy

Abstract

GRW theory offers precise laws for the collapse of the wave function. These collapses are characterized by two new constants, λ and σ. Recent work has put experimental upper bounds on the collapse rate, λ. Lower bounds on λ have been more controversial since GRW begins to take on a many-worlds character for small values of λ. Here I examine GRW in this odd region of parameter space where collapse events act as natural disasters that destroy branches of the wave function along with their occupants. Our continued survival provides evidence that we don't live in a universe like that. I offer a quantitative analysis of how such evidence can be used to assess versions of GRW with small collapse rates in an effort to move towards more principled and experimentally-informed lower bounds for λ.

1 Introduction

One central point of disagreement in the foundations of quantum mechanics is whether the collapse of the wave function is a genuine physical process. If collapse is to be taken seriously, we should seek to determine physical laws that might govern this process. Ghirardi-Rimini-Weber theory (GRW) offers possible precise laws which guarantee that the wave function collapses during familiar quantum measurements. However, observers and measurements have no special status in the theory; collapses happen whether or not scientists are watching.

The laws of GRW include two new fundamental constants not present in textbook discussions of quantum mechanics. One parameter, σ, characterizes the precision of the collapse events and the other, λ, the rate at which collapses occur. If these parameters are chosen properly, the theory appears to succeed in generating the correct probabilistic predictions for experiments taken to be within the purview of non-relativistic quantum mechanics. However, as more experiments are conducted we continue to shrink the space of possible values for σ and λ. Potentially, the allowed region could shrink so much it disappears and GRW could be ruled out. Alternatively, new experiments might confirm GRW over its competitors. As of now, there seems to be a fair amount of leeway as to what values we may assign to the parameters (figure 1). Focus on the collapse rate λ. It is fairly well-understood how we can put experimental *upper* bounds on the collapse rate. If collapse events were too frequent, interference patterns would be destroyed by particles collapsing mid-experiment, isolated systems would heat up, undisturbed atoms would

spontaneously emit photons, and in other varied ways the experimental predictions of the theory would be corrupted (these constraints have been reviewed recently in Adler, 2007; Feldmann & Tumulka, 2012; Bassi *et al.*, 2013).

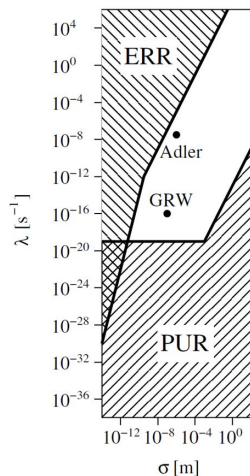Figure 1: Parameter diagram of GRW theory from Feldmann & Tumulka (2012). ERR is the "empirically refuted region." PUR is the "philosophically unsatisfactory region." The points labeled "GRW" and "Adler" indicate the values suggested in Ghirardi *et al.* (1986) and Adler (2007) respectively. It should be noted that Adler's proposal was made in the context of CSL, not GRW.

In this article, I would like to explore how we might put experimental *lower* bounds on the collapse rate λ. The trend in the literature has been to dismiss low values of λ for non-empirical reasons or for reasons that presuppose the failure of the many-worlds interpretation. When λ is very small GRW becomes an odd theory. Macroscopic objects are not prevented from entering superpositions and the theory takes on a many-worlds character (§3). Such versions of GRW have been rejected as philosophically unsatisfactory. Surely they are. But, there has been disagreement about exactly where the problems arise. Feldmann & Tumulka (2012) give the criterion, "We regard a parameter choice (σ, λ) as philosophically satisfactory if and only if the PO [primitive ontology] agrees on the macroscopic scale with what humans normally think macroscopic reality is like." Bassi *et al.* (2010) impose the requirement that "any superposition reaching the eye must be reduced before it is transformed into a perception in the brain.", building on a suggestion in Aicardi *et al.* (1991). Adler (2007) and Gisin & Percival (1993) argue that the formation of a microscopic latent image in a detector counts as a measurement even before this image is amplified to macroscopic scale. They believe that the collapse rate must be high enough that even these latent images do not enter superpositions.

I will argue that very small values of λ are not just *philosophically* problematic, they are *empirically* unacceptable *even if* the many-worlds interpretation is viable. In doing so, I hope to begin shifting the burden from philosophical considerations to empirical ones and to lay the foundation for a principled and experimentally informed approach to determining lower bounds on λ. Although the paper will focus on GRW throughout, many of the lessons could be applied *mutatis mutandis* to other collapse theories.

2   GRW Theory

In GRW theory, the evolution of the wave function is typically governed by the familiar Schrödinger equation,

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = \widehat{H} |\Psi(t)\rangle \ .$$

(2.1)

At some instants, the evolution of the wave function is discontinuous and not in accord with the Schrödinger equation. The wave function collapses. According GRW, collapse is a real physical process governed by well-defined laws and occurring frequently, not just during measurements. Humans and other observers play no spooky role, they are just particularly intelligent and perceptive collections of particles.

When a collapse occurs a randomly chosen particle has its position become extremely well-localized. Collapses occur randomly at a rate of $N\lambda$ where N is the total number of particles. That is, once a collapse occurs at T_1 the probability that the next collapse, at T_2, will happen within time interval Δt is given by

$$P(T_2 - T_1 < \Delta t) = 1 - e^{-N\lambda\Delta t} \ . \tag{2.2}$$

The collapse rate λ is one of two new constants of the theory, originally suggested to be on the order of $10^{-16}\mathrm{s}^{-1}$ (Ghirardi *et al.* , 1986).[1] The collapse localizes particle I (randomly chosen) around location $\boldsymbol{X}$, where $\boldsymbol{X}$ is chosen randomly with probability density

$$\rho_I(\boldsymbol{x}) = \lim_{t \nearrow T} \langle \Psi(t)| \Lambda_I(\boldsymbol{x}) |\Psi(t)\rangle \ . \tag{2.3}$$

"$\lim_{t \nearrow T}$" denotes the limit as t approaches the time of collapse, T, from below. $\Lambda_i(\boldsymbol{x})$ is the collapse operator defined by

$$\Lambda_i(\boldsymbol{x}) = \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-\frac{(\widehat{\boldsymbol{x}}_i - \boldsymbol{x})^2}{2\sigma^2}} \ , \tag{2.4}$$

where $\widehat{\boldsymbol{x}}_i$ is the position operator for particle i. The wave function after the collapse is given by the pre-collapse wave function multiplied by a tightly peaked three-dimensional Gaussian centered about $\boldsymbol{X}$ and normalized,

$$\lim_{t \searrow T} |\Psi(t)\rangle = \lim_{t \nearrow T} \frac{\Lambda_I(\boldsymbol{X})^{1/2} |\Psi(t)\rangle}{\langle \Psi(t)| \Lambda_I(\boldsymbol{X}) |\Psi(t)\rangle^{1/2}} \ . \tag{2.5}$$

The second new constant in GRW, σ, appears in (2.4) and characterizes the width of the Gaussian that localizes the particle. It was originally proposed to be on the order of $10^{-7}\mathrm{m}$ (Ghirardi *et al.* , 1986). In the remainder of the paper different values of λ will be considered, but σ will be kept fixed at about $10^{-7}\mathrm{m}$.

