Can Many-Worlds Survive a Quantum Doomsday?

Brett M. Bevers

Abstract

A novel puzzle for the notion of probability in the Many-Worlds interpretation of quantum mechanics is presented. The puzzle makes use of a thought experiment that some have claimed would provide empirical support for many-worlds over alternatives. It is argued that, if the predictions of Many-Worlds do indeed differ from other interpretations as claimed, then Born's rule must generally be invalid in Many-Worlds. It is shown that the thought experiment provides a counter example for recent decision-theoretic arguments that purport to establish Born's rule. Finally, it is shown that the puzzle can be grounded in general considerations regarding the nature of prediction in Many-Worlds.

The *Many-Worlds interpretation* (MW) of quantum mechanics is routinely mentioned alongside various hidden-variable, collapse and modal formulations as a proposed solution to the quantum measurement problem. The MW is unique among these alternatives, however, for the fact that there is serious contention over what predictions the theory makes. Though Hugh Everett III is credited with first proposing a many-worlds interpretation, the overwhelming consensus is that Everett's own account of quantum measurement was incomplete. Bryce DeWitt, who first promoted Everett's proposal under the name 'many-universes', identified lacunae that have remained central for both proponents and critics. One crucial open question concerns the connection between Everett's model of measurement and the observed quantum statistics. Everett's principal move was to eliminate the standard collapse dynamics and, consequently, the objective probabilities stipulated therein. Yet MW purports to account for the very same phenomena. Since it is not obvious that the many-worlds picture places any alternative constraint on expectations, we are owed some account of how the observed quantum statistics end up being the *right* statistics. Work on this subject has been disparate; but it has been called the *problem of probability*, because it is most often posed as the problem of showing how objective probabilities arise in MW.

Despite differences in how the problem of probability is approached, the stakes for MW are always the same. The empirical content of the standard (textbook) formulation of quantum mechanics consists of probabilistic assertions derived according to *Born's rule*. The majority of the evidence that we have for quantum mechanics consists of the observed statistics in repeated experiment trials or large ensembles of object systems, which (as far as we can tell) are *random*. Hence, the quantum probabilities are indispensable. Born's rule must be validated in any formulation of quantum theory that enjoys at least as much empirical success as the standard formulation.

Quantum suicide (also called *quantum Russian roulette* or *quantum immortality*) is a compelling thought experiment that exercises one's understanding of probability and prediction in the context of MW. Max Tegmark was first to use the term 'quantum suicide' in print, and to claim that MW makes novel predictions regarding such an experiment (1998). Tegmark concludes that quantum suicide could actually serve to confirm MW over alternative formulations. Only limited work has been done regarding quantum suicide since (Lewis, P. 2000; Papineau 2003, 2004; Lewis, D. 2004; Tappenden 2004, 2009). Consequently, the significance of that thought experiment for the problem of probability has not been generally recognized. In the present paper I aim to clarify the reasoning that leads to novel predictions and provide a statement of the problem that they pose for the notion of probability in MW.

1 Quantum Suicide

The essential components of quantum suicide are a *quantum experiment* and a *lethal weapon*. With some ingenuity, one might be able to tailor the thought experiment to any pair of quantum phenomenon and cause of death, but it will be helpful to make some simplifying assumptions. We suppose that the experiment can be performed at will on an easily prepared system, and that the measured observable has a discrete spectrum. It should make sense to talk about the execution of this experiment and each of its possible outcomes as a distinct event. We also suppose that the weapon produces near-instantaneous destruction on whatever scale is required. As far as the participants in quantum suicide are concerned, the weapon's sole product is *oblivion*—the weapon immediately destroys everything of consequence for those involved. Given such a pair of quantum experiment and deadly device, one prepares for quantum suicide by rigging the experiment apparatus so that, instead of producing a readable measurement record, the apparatus triggers the weapon *in all but one* of the possible experiment outcomes. Finally, one performs quantum suicide by turning the weapon on oneself while running the experiment.