In the simplest version of GRW, GRW0, the wave function is all there is and its evolution is determined by the Schrödinger equation (2.1) and the collapse process (2.2, 2.3, 2.5). In the limit where λ is taken to zero, collapse never occurs and GRW0 becomes Everettian quantum mechanics (a.k.a. the many-worlds interpretation or S0). All there is is the wave function and it always evolves in accordance with the Schrödinger equation. Defenders of Everettian quantum mechanics tend to view GRW0 as the right way to think about GRW theory since they think that our experiences of reality can emerge from patterns in wave functions. For Everettians and others who prefer GRW0 to the alternatives below, this paper can be read as a discussion of GRW0 in the strange regime where it approaches Everettian quantum mechanics.

For some, GRW0 is unsatisfactory (e.g., Allori *et al.* , 2008, §4.3; Maudlin, 2010).

---

[1]It has been suggested that different particles might collapse at different rates depending on their masses (Pearle & Squires, 1994). The analysis presented here could be applied to such a formulation.

According to GRW0 there are no objects in familiar three-dimensional space, there is only a wave function in an abstract space: a vector in Hilbert space, a complex-valued function on configuration space, or some other exotic beast. In GRWm, the universe contains a wave function which obeys the above dynamics, but that's not all there is, and, in some sense, that's not the important stuff. In particular, it's not the stuff we're made of. In addition to the wave function, there also exists a distribution of matter in three-dimensional space specified by a density,

$$m(\boldsymbol{x}, t) = \langle \Psi(t) | \, \widehat{M}(\boldsymbol{x}) \, | \Psi(t) \rangle \ . \tag{2.6}$$

Here $\widehat{M}(\boldsymbol{x})$ is the mass density operator defined by

$$\widehat{M}(\boldsymbol{x}) = \sum_{i=1}^{N} m_i \, \delta^3(\widehat{\boldsymbol{x}}_i - \boldsymbol{x}) \ . \tag{2.7}$$

In the limit as λ goes to zero, there is no collapse and GRWm becomes Sm, Schrödinger evolution with a mass density (discussed in Allori *et al.* , 2011). Sm is a many-worlds theory much like Everettian quantum mechanics, but where the universe contains a distribution of mass in three-dimensional space in addition to the unitarily evolving wave function. Some think that GRW0 and S0 are unsatisfactory because such laws would not give rise to creatures with conscious experiences like ours, perceiving an apparently three-dimensional world. Readers who think GRW0 is unsatisfactory can understand this paper as a discussion of GRWm in the awkward bit of parameter space where it approaches Sm. In the following sections, I will not differentiate between GRW0 and GRWm. Read GRW in whichever way you think makes it the stronger theory. Read MWI as S0 if you're reading GRW as GRW0, as Sm if you're reading GRW as GRWm.

Perhaps neither S0 nor Sm really are many-worlds theories. Without collapse, one might argue, it's not that every outcome of a quantum experiment is observed by a separate copy of the experimenter but that the single experimenter somehow experiences all outcomes at once or otherwise ceases to have a normal mental state. If this is the problem with these theories, GRW will become empirically inadequate as it approaches S0 or Sm. However, determining when and how it becomes inadequate would require a specific account of the abnormality that should be expected in the absence of collapse. For the purposes of this article, I will assume that the theory GRW limits to as the collapse rate is taken to zero really is a many-worlds theory.

There is a third version of GRW, GRWf. Here one supplements the wave function with a primitive ontology of flashes. Taking λ to be small in this version of the theory raises entirely different concerns from those faced by GRW0 and GRWm. The problem for GRWf when λ is small is not that human lives are constantly ending, but that such life may be absent altogether. Understanding the empirical adequacy of GRWf in this region of parameter space would require a very different kind of analysis and for that reason GRWf will not be discussed in the remainder of the article. A brief discussion of GRWf in this regime can be found in Feldmann & Tumulka (2012, §4).

3 Branches and Stumps

GRW was originally formulated with the rate of collapse $\lambda \approx 10^{-16} \text{s}^{-1}$. With this rate, when a measurement occurs the wave function just starts to branch into a superposition

of outcomes when, with very high probability, the wave function collapses to a single definite outcome.[2] This is how GRW solves the measurement problem: a single definite outcome is guaranteed by the rapid collapse of the wave function and the fact that probabilities for collapsing to different outcomes are (approximately) given by the Born rule is a non-trivial consequence of the collapse process (2.2, 2.3, 2.5). If the rate of collapse is taken to zero, then collapses never occur and GRW becomes MWI. In MWI, every possible outcome of a quantum measurement actually occurs.

What if λ is chosen so that it is not quite zero, but is very small ($\lambda \ll 10^{-16}\text{s}^{-1}$, keeping $\sigma \approx 10^{-7}\text{m}^3$)? In this regime collapses occur, but only very rarely. When a collapse occurs, the results are catastrophic. After a spin measurement, the laboratory enters into a superposition of a world in which the scientists record an up result and another in which they record down. Later, if any of the particles that compose the scientists or the measurement readout collapse, one of the worlds will be destroyed. Imagine 15 minutes pass between the moment when the measurement occurred and the time when collapse chooses a world to eliminate.[4] In this time, the scientists in both worlds can walk, think, and talk. After collapse, only one world remains. When a collapse like this occurs, all of the inhabitants of the other world are instantaneously and painlessly killed. Or, maybe the collapse doesn't cause the other world to go out of existence, but instead the tail of the Gaussian distorts the world and alters its evolution so that it is inhospitable to human life.[5] In this case, death is fairly quick but perhaps not instantaneous. Either way, in this region of parameter space collapses are not helpful shifts which prevent macroscopic superpositions from forming, they're colossal natural disasters.

The way the universe—the totality of quantum worlds—evolves in each of these three regions of parameter space is depicted in figure 2. With λ at or near zero, worlds branch every time a measurement occurs and each outcome happens on some branch. For standard values of λ, branching is prevented by the collapse of the wave function and each measurement has a definite outcome. For small values of λ branching occurs before collapse is able to prevent it; collapse events occur after branching. Living in such a universe is extremely dangerous as entire worlds are constantly being obliterated. If you are lucky enough to find yourself living a long life, you should be shocked. Repeated improbable occurrences often indicate failure of a theory. This is no exception. The data you receive from your survival provides strong *empirical evidence* against the theory.

---

[2]There has been some debate over whether the destruction of other branches is successful; see the literature on the problem of tails. Here I assume that the problem can be solved. If it cannot, GRW is not a viable solution to the measurement problem. In particular, I will assume that if collapse chooses one part of the state and massively shrinks the rest, it is not merely improbable to find oneself in a part of the state that was not fortunate enough to be the center of the collapse, it is impossible. There is no life in those other parts soon after collapse.

[3]This ensures that, in general, a single collapse will be sufficient to destroy branches in which the measurement turned out differently.

[4]This would be typical if we choose λ to be on the order of 10^{-33}s^{-1} and assume that there are about 10^{30} fundamental particles brought into an entangled superposition by the experiment (using (2.2)).

[5]See Wallace (2014); Vaidman (2014b, §8) for "solutions" to the tails problem along these lines (also briefly discussed in Allori *et al.* (2011, §4)).
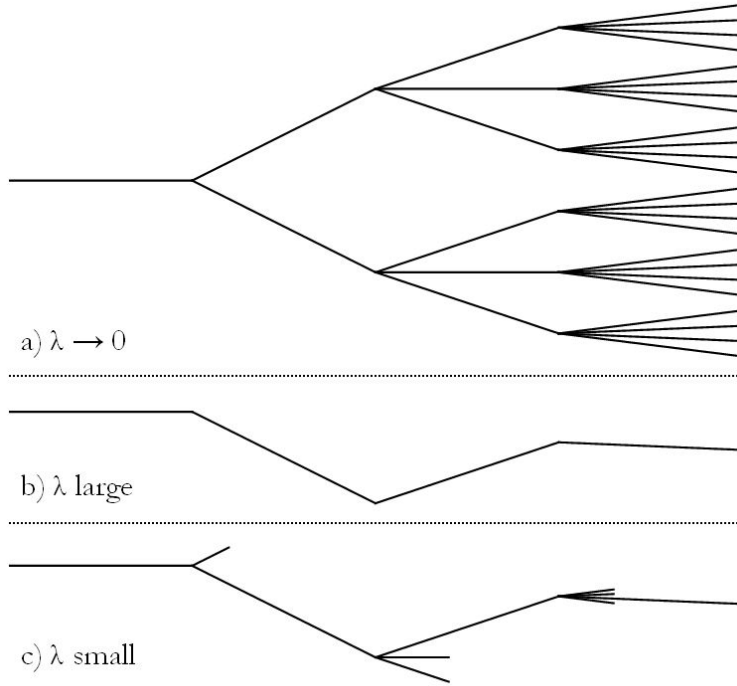
Figure 2: Plot of GRW evolution for a sequence of three measurements for different values of λ.

4 The Rarity of Longevity

To judge the empirical adequacy of a given theory, I will focus on the likelihood of the evidence given the theory, $P(\mathcal{E}|\mathcal{T})$. If, for some evidence $\mathcal{E}$ and theories $\mathcal{T}_1$ and $\mathcal{T}_2$, $P(\mathcal{E}|\mathcal{T}_1) > P(\mathcal{E}|\mathcal{T}_2)$, then the evidence $\mathcal{E}$ confirms $\mathcal{T}_1$ over $\mathcal{T}_2$. If one updates on $\mathcal{E}$ by Bayesian conditionalization, then for any theory $\mathcal{T}$, the credence assigned to $\mathcal{T}$ after gaining the evidence can be expressed in terms of the prior probabilities as $P_{post}(\mathcal{T}) = P(\mathcal{T}|\mathcal{E})$.[6] It follows from $P(\mathcal{E}|\mathcal{T}_1)$ being greater than $P(\mathcal{E}|\mathcal{T}_2)$ that, if one changes their credences in response to $\mathcal{E}$ by Bayesian updating, the ratio of one's credence in $\mathcal{T}_1$ to their credence in $\mathcal{T}_2$ will rise,

$$\frac{P_{post}(\mathcal{T}_1)}{P_{post}(\mathcal{T}_2)} = \frac{P\left(\mathcal{E}|\mathcal{T}_1\right)}{P\left(\mathcal{E}|\mathcal{T}_2\right)}\frac{P(\mathcal{T}_1)}{P(\mathcal{T}_2)} > \frac{P(\mathcal{T}_1)}{P(\mathcal{T}_2)} \tag{4.1}$$

---

[6]Although I expect that this straightforward account of theory confirmation applies to the cases under discussion, one might reasonably be concerned. The situations considered involve *self-locating uncertainty* (see Sebens & Carroll, 2014; Vaidman, 2014a, §4.2) and Bayesian conditionalization must be somehow modified to handle such cases (see Arntzenius, 2003). Some modifications will vindicate the use of conditionalization here, others will not. To avoid controversy, I focus primarily on the probability of the evidence given the theory and not the posterior probabilities that result from updating on the evidence. I approach the problem from the familiar diachronic perspective, taking one's previous beliefs and evidence to together determine what one's current beliefs should be. Alternatively the problem could be approached synchronically, taking one's evidence together with what Meacham (2010) calls an "epistemic kernel" to determine what one's current beliefs should be (there are several competing ways of implementing this approach; see Manley, 2014).

Theories that are empirically equivalent will assign the evidence equal probability and the data that comes in will not discern between them.

The theories which will be compared are: versions of GRW with different parameter values, e.g., $\text{GRW}_{\lambda=10^{-16}\text{s}^{-1}}$; the many-worlds interpretation, MWI; and some unspecified theory which gives the correct Born rule probabilities and guarantees survival, QM.[7] The constraint that QM gives the Born rule probabilities is the constraint that: the probability of seeing the outcome corresponding to eigenvalue O_i of the observable operator $\widehat{O}$ is given by

$$P(O_i|\text{QM}) = |\langle O_i|\Psi\rangle|^2 . \tag{4.2}$$

Throughout I'll assume that the agent knows whatever is useful to know about the universal wave function, Ψ, including $|\langle O_i|\Psi\rangle|^2$ for all i. This allows us to focus on the confirmation of alternate dynamical theories without concerning ourselves with the way agents learn about the universe's wave function.