As an instance of quantum suicide consider the following. Imagine that some care has been taken in setting up an apparatus that is ready to perform a *Stern-Gerlach spin measurement* on a single electron. Since an electron is a spin-$\frac{1}{2}$ particle, there are two spin values for any given orientation, which we will label '↑' for *spin-up* and '↓' for *spin-down*. The Stern-Gerlach apparatus "observes" the spin value of the electron by detecting its position after passing through an inhomogeneous magnetic field, attributing ↑ if the electron is found deflected upward and ↓ if it is found deflected downward. Now also imagine that great care has been taken in constructing a *doomsday device* that can—in just a fraction of a millisecond—convert everything within a certain radius of the Stern-Gerlach apparatus into a chaotic swarm of particles and radiation. This doomsday device is composed, if you like, of an array of strategically placed and suitably powerful bombs that are timed to go off simultaneously. The device has been designed to be covertly triggered by the Stern-Gerlach type apparatus. The usual instrument readout, or "pointer", has been omitted so that there is no discernible record of the electron's spin value; instead, the apparatus detonates the doomsday device if and only if the measurement is performed and the value ↓ is observed. Importantly, no readable measurement record is produced in the process of triggering the doomsday device. We might suppose that the whole process happens so quickly that observers are prevented from interacting with the apparatus or weapon in the time that it takes to make the measurement and detonate the device. Under these conditions, performing the Stern-Gerlach experiment is a clear example of quantum suicide. This scenario incorporates the essential features as well as the simplifying assumptions identified above, and it will serve as our prototype.

Quantum suicide is intended as an empirical test for the various formulations of quantum mechanics. If someone is in a position to perform quantum suicide, and they can trust the engineering of the apparatus and weapon, then what quantum mechanics says about the interaction of the object system and apparatus has vital consequences. Given that the interaction is itself a clear-cut quantum experiment, quantum theory certainly has something to say. The usual method of making predictions straightforwardly applies in the case of our prototypical quantum suicide. Spin is treated as an intrinsic property of the electron that is manifest in its deflection; so, in practice, we can model the Stern-Gerlach experiment by postulating some initial spin state for the electron and applying Born's rule directly. Let $\{|\uparrow\rangle, |\downarrow\rangle\}$ be an orthonormal eigenbasis for the measured spin observable, then we may represent the state of the electron just before the measurement interaction as

$$\psi_e = a_1|\uparrow\rangle + a_2|\downarrow\rangle\,,$$

where a_1 and a_2 are complex amplitudes. Born's rule stipulates that the electron will exhibit *one* of the two spin values, and that the probability of each outcome (denoted $\Pr(\uparrow)$ and $\Pr(\downarrow)$ respectively) is given by

$$\Pr(\uparrow) = |a_1|^2 \quad \text{and} \quad \Pr(\downarrow) = |a_2|^2.$$

By hypothesis, the doomsday device remains inactive if the observed value is ↑ and is triggered if the observed value is ↓. Moreover, immediate oblivion is the assured consequence of triggering the doomsday device. Therefore, the usual method yields a precise description of how likely it is that performing quantum suicide will be fatal. The probability of *survival* and of *annihilation* are

$$\Pr(\ddot\smile) = |a_1|^2 \quad \text{and} \quad \Pr(\mathbf{\Omega}) = |a_2|^2$$

respectively. Depending on the initial spin state of the electron, performing quantum suicide may be absolutely safe, certainly fatal, or anything in between.

On the face of it, $\Pr(\ddot\smile)$ and $\Pr(\mathbf{\Omega})$ are the same sort of probabilities with which quantum theory has successfully predicted a wide variety of phenomena. The standard formulation does not give special consideration to "dangerous" interactions—it might (in principle) be invoked to account for the functioning of the weapon itself. But some have argued that, if MW turns out to be correct, the usual method for making predictions *does not apply* to the case of quantum suicide (Tegmark 1998; Lewis, P. 2000; Lewis, D. 2004; Tappenden 2004, 2009). And this is not because the usual method could end up being inaccurate or misleading about particular details. Rather, it is claimed that MW yields *an entirely different kind of prediction.*

MW is a *no-collapse* formulation. No-collapse formulations treat measurement as a unitary interaction between an object system and apparatus; that is, the apparatus is also treated as a quantum system, and the joint object-apparatus system evolves unitarily. The immediate consequence is that a measurement will generally cause the apparatus to become entangled with the object system, no matter how macroscopic the apparatus is. Given a no-collapse formulation, the vector $\psi = a_1 \left|\uparrow\right\rangle + a_2 \left|\downarrow\right\rangle$ not only represents the spin state of the electron, it is also indicative of the entangled post-measurement state of the object-apparatus system. We might write the post-measurement state as

$$\psi_{e+A} = a_1 \left|\uparrow, \text{``}\uparrow\text{''}\right\rangle + a_2 \left|\downarrow, \text{``}\downarrow\text{''}\right\rangle ,$$

where $\left|\uparrow, \text{``}\uparrow\text{''}\right\rangle$ is understood as an eigenstate for the joint system in which the electron is deflected upward and the apparatus attributes $\uparrow$, and $\left|\downarrow, \text{``}\downarrow\text{''}\right\rangle$ is an eigenstate in which the electron is deflected downward and the apparatus attributes $\downarrow$.[1] The most distinctive feature of MW is the way in which the entangled object-apparatus state is interpreted. Everett's original claim, which has been variously construed, was that the components of the entangled post-measurement state are all "equally real." Subsequent many-worlds theorists have supplemented this claim with an account of how distinct *worlds* come to correspond to each component of the predicted post-measurement state (as written in the measurement basis.) Typically, the surrounding environment (including human observers) will quickly become entangled with the relatively stable state of the apparatus, so that the measurement outcome is recorded redundantly throughout the environment. Measurement interactions are said to cause a world to "split" or "branch" into several worlds, in which each possible measurement outcome has occurred. Going back to our example, if the electron is initially in a superposition of spin eigenstates, after the measurement there will be one component of the electron-apparatus state in which that electron has been deflected fully upward and another component in which it has been deflected fully downward. These components are said to correspond to distinct worlds, which are meant to account for our experience (under normal circumstances) of a determinate spin value, either $\uparrow$ or $\downarrow$.

Many-worlds theorists maintain that it is coherent to use Born's rule to derive predictions from MW. More precisely, they hold that one should set their *expectations* according to probabilities like $\Pr(\uparrow)$ and $\Pr(\downarrow)$ in a splitting universe, and that this carries the same

---

[1] Realistically, we would also expect there to be terms in which the experiment does not proceed as planned because the electron fails to enter the apparatus, parts of the apparatus behave erratically, etc. This will not be important in what follows.

implications for the observed quantum phenomena and statistics as in the standard formulation. However, some proponents of MW consider quantum suicide to be an exception, and they argue that MW and the standard formulation make different predictions regarding such circumstances. In a universe described by MW, the state of the doomsday device quickly becomes entangled with the state of the electron and Stern-Gerlach apparatus. The state of the object-apparatus-weapon system becomes

$$\psi_{e+A+D} = a_1 \left|\uparrow, \text{``}\uparrow\text{''}, \ddot{\smile}\right\rangle + a_2 \left|\mathbf{\Omega}\right\rangle \, ,$$

where $\left|\uparrow, \text{``}\uparrow\text{''}, \ddot{\smile}\right\rangle$ is a state in which the electron is deflected upward, the apparatus attributes ↑, and the doomsday device remains inactive, and $|\mathbf{\Omega}\rangle$ represents the chaotic aftermath of triggering the doomsday device. Suppose we now ask what this state tells us about what the participants in quantum suicide should expect as an outcome. They should clearly expect to become entangled with the object-apparatus-weapon system, but how should they anticipate this state of affairs? Under MW, one thing that can be read off of the quantum state is that there is only *one* set of participants before, during, and after performing quantum suicide; that is, the number of worlds in which there are such individuals at work does not multiply at any point. This is a somewhat delicate point; but, recall, we have assumed that the participants are effectively isolated from the object-apparatus-weapon system before the weapon detonates. Because $|\mathbf{\Omega}\rangle$ is not a state in which there are any individuals, it never happens that the participants are described in multiple components of the resulting entangled superposition. What is more obvious is that any *experiences* that can be ascribed to the participants form a single, unbroken narrative. $|\mathbf{\Omega}\rangle$ is unquestionably not a state in which any kind of experience is being had; hence, each participant only experiences worlds in which she initiates the experiment, awaits the outcome, and leaves unscathed. And, it can be argued that this reading of the quantum state entails that the participants should fully expect to become those unique individuals and that their experiment will proceed according to that narrative. In other words, MW might be taken to predict (with deterministic certainty) that *the electron will be deflected upward, the doomsday device will remain inactive, and (consequently) one will survive quantum suicide.*

2 A Problem for Probability

What has not been generally recognized is that quantum suicide presents a problem for the notion of *probability* in the context of MW. Suppose for the moment that MW does entail that one is certain to survive quantum suicide. The standard formulation predicts, by way of Born's rule, that the probability of surviving is given by Pr(☺). If the object system is initially in a superposition of spin values, Pr(☺) can take a value anywhere between 0 and 1. Hence, Born's rule is inconsistent with the claim that the participants in quantum suicide should expect (to the fullest degree) that the weapon will remain inactive. Yet, in any mundane quantum experiment, the standard probabilistic predictions are supposed to apply—and they *must* apply if MW accounts for quantum phenomena.