I will initially suppose that MWI is capable of recovering the Born rule probabilities.[8]

CONVENIENT CONJECTURE In MWI, after a measurement of the observable $\widehat{O}$ has been made and before the outcome is observed, the probability one ought to assign to seeing the outcome corresponding to eigenvalue O_i is given by $P(O_i|\text{MWI}) = |\langle O_i|\Psi\rangle|^2$.

This is a highly controversial assumption, so let me clarify the spirit in which I am introducing it. In order to put empirical lower bounds on λ we need to consider cases where GRW becomes more and more like MWI. If we don't have quantitative predictions from MWI, it will not be possible to quantify the success of GRW in these bits of parameter space. Later I'll discuss how things change if the conjecture is false (§5).

In the notation used here, $\text{GRW}_{\lambda=0}$ is MWI. So, when a measurement is made, $P(O_i|\text{MWI}) = P(O_i|\text{GRW}_{\lambda=0})$. Thus if we are assuming that the CONVENIENT CONJECTURE is true and thereby that MWI is empirically adequate, it follows that $\text{GRW}_{\lambda=0}$ is empirically adequate as well.

The question, then, is for what values of λ is GRW approximately empirically equivalent to QM and for what values do the predictions of GRW and QM diverge? If the predictions diverge significantly, GRW becomes empirically inadequate—the data we actually have fits the predictions of QM.[9] For the remainder of this section, take the rate of collapse λ to be sufficiently small that whenever a measurement occurs we can expect there to be copies of the experimenter who record each outcome. From the CONVENIENT CONJECTURE and the fact that the dynamics are the same in GRW and MWI before collapse, it is reasonable to suppose that for these small values of λ the

---

[7]What wonderful theory succeeds in recovering the Born rule, as is demanded of the theory I've called "QM"? This will be a matter of disagreement. Let QM stand in for your favorite theory, whichever you think recovers the right probabilities, be it MWI, $\text{GRW}_{\lambda=10^{-16}\text{s}^{-1}}$, Bohmian mechanics, or something else. $\text{GRW}_{\lambda=10^{-16}\text{s}^{-1}}$ predicts deviations from the Born rule for certain yet-to-be-conducted experiments involving, e.g., macroscopic superpositions (which, even if perfectly isolated from the environment, would be predicted to be unstable). However, for the already-conducted experiments typically taken to provide support for quantum mechanics the predictions should (approximately) match those of the Born rule (setting aside the concerns raised in §6).

[8]For an extended defense of this conjecture, see Wallace (2012). See also Carroll & Sebens (2014); Sebens & Carroll (2014).

[9]It's fine if the predictions for certain *future* experiments diverge (see footnote 7) since the data might (for all we know) support GRW over alternative formulations of quantum mechanics.

probability of seeing each result is given by

$$P\left(O_i | \text{GRW}_\lambda\right) = |\langle O_i | \Psi \rangle|^2 . \tag{4.3}$$

But, the observed experimental outcome is *not* the only data one has to update on. The experimenter should also take into account the fact that she has survived for a time Δt beyond the moment when the measurement was performed. The probability for surviving to Δt can be calculated as

$$P\left(\Delta t | \text{GRW}_\lambda \& O_i\right) = 1 - P\left(\text{fatal collapse by } \Delta t | \text{GRW}_\lambda \& O_i\right)$$
$$= 1 - P\left(\text{death} | \text{collapse by } \Delta t \& \text{GRW}_\lambda \& O_i\right) \times P\left(\text{collapse by } \Delta t | \text{GRW}_\lambda \& O_i\right) . \tag{4.4}$$

The probability of a collapse occurring by Δt can be approximated using (2.2) along with the simplifying assumption that there are N_S particles whose collapse would cause a jump to a single outcome: $P\left(\text{collapse by } \Delta t | \text{GRW}_\lambda \& O_i\right) = 1 - e^{-N_S \lambda \Delta t}$.[10] The probability of dying in the event of such a collapse is just the probability that the collapse is centered around some branch other than one's own: $1 - |\langle O_i | \Psi \rangle|^2$.[11,12] Inserting these two expressions into (4.4) yields

$$P\left(\Delta t | \text{GRW}_\lambda \& O_i\right) = |\langle O_i | \Psi \rangle|^2 + e^{-N_S \lambda \Delta t} - |\langle O_i | \Psi \rangle|^2 e^{-N_S \lambda \Delta t} . \tag{4.5}$$

The probability of the total evidence can be assessed by combining (4.3) and (4.5),[13]

$$P\left(O_i \& \Delta t | \text{GRW}_\lambda\right) = P\left(\Delta t | \text{GRW}_\lambda \& O_i\right) \times P\left(O_i | \text{GRW}_\lambda\right)$$
$$= \left(|\langle O_i | \Psi \rangle|^2 + e^{-N_S \lambda \Delta t} - |\langle O_i | \Psi \rangle|^2 e^{-N_S \lambda \Delta t}\right) |\langle O_i | \Psi \rangle|^2 . \tag{4.6}$$

We can better understand this formula by considering a simple case. Imagine $\lambda \approx 10^{-33} \text{s}^{-1}$ and $N_S \approx 10^{30}$ so that the experimenter can expect to have approximately 15 minutes between measurement and collapse (as in footnote 4). In this time, she can form expectations about what will happen and look around. Suppose she sees an outcome,

_____

[10]More realistically, N_S would increase as a function of time.

[11]This is an optimistic estimate. In fact there will usually be many worlds corresponding to each outcome and thus even when a collapse is centered on the right outcome O_i, one's world might well be destroyed.

[12]Here, to keep things simple, it is assumed that life on the branches not selected by collapse ends immediately (setting aside the possibility of delayed death mentioned in the previous section and footnote 5).

[13]Two clarifications: First, the proposition signified by "$O_i \& \Delta t$" in (4.6) should be understood as the indexical claim "*I* am alive Δt after the experiment and in *my* world the result of the experiment is O_i." not the weaker claim that "There exists a copy of me who is alive Δt after the experiment and in a world where the result of the experiment was O_i." Long after the experiment, the probability of the second claim is given by $|\langle O_i | \Psi \rangle|^2$ since it is just the probability that the GRW collapse will select outcome O_i. Thus the weaker claim might appear friendlier to GRW with small λ. Why focus on the stronger claim? The weaker claim does not take into account one's full (indexical) evidence and using it to update probabilities in GRW and MWI leads to unacceptable results (see footnote 24).

Second, the probability of "$O_i \& \Delta t$" (stronger version) is difficult (perhaps impossible) to assess before the measurement since it is unclear whether one, all, or none of the post-branching copies are identical to the original experimenter. Fortunately, we can focus on the probability assigned to "$O_i \& \Delta t$" *immediately after* branching. Since the experimenter doesn't yet know which branch they are on or whether they will survive, it makes sense to assign probabilities at this point. Further, these later probabilities are what matter for theory confirmation as these are the probabilities assigned to the evidence right before the evidence is acquired.