It would not be hard to adjust Born's rule to accommodate situations like quantum suicide. After *reducing* and *renormalizing* the actual quantum state given what one knows about their branch, one might *conditionalize* so as to obtain a superposition over only those outcomes that are of interest. But, if the evolution of the quantum state is always described by Schrödinger's equation, superpositions never *really* undergo such transformations. Conditionalization is just as foreign to unitary quantum mechanics as reduction and renormalization, and it removes the predictions of MW yet another step from the actual physics. After all, conditionalization is an operation that is naturally applied to *probability distributions*, and it would be question-begging to apply it to any other quantity without adequate justification. Such a move would further complicate the question that is raised by the problem of probability: Why should the amplitudes of vector components in the quantum state bear any relation to one's expectations under MW? In particular, why should Born's rule *ever* apply?

Quantum suicide is a circumstance in which amplitudes appear to be irrelevant. One's expectations are not affected by the fact that the branch in which one survives is atypical under the measure stipulated by Born's rule. But how could this be the case if that measure represents objective probability? Even if it is known beforehand that exactly one branch will contain survivors, shouldn't those survivors be *surprised* to find themselves in an atypical situation? If surviving quantum suicide cannot be *unexpected* in the relevant sense, then it is unclear why it would be unexpected for one to record only '↑' in a long sequence of independent spin measurements. There does not appear to be any relevant physical difference between those circumstances. They are both produced by the same physical process, over which we (human beings) have the same degree of control. Why should the occurrence of the later outcome be a matter of greater objective uncertainty? Regardless of whether the usual method determines that the chance of survival is 80% or 20%, the world in which the participants survive exists *just the same*—and the "amplitude" of that world is not relevant to those participants. It is not clear why amplitudes should matter any more in an ordinary spin measurement. In short, if amplitudes can be ignored in the case of quantum suicide, then it is less clear why they should be crucially important in any other circumstance.

The problem that quantum suicide poses becomes more apparent when one considers how many-worlds theorists have sought to validate Born's rule. Recently, there have been several novel attempts to vindicate Everett's original claim that Born's rule can be derived or proven from the remaining quantum formalism. David Deutsch led this new wave with his *decision-theoretic proof* (1999), which has been defended and modified by David Wallace (2003b, 2007, 2010). Their arguments inspired Simon Saunders' *derivation from operational assumptions* (2003) and Hilary Greaves' adaptation of Bayesian arguments (2004, 2007; Greaves & Myrvold 2010). I will refer to these efforts collectively as the *decision-theoretic approach.*

The problem generated by quantum suicide translates easily into the language of decision theory. The dispositions of an *actor* are modeled in the theory by a *preference ordering* on a set of hypothetical *actions.* This relation is usually a *linear ordering*, so that it represents a determinate preference (or indifference) given a choice of several actions. Actions are often given intrinsic structure; e.g., a simple *wager* may be taken to consist of a *proposition* and the *outcome* if that proposition turns out to be true. Actions may also be composed of other actions, by undertaking multiple actions simultaneously or in a particular sequence. The principal application of this conceptual machinery is in modeling decisions made under conditions of uncertainty, and it has played a role in important analyses of the nature of probability and statistical reasoning. The foundational insight is that actions often cannot be distinguished by their consequences; that is, actions generally have indefinite outcomes. Thus, the inherent uncertainty of an action must be reflected in the preferences of a (rational) actor. The decision-theoretic approach to probability in MW aims to show that the preferences of a rational actor regarding actions *that cause worlds to split* must have the same structure as preferences regarding conventional actions with uncertain outcomes. Moreover, it is argued that Born's rule can be justified by showing that a rational actor is compelled to act as though they judge the likelihood of an outcome according to the usual method. It should be no surprise, then, that a situation in which rational expectations run contrary to the standard predictions might be problematic for the decision-theoretic approach.