O_A, with low Born rule probability, $|\langle O_A|\Psi\rangle|^2 = \frac{1}{10}$. She should be somewhat surprised and also afraid. Now she knows that she only has a one in ten chance of survival. If she makes it through the day, she should be surprised again. The probability assigned to the total evidence (surviving and seeing that outcome) is $\frac{1}{10} \times \frac{1}{10} = \frac{1}{100}$, which follows from (4.6) with $\Delta t \gg \frac{1}{N_S \lambda}$.

Consider a variation of the previous case in which the experimenter does not observe the outcome until long after the measurement. Assume for simplicity that there are just two possible outcomes, O_A and O_B, and one branch corresponding to each (see footnote 11). In this case her survival should not be much of a surprise; the probability is 82%. The probability of O_A is 10% and the chance of survival given O_A is 10%. The probability of the other outcome, O_B, is 90% and the chance of survival given O_B is 90%. Thus the total chance of survival is $\frac{1}{10} \times \frac{1}{10} + \frac{9}{10} \times \frac{9}{10} = \frac{82}{100}$. The probability she should assign to O_A given that she survived can be calculated by Bayes' theorem as the probability of survival conditional on O_A, $\frac{1}{10}$, times the probability of O_A, $\frac{1}{10}$, divided by the probability of survival, $\frac{82}{100}$. This yields $\frac{1}{82}$. The probability assigned to her total evidence is the probability of surviving times the probability of seeing O_A upon surviving, $\frac{82}{100} \times \frac{1}{82} = \frac{1}{100}$ (the same result as was obtained in the first case).

In the first case—observation before collapse—the probability assigned to seeing O_A was correct but the subsequent probability of survival was in disagreement with QM. In the second case—collapse before observation—the probability of survival was in disagreement with QM *and* the subsequent probability of O_A was incorrect. The fact that the probability of survival was less than one in the second case shows that GRW_λ (with λ very small) could be disconfirmed by repeated experiments even if no one bothers to look at the results of the experiments. In the second case, unlike the first, O_A is assigned a probability in disagreement with the Born rule. (The reason for this disagreement is that outcomes which were already improbable get further penalized for poorly predicting the agent's survival.) Thus the problem for small values of λ is not *merely* that the probability of survival is low but also that supposing one has survived these dangerous collapse events leads to poor predictions about the outcomes of measurements that have already been made. Note that by focusing on the probability of the total evidence, (4.6), we need not worry about whether collapse happens before or after observation of the outcome.

If λ is so small that no collapses are expected to occur within any reasonable length of time Δt and the CONVENIENT CONJECTURE holds, the predictions of GRW_λ approximately match those of QM. However, as has been noted (Feldmann & Tumulka, 2012, §4), there would be little motivation for such a theory. It would be simpler to just set λ to zero and remove the collapses all together, yielding MWI. As λ grows it becomes more likely that a collapse will have occurred within Δt and the disagreement between GRW_λ and QM gets worse. QM predicts that you will be alive whereas GRW_λ assigns a certain probability to your death. For fixed λ, the larger Δt is the larger the disagreement between QM and GRW_λ; see (4.6). However, once λ is sufficiently large the assumption that branching precedes collapse becomes invalid. In the next section I'll consider cases in which branching is prevented by collapse.

The fact that one's own continued survival is used as evidence for assessing theories is undeniably odd. Experimenters don't typically keep track of the time elapsed since the experiment was performed. But, epistemologists have contemplated cases much like this where survival *is* relevant data. Consider the following much-discussed example (Leslie, 1989; Swinburne, 1990):

Firing Squad Suppose that a dozen well-trained shooters are ordered to execute you by firing 12 shots each. While blindfolded you hear 144 shots ring out but you survive unscathed.

In such a scenario, your own survival provides evidence that the shooters intentionally let you live over the alternative hypothesis that you got lucky because each of the 144 shots missed its intended target.

The situation here is similar to *Firing Squad*. The hypothesis that the squad intentionally misses is like the hypothesis that QM is true and there are no cataclysmic collapse events. The hypothesis that the shooters were attempting to kill you is like the hypothesis that GRW$_\lambda$ is true for some troublesome small-but-not-too-small choice of λ where worlds are constantly snuffed out quickly and without warning. However, there is an important difference: In *Firing Squad*, the target will either survive or be killed. In GRW$_\lambda$ with troublesome λ, there will be many versions of the experimenter that are killed and always at least one that survives. A closer non-quantum analogy is:

Prison Poisoning On New Year's Day you wake up in a nondescript prison cell, #27. A coin was flipped. On New Year's Eve, you were blindfolded and shipped either to *Alcatraz*, if heads, or *Arkham*, if tails. Each prison contains 100 numbered cells and you were randomly assigned to #27.[14] While you slept in your cell the new year began with a randomly chosen 99 of the 100 cells in Arkham being filled with deadly poison gas. Those in Alcatraz were safe. You knew the plan all along.

In this case, you should initially think it equally likely that you ended up in either prison. After surviving the night you should come to believe that you were probably shipped to Alcatraz since being shipped to Arkham would have likely resulted in your death. It was guaranteed that one of the prisoners in Arkham would survive, but it was not likely to be the one in cell #27. Alcatraz is like MWI and Arkham is like GRW with troublesome λ. The numbered cells represent 100 possible results of a measurement and the gas plays the role of collapse. For an analogue of GRW with a normal collapse rate, one could introduce the possibility of being sent to a third prison with a single cell, randomly numbered and free of poison.[15]

Cases like *Firing Squad* and *Prison Poisoning* have a curious feature: in both scenarios, one hypothesis cannot be confirmed by the subject. If the bullets and poison kill instantly, no course of experience would support the hypothesis that the squad was trying to kill you or that you were sent to Arkham. Similarly, if collapse kills instantly there are no experiences one could have that would provide evidence for GRW with troublesome λ over QM (*if* the CONVENIENT CONJECTURE holds). In *Firing Squad*, the problem with the hypothesis that the squad is trying to kill you is not that it predicts odd experiences but that it predicts your experiences will end. In GRW with small λ too the problem is not odd experiences. In fact, in a theory like GRW$_{\lambda=10^{-33}\text{s}^{-1}}$ if the outcomes of many repeated experiments are recorded one can expect[16] with high probability that at any given time the record will show a sequence of results that looks

---

[14]For the closest analogy, imagine that each cell of your prison is occupied by a copy of you that resulted from a 1-to-100 fission midday on New Year's Eve.