We begin by considering the set of "quantum actions" that might be performed in a universe described by MW. Expectedly, there will be actions that involve branching events—we will refer to these generically as *measurements*. But the set of quantum actions should also include a variety of actions in which branching is not important, and we will refer to these as *quasi-classical actions*. The difference between the two is that the possible outcomes of a measurement must occur in distinct worlds, while the outcome of a quasi-classical action is not dependent on any particular branching event. For simplicity, we will suppose that any branches that emerge during a quasi-classical action realize nearly identical macroscopic states.[2] Now, there is no reason that a quasi-classical action couldn't involve *chance events* in the classical sense. Something like a coin-flip or throw of the dice need not produce any macroscopic branching. And there can certainly be ignorance regarding the outcome of a quasi-classical action—the action might be *ambiguous* in some way. Therefore, there are actions that can be performed in a splitting universe whose outcomes are uncertain in the most ordinary sense. In particular, one can perform *quasi-classical Russian roulette*; i.e., one can arrange for the triggering of a weapon to be contingent on the uncertain outcome of an quasi-classical action.

The problem that quantum suicide presents for the decision-theoretic approach can be compactly stated in the following claim.

QS: It is rational for an actor to prefer performing any instance of quantum suicide to performing any instance of quasi-classical Russian roulette.

While an actor is certain to survive quantum suicide unharmed, there is a chance that the actor is presently on a branch in which performing quasi-classical Russian roulette will be

---

[2] The decision-theoretic approach is already committed to the existence of quasi-classical actions. Events like *payoffs* or *rewards* are conceived of as non-branching actions, and it is also assumed that there are "actions" in which the macroscopic state does not change at all.

lethal. If so, then the actor can expect that their last experience will be of undertaking the relevant quasi-classical action. This is an undesirable outcome if there is any utility in having further experiences. According to classical decision theory, any risk of an undesirable outcome counts against an action; hence, there is reason to prefer quantum suicide to quasi-classical Russian roulette (other things being equal.) We could even strengthen QS if we like. Quantum suicide may be preferred to any action for which there is *some chance* that it is in fact an instance of quasi-classical Russian roulette, since this also implies that there is some risk that the actor will cease to exist.

If the preferences of a rational actor must reflect the uncertainty that the actor has regarding the outcome of an action, then QS implies that *a rational actor cannot have nontrivial credences regarding the outcome of both branching and quasi-classical events.* In other words, an actor that adheres to QS must either treat the outcome of measurements as certain or treat the outcome of quasi-classical events as certain. This claim is established by the following decision-theoretic argument: Both quantum suicide and quasi-classical Russian roulette have the form of a wager on the outcome of a neutral action. The action being wagered on is a measurement in the former case and a quasi-classical action in the later. If the outcome of either sort of event can be uncertain, then it can (presumably) be uncertain to any degree. Hence, there ought to be a measurement for which a particular outcome, lets call it *outcome A*, is less likely to occur than a particular outcome of a quasi-classical event, call it *outcome B*. We will write $0 < Pr(A) < Pr(B) < 1$. Consider two wagers: a wager that outcome A will occur and a wager that outcome B will occur, each with the payoff that our doomsday device remains inactive (as opposed to being triggered.) The first wager is an instance of quantum suicide and the second is an instance of quasi-classical Russian roulette. In this simple case the *expected utility*, E, of each wager is calculated with respect to an individual's particular *utility function*, U, according to the formulas

$$\begin{aligned} E(W_A) &= Pr(A)U(\ddot\smile) + (1 - Pr(A))U(\Omega)\,, \\ E(W_B) &= Pr(B)U(\ddot\smile) + (1 - Pr(B))U(\Omega)\,. \end{aligned}$$

It follows that, if $0 \leq Pr(A) < Pr(B) \leq 1$ and $U(\Omega) < U(\ddot\smile)$, then $E(W_A) < E(W_B)$. That is, if outcome A is less likely than outcome B and ceasing to exist has less utility than the status quo, then performing quasi-classical Russian roulette must have a *greater* *expected utility* than performing quantum suicide. If the preferences of a rational actor can be recovered from the expected utility of each action, understood as a ranking, then a rational actor will prefer performing that instance of quasi-classical Russian roulette to that instance of quantum suicide. But this contradicts QS. Therefore, either the preferences of an ideally rational actor (in the context of MW) is not necessarily representable by an expected utility function, or measurements and quasi-classical actions cannot both have uncertain outcomes. Either conclusion is problematic for the decision-theoretic approach.