[15]In this case, the fission in footnote 14 should not be supposed.

[16]That is, looking at the probabilities derived from the collapse process this is what one should expect. As discussed earlier in this section, if one performs experiments, survives for a long time, and doesn't look at the outcome, the probabilities that should be assigned to the different possible outcomes are not the standard Born rule probabilities. Seeing a sequence that fits *these* expectations should shift one's credence towards GRW with small lambda—but, only after the theory has been significantly

randomly generated with each outcome's probability weighted by $|\langle O_i|\Psi\rangle|^2$.[17] If a branching has just occurred, there may be multiple distinct versions of the record but each will show the same long-run frequencies for the various outcomes as the records will only differ in the last few entries. At any time, a typical observer will remember, and have records of, measurement results that fit the predictions of QM. This feature of the theory might cause one to doubt whether we could have empirical evidence against $\text{GRW}_{\lambda=10^{-33}\text{s}^{-1}}$, but it shouldn't.[18] As in *Prison Poisoning* and *Firing Squad*, the fact that one has survived *is relevant evidence* in determining which hypothesis to believe. In contrast to GRW with small λ, MWI predicts that there will at any time be many observers whose memories and records don't fit the predictions of QM—every sequence of quantum measurement results is observed by the inhabitants of some quantum world. The theory then faces the challenge of explaining why we should not expect to be one of these observers (the challenge of establishing the truth of the CONVENIENT CONJECTURE).

Those who are attracted to the idea of quantum immortality may object to the conclusions reached in this section. Consider a dangerous branching event from the perspective of the many-worlds interpretation (a "quantum suicide" scenario). Suppose you will survive on one branch and die immediately, or quickly, on all others. It is tempting to think you should expect survival with certainty. As Lewis (2004) put it, "The experience of being dead should never be expected to any degree at all, because there is no such experience." If death is indeed immediate on all branches but one, the thought has some plausibility. But if there is any delay it should be rejected. In such a case, there is a short period of time when there are multiple copies of you, each (effectively) causally isolated from the others and able to assign a credence to being the one who will live.[19] Only one will survive. Surely rationality does not compel you to be maximally optimistic in such a scenario.[20] The situation in GRW with a troublesome collapse rate is just like the delayed-death version of the above quantum suicide scenario and, as in that case, survival should not receive probability one.[21] If the collapse rate is raised so that the agent never splits into multiple copies, there is no danger of death and survival can be expected with certainty.

---

disconfirmed by one's survival. Thus unlike *Firing Squad* and *Prison Poisoning*, there is in fact a way to get a piece of evidence that points towards the dangerous hypothesis. Still, one's total course of experience will never favor GRW with small λ over QM; as can be seen by noting that the expression in parentheses in (4.6) is at most one.

[17]If the collapse rate is much smaller, there will be many records only some of which show sequences deemed probable by the Born rule (as in MWI).

[18]Wallace (2014) in considering a similar situation seems to find this—"strictly speaking"—sufficient empirical success as the theory does manage to "explain why the scientific community has so far observed statistical results in accord with quantum mechanics (via the anthropic fact that worlds in which violations were observed are now radioactive deserts [the fate he believes befalls worlds in the tails]). And it explains why it is rational to act as if the predictions of quantum mechanics were true (because in those worlds where they turn out false, we're all doomed anyway)." Vaidman (2014b, §8) also seems untroubled by the possibility of death in the tail branches.

[19]Do the copies need to last long enough to have thoughts to cause trouble? I think not. If you survive, you can consider what credences you should have assigned during the short period after splitting when you coexisted with the other copies.

[20]The situation here is like that of the prisoner in Arkham if the period between the splitting event (see footnote 14) and the deaths were made much shorter.

[21]Anticipating an upcoming branching and subsequent collapse, can one assign a probability to survival *before* splitting? If so, does the fact that some successor will survive the collapse mean that before splitting survival is certain? These questions need not be answered as the relevant probabilities are those assigned immediately after branching (footnote 13).

5   Averting Branching

If collapse occurs sufficiently soon after a measurement, branching can be averted. As the other branches of the universe where the outcome was different are just beginning to form the collapse event occurs, ensuring that the macroscopic readout gives a definite result and the experimenter sees a single outcome. The simplest way to incorporate this feature of the theory is by introducing a cutoff characterizing the amount of time that passes before branching occurs if there is no collapse. If a collapse happens within τ, branching is averted and a single outcome occurs. If collapse does not occur until after τ, then there is a branching of worlds before the collapse, as in the previous section.[22] Let $C_{<\tau}$ indicate that collapse occurs before the cutoff, $C_{>\tau}$ indicate after. Including both of these possibilities, the probability of the data given the theory can be expressed as

$$
P\left(O_i \& \Delta t | \mathrm{GRW}_\lambda\right) = \overbrace{P\left(O_i \& \Delta t | \mathrm{GRW}_\lambda \& C_{>\tau}\right)}^{①} \times \overbrace{P\left(C_{>\tau} | \mathrm{GRW}_\lambda\right)}^{②}
$$
$$
+ \underbrace{P\left(O_i \& \Delta t | \mathrm{GRW}_\lambda \& C_{<\tau}\right)}_{③} \times \underbrace{P\left(C_{<\tau} | \mathrm{GRW}_\lambda\right)}_{④} . \qquad (5.1)
$$

The first piece, ①, is just as in (4.6) where it was assumed that branching preceded collapse. The fourth piece, ④, is the probability that a collapse happens by τ. This follows directly from (2.2), $④ = 1 - e^{-N_S \lambda \tau}$. The second piece is simply the probability that a collapse does *not* occur, $② = 1 - ④$. The third piece, ③, is the probability that a given outcome resulted from the GRW collapse process in a case where branching does not occur. Here we have GRW working as intended and the probability should be in approximate agreement with the Born rule provided λ is not so large as to push us into the empirically refuted region of parameter space (figure 1), $③ \approx |\langle O_i | \Psi \rangle|^2$.[23] Inserting these expressions in (5.1) and rearranging gives,

$$
P\left(O_i \& \Delta t | \mathrm{GRW}_\lambda\right) = |\langle O_i | \Psi \rangle|^2 - \left(1 - |\langle O_i | \Psi \rangle|^2\right)\left(1 - e^{-N_S \lambda \Delta t}\right)|\langle O_i | \Psi \rangle|^2 e^{-N_S \lambda \tau} ,
$$
$$
(5.2)
$$

which limits to the Born rule probabilities as λ goes to zero or infinity. (5.2) is not valid if λ is large enough that the probabilities in ③ deviate significantly from those given by the Born rule. It cannot be extended in a simple and general manner as the way in which ③ deviates from $|\langle O_i | \Psi \rangle|^2$ will be depend on the particular experiment under consideration.

In this simplified story, the probability of surviving to Δt and seeing a certain outcome O_i depends dramatically and discontinuously on whether collapse happens before or after branching. The expressions for ① and ③ are quite different. A more careful analysis would ideally give a smooth transition or justify a precise cutoff, but this would require wading into the murky territory of collapses that occur *during* branching and settling questions of personal identity there (in particular, when exactly personal fission occurs and whether it can, in any relevant sense, partially occur). It might be seen either as intriguing or disconcerting that we must answer questions of personal

---

[22]The cutoff τ is not a free parameter and not derived from the collapse process. It could be calculated by determining when branching occurs in the absence of collapse (see §6).

[23]Here it is also assumed that we're considering familiar experiments, not future ones that probe smaller values of λ (see footnote 7).

identity in the context of MWI to put precise lower bounds on λ in GRW.

To recap: If λ is so extremely small that you should not expect (relevant) collapses to have occurred in your lifetime (figure 2.a), then GRW_λ is empirically adequate if the CONVENIENT CONJECTURE holds. If λ is large enough that collapses must be considered but small enough that branching typically precedes collapse (figure 2.c), then early death is the norm and one's continued survival provides strong evidence against the theory. If λ is increased to around the initially proposed value of $10^{-16}\mathrm{s}^{-1}$ (figure 2.b), the theory may again be empirically adequate as branching is prevented by collapse and the collapse process ensures that the probabilities of various outcomes are given by the Born rule. If λ is increased even further, so that $\lambda > 10^{-8}$, the theory is again empirically inadequate as collapses occur too frequently. Superpositions are destroyed mid-experiment and other maladies ensue (see Feldmann & Tumulka, 2012; Bassi *et al.* , 2013).

What happens if the CONVENIENT CONJECTURE is false and MWI gives different probabilities from QM? Then, $\mathrm{GRW}_{\lambda=0}$ is empirically inadequate as $\mathrm{GRW}_{\lambda=0}$ is MWI. This failure also rules out GRW_λ for very small λ where collapses can be neglected. For larger values of λ where collapse is rare but relevant, there are now two ways in which the theory fails: the probabilities of the various outcomes are incorrect and there is, in general, some probability that one would not have survived to Δt.[24] For still larger values of λ that successfully avert branching, the theory again has a chance of being empirically adequate since the probabilities of outcomes are now determined by the collapse process and the MWI probabilities are irrelevant.

6 The Race: Decoherence versus Collapse

For GRW to be tenable, there must be values of λ for which the theory is empirically adequate. On the one hand, λ must be large enough that collapse practically never occurs after the experimenter has branched into multiple copies. Otherwise, one's continued survival empirically refutes GRW_λ, (5.2). On the other hand, λ must be small enough that collapses do not spoil the predictions for experiments that have already been performed. That is, λ must lie below the experimentally refuted region of figure 1. But, are there any values in this range? To answer this, we need to determine whether decoherence-induced branching tends to occur before or after collapse.[25]

We know that for values of λ near the originally suggested value, $10^{-16}\mathrm{s}^{-1}$, the experiment readout and the experimenter are in a well-defined state corresponding to a single outcome very soon after the measurement occurs. But, what is not clear is which of two possibilities occurred immediately after the measurement (figure 3): (a) the world briefly branched and then a collapse event destroyed some of the copies of the experimenter, or (b) there was never a branching event because collapse prevented the microscopic superposition from causing the experimenter to enter into a superposition.

A proper analysis is warranted, but beyond the scope of this paper. Here is a *very* rough calculation of how quickly collapse would have to occur to prevent

---

[24]When Δt is large, the probability of the total evidence given by (4.6) is $|\langle O_i|\Psi\rangle|^4$ because the probability of seeing O_i is $|\langle O_i|\Psi\rangle|^2$ and the probability of subsequently surviving is also $|\langle O_i|\Psi\rangle|^2$. This suggests a crafty maneuver to save GRW with small λ. What if the probability of each outcome were one—after all, each outcome will be recorded by someone—so that the probability of total evidence is $|\langle O_i|\Psi\rangle|^2$ as in QM? This proposal faces a fatal problem: practically every experiment would confirm MWI over single-world theories (see Greaves, 2007, §4).

[25]See Schlosshauer (2005, §IV.E); Bacciagaluppi (2012, §3.1.2) for discussion of decoherence in GRW.
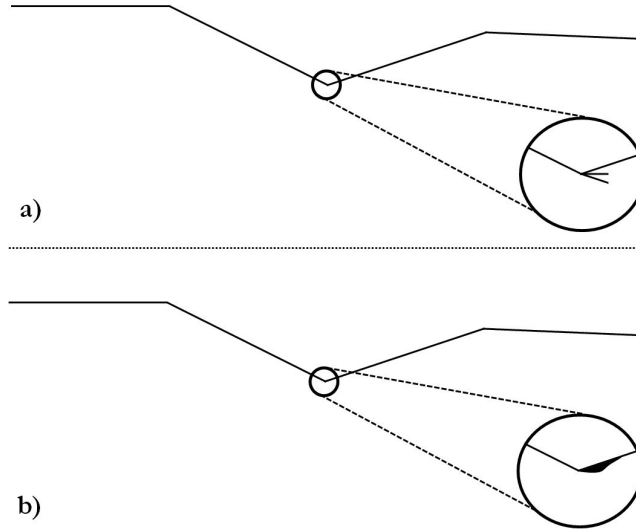
Figure 3: Two potential close-ups of figure 2.b.

decoherence-induced branching: Decoherence is fast. A slow estimate might be 10^{-23}s for 1 gram of matter at room temperature in a superposition of two locations separated by one centimeter (Zurek, 2003). To ensure a 95% probability of collapse by 10^{-23}s, λ would have to be at least 3 s^{-1} (from (2.2), assuming the number of particles is on the scale of moles, $N = 10^{23}$). But, experiments restrict λ to being at most 10^{-8}s^{-1} (figure 1). This calculation suggests trouble. There may not be a safe region of parameter space.[26]

Let me highlight one particularly pernicious simplification in this rough calculation: It is assumed that the bit of matter starts in a superposition. In actuality, it would take time for the matter to enter a superposition and a collapse event could occur in this interval, preventing the macroscopic superposition from forming. It would (normally) take even longer for a large object like a human to enter a superposition of spatially disjoint states (although this is certainly asking too much; the presence of a far less dramatic superposition should be sufficient for the human to branch).