The treatment of rational preferences among ambiguous actions is essential to the decision-theoretic analysis of probability, and the Everettians cannot ignore the implications of their approach in that domain. Nor can they accept the conclusion that one cannot make sense of rational action under conditions of ordinary ignorance in a universe described by MW. The aim of the decision-theoretic approach is to show that the *indeterminacy* due to an *actual* plurality of measurement outcomes must be regarded in the same way (for the purposes of rational action) as the *uncertainty* due to a plurality of *possible* outcomes. The quantum suicide thought experiment suggests that these two notions are incommensurable. Specifically, they cannot both be quantified and treated as *probabilities* in a calculation of expected utility, if QS is true. Therefore, the worry raised by quantum suicide is a problem for Everettian interpretations more generally. QS not only contradicts the standard quantum predictions, but it can be marshaled into an argument *against* attributing nontrivial probabilities (expectations, credences, weights, etc.) to the outcomes of measurements under MW. We have, in effect, just given such an argument in decision-theoretic terms.

Everettians must either find a way to deny that one is certain to survive quantum suicide, while maintaining the intelligibility of the theory, or explain how to reconcile Born's rule with that very different kind of prediction. The later task is made more difficult by the fact that the validity of Born's rule is already in question under all circumstances and not only in extraordinary ones. MW already suffers from a crisis regarding the notion of probability, and QS appears to confound the more promising attempts to address this problem. The more direct approach would be to deny that the predictions of MW differ from those of the standard formulation with respect to quantum suicide. In that case, the burden of proof is on critics to show that MW has not been misunderstood or misapplied.

So, we will next examine the reasoning that has led us to problematic conclusions.

3 Surviving Quantum Suicide

Before we lay out our line of reasoning in more detail, there is a straw man that should be immediately dispelled. Since death is nothingness, there is a sense in which one cannot *expect* to be dead. If there is nothing that *it is like* to be dead, then one cannot anticipate *being* dead. This is really just a matter of semantics—just as one might insist on omitting the funeral from someone's "life story." Yet it might appear that the argument turns on this very point—especially since commentators have said almost exactly the same while arguing that one should expect to survive quantum suicide. For example, Tegmark reasons as follows: "This time the [usual] recipe is inapplicable, *since probabilities have no meaning for an observer in the dead state* $|\overset{\mathrm{x}\,\mathrm{x}}{\frown}\rangle$, and the [interpretations] will differ in their predictions" (Tegmark 1998, p. 861; my italics). David Lewis lays out the same argument in full, using Schrödinger's cat as an example.

> What should the cat expect to experience, if it's a very smart cat and knows the set-up, and if it knows there are no collapses? The intensity rule says: expect branches according to their intensities. The intensities are equal. So the cat should equally expect to experience life and death.
>
> But that's nonsense! There's nothing it's like to be dead. Death is oblivion. (Real death, I mean. Afterlife is life, not death.) The experience of being dead should never be expected to any degree at all, because there is no such experience. So it seems that the intensity rule does not work for the life-and-death branching that the cat undergoes.
> (2004, pp. 16-7)

It appears that contingencies are being ruled out on the grounds that one cannot literally *witness* them. If this is the entirety of the argument, then it surely fails. The fact that one cannot have a conscious experience of death does not imply that one should expect immortality! That inference equivocates on two different meanings of 'expect', one that connotes conscious experience and one that does not. Moreover, this reasoning would clearly lead to absurd conclusions if applied to a classical universe, even one in which some kind of splitting regularly occurs.

Every student of metaphysics is familiar with various thought experiments in which individuals undergo classical "splitting." Derek Parfit's *split-brain* thought experiment (Parfit 1984, pp. 254-55) is well-known and apt for our purposes, since many-worlds theorists have commonly appealed to a Parfitian analysis of rationality (Papineau 1996, 2004; Saunders 1998; Wallace 2003; Greaves 2004; Tappenden 2004; Saunders & Wallace 2008). In the symmetric version, the two hemispheres of an individual's brain are separated and successfully transplanted into two identical bodies (from twin donors.) It is presumed that the hemispheres are functionally equivalent, so that both transplant recipients are *psychologically continuous* with the brain donor. We must modify this scenario in two ways to make it as analogous as possible to quantum suicide. First, we must stipulate that the very act of separating the two hemispheres instantly obliterates one of them. Second, we somehow arrange that the brain donor remains fully conscious throughout the (painless) procedure. Now there are several similarities between this thought experiment and quantum suicide. In such a circumstance, one wants to say that *fission* never really occurs—one has not *split* the brain so much as *destroyed* one half of it. Moreover, there is one seamless experience connecting the brain donor to the unique transplant recipient, and there is certainly no experience of being destroyed. The brain donor should, arguably, expect to become that unique recipient and to occupy the perspective of the surviving hemisphere throughout. However, it would be absurd to conclude that the brain donor should expect that *nothing is destroyed.* If the donor is informed about the procedure, then she will rightfully expect that half of her brain will be destroyed. Even if we consider what the donor should expect to experience, one really ought to clarify that her "cognitive apparatus" will undergo a severe reorganization and that, as a result, only half of it will be able to support conscious thought. The point is that, in a classical universe, one expects to survive such a procedure *despite* having half of one's brain destroyed; and this is very different from expecting to remain untouched because a weapon has remained inactive. There is an explicit disanalogy between the alleged prediction of MW and rational expectations in our modified split-brain thought experiment. The fact that *one does not expect to experience x* does not imply that *one should not expect x* in a classical universe. If the argument regarding quantum suicide succeeds at all, it must appeal to some further peculiarity of MW—and I contend that it does.

The additional factor regards the way in which *worlds* are supposed to enter into the *predictions* that MW makes. It turns out that this is a rather subtle matter. Superficially, the claim is that Born's rule still applies; i.e., each outcome should be expected according to the norm-squared amplitude of the corresponding component in the universal quantum state (the joint state of all systems involved) at that time.[3] But this cannot strictly be what is intended. One cannot be concerned with the *actual* amplitude of a world, since one could never be in a position to *know* the amplitude of a component within the universal quantum state. This is because worlds are constantly splitting. Worlds split when nobody is performing an experiment, they split when nobody is looking, and they have been splitting since before anyone had thought to keep track. Countless quantum systems are constantly interacting in ways that contribute to relatively stable macroscopic phenomena; and this means that there ought to be worlds in which all sorts of things that could have happened actually did happen. Even the approximate amplitude of a branch depends on myriad, unknown factors. Furthermore, if there is a universal quantum state, then the norm-squared amplitudes of all these worlds sum (approximately) to one. Consequently, we should expect the amplitude of any given world to be, numerically, exceedingly small. Born's rule, as stated above, is defunct in a universe described by MW.

At least two changes must be made to Born's rule, though they are only occasionally made explicit. First, the application of Born's rule must be restricted to the future worlds that are *causally connected* to the world that one presently inhabits, so that one does not form expectations regarding branches that have already effectively split from one's own. In other words, one must restrict attention to future worlds whose history is a continuation of one's own. The second change is to modify Born's rule so that it compels one to expect a future world according to the *ratio* of its norm-squared amplitude to that of the world they presently inhabit. The fully modified Born's rule reads:

Modified Born's Rule: Each outcome that is causally connected with one's present world should be expected according to the ratio of its norm-squared amplitude to that of the present world.

---

[3] According to the standard collapse formulation, the actual state of a quantum system "jumps" to a normalized eigenstate of the observed value the moment that a "measurement" is performed. This postulated *physical* process justifies the formal procedures of reduction and renormalization in the context of the standard formulation.

In practice, these changes amount to reducing and then renormalizing the quantum state, given what one knows about their own world, before applying Born's rule.

These steps must be taken so that one arrives at the same predictions as in the standard (collapse) formulation, but the many-worlds theorist must give a very different rationale for the procedure. Consider just the first step. Why shouldn't MW be taken to make predictions about sequences of events that diverged from one's own history sometime in the past? For instance, since there is an alternate history in which my (somewhat feeble) alarm clock stopped working before it could wake me, why shouldn't I expect to find myself (five minutes from now) still in bed? The answer to this question is closely related to the reason that one is not permitted to *expect* a state of affairs that combines facts from distinct worlds. One is not supposed to expect, for example, that both ↑ and ↓ will be attributed in a spin measurement, even though there are "equally real" worlds in which each of those facts obtain. This sort of expectation is inconsistent with the usual quantum probabilities; not only because ↑ and ↓ are supposed to be mutually exclusive, but also because one would expect these values to occur with a fixed frequency (i.e., *every* trial.)

Simon Saunders explains why it is a mistake to "mix worlds" when deriving predictions from MW in the following way:

> As Lewis and others have taught us, chance had better relate to credence. We give credence to what concerns us, our possible fates. Probabilities here, as in [standard] quantum mechanics, had better relate to what we care for, and to what we can perceive