I'll close by summarizing the key lessons of the analysis. First, to determine precise experimental bounds on the parameters λ and σ in GRW, we must answer metaphysical and epistemological questions about MWI: When/how does branching occur in the absence of collapse (§5 & 6)? What probabilities should be assigned to different outcomes in MWI (§4)? This provides additional motivation for that ongoing research program. Second, even if the CONVENIENT CONJECTURE holds and MWI is empirically adequate, some of the philosophically unsatisfactory region of parameter space is also empirically refuted (§3, 4, & 5). Surprisingly, it is not refuted by the outcomes we observe, but by the fact that we live long enough to observe so many of them. Third, it is not clear how to draw a principled border for the philosophically unsatisfactory region if our dissatisfaction is purely "philosophical" (§1). But, with the realization that small values of the collapse rate λ are empirically refuted, we now have a method to begin drawing principled lower bounds on λ: determine whether the experimenter branches before or after collapse (§5 & 6). Simple calculations suggest that the lower bound generated

---

[26]This concern is corroborated by the calculations in Tegmark (1993); Benatti *et al.* (1995).

from empirical considerations will be stronger than the bound generated from a distaste for long lasting macroscopic superpositions, perhaps strong enough to rule out GRW entirely (§6). This merits further study.

7 Acknowledgments

References

Adler, Stephen L. 2007. Lower and Upper Bounds on CSL Parameters from Latent Image Formation and IGM Heating. *Journal of Physics A: Mathematical and Theoretical*, **40**(12), 2935.

Aicardi, Franca, Borsellino, Antonio, Ghirardi, Gian Carlo, & Grassi, Renata. 1991. Dynamical Models for State-Vector Reduction: Do they ensure that measurements have outcomes? *Foundations of Physics Letters*, **4**(2), 109–128.

Allori, Valia, Goldstein, Sheldon, Tumulka, Roderich, & Zanghì, Nino. 2008. On the Common Structure of Bohmian Mechanics and the Ghirardi–Rimini–Weber Theory. *The British Journal for the Philosophy of Science*, **59**(3), 353–389.

Allori, Valia, Goldstein, Sheldon, Tumulka, Roderich, & Zanghì, Nino. 2011. Many Worlds and Schrödinger's First Quantum Theory. *The British Journal for the Philosophy of Science*, **62**(1), 1–27.

Arntzenius, Frank. 2003. Some Problems for Conditionalization and Reflection. *The Journal of Philosophy*, **100**(7), 356–370.

Bacciagaluppi, Guido. 2012. The Role of Decoherence in Quantum Mechanics. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, winter 2012 edn.

Bassi, Angelo, Deckert, Dirk-André, & Ferialdi, Luca. 2010. Breaking Quantum Linearity: Constraints from Human Perception and Cosmological Implications. *EPL (Europhysics Letters)*, **92**(5), 50006.

Bassi, Angelo, Lochan, Kinjalk, Satin, Seema, Singh, Tejinder P, & Ulbricht, Hendrik. 2013. Models of Wave-Function Collapse, Underlying Theories, and Experimental Tests. *Reviews of Modern Physics*, **85**(2), 471.

Benatti, F, Ghirardi, Gian Carlo, & Grassi, R. 1995. Quantum Mechanics with Spontaneous Localization and Experiments. *Pages 263–279 of:* Beltrametti, Enrico, & Lévy-Leblond, Jean-Marc (eds), *Advances in quantum phenomena*. Springer.

Carroll, Sean M., & Sebens, Charles T. 2014. Many Worlds, the Born Rule, and Self-locating Uncertainty. *In:* Struppa, Daniele, & Tollaksen, Jeff (eds), *Quantum Theory: A Two-Time Success Story: Yakir Aharonov Festschrift*. Springer.

Feldmann, William, & Tumulka, Roderich. 2012. Parameter Diagrams of the GRW and

CSL Theories of Wavefunction Collapse. *Journal of Physics A: Mathematical and Theoretical*, **45**(6), 065304.

Ghirardi, GianCarlo, Rimini, Alberto, & Weber, Tullio. 1986. Unified Dynamics for Microscopic and Macroscopic Systems. *Physical Review D*, **34**(2), 470.

Gisin, Nicolas, & Percival, Ian C. 1993. The Quantum State Diffusion Picture of Physical Processes. *Journal of Physics A: Mathematical and General*, **26**(9), 2245.

Greaves, Hilary. 2007. On the Everettian Epistemic Problem. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **38**(1), 120–152.

Leslie, John. 1989. *Universes*. Routledge.

Lewis, David. 2004. How Many Lives has Schrödinger's Cat? *Australasian Journal of Philosophy*, **82**(1), 3–22.

Manley, David. 2014. On Being a Random Sample. Manuscript.

Maudlin, Tim. 2010. Can the World be Only Wavefunction? *Pages 121–143 of:* Saunders, Simon, Barrett, Jonathan, Kent, Adrian, & Wallace, David (eds), *Many Worlds?: Everett, Quantum Theory, & Reality*. Oxford University Press.

Meacham, Christopher J. G. 2010. Unravelling the Tangled Web: Continuity, Internalism, Uniqueness and Self-Locating Belief. *Pages 86–125 of:* Szabó Gendler, Tamar, & Hawthorne, John (eds), *Oxford Studies in Epistemology*, vol. 3.

Pearle, Philip, & Squires, Euan. 1994. Bound State Excitation, Nucleon Decay Experiments and Models of Wave Function Collapse. *Phys. Rev. Lett.*, **73**, 1–5.

Schlosshauer, M. 2005. Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics. *Reviews of Modern Physics*, **76**(4).

Sebens, Charles T., & Carroll, Sean M. 2014. Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. arXiv:1405.7577 [quant-ph].

Swinburne, Richard. 1990. The Argument from the Fine Tuning of the Universe. *In:* Leslie, John (ed), *Physical Cosmology and Philosophy*. MacMillan, New York.

Tegmark, Max. 1993. Apparent Wave Function Collapse Caused by Scattering. *Foundations of Physics Letters*, **6**(6), 571–590.

Vaidman, Lev. 2014a. Many-Worlds Interpretation of Quantum Mechanics. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, spring 2014 edn.

Vaidman, Lev. 2014b. Quantum Theory and Determinism. *Quantum Studies: Mathematics and Foundations*, **1**(1-2), 5–38.

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

Wallace, David. 2014. Life and Death in the Tails of the GRW Wave Function. arXiv:1407.4746 [quant-ph].

Zurek, Wojciech H. 2003. Decoherence and the Transition from Quantum to Classical–REVISITED. quant-ph/0306072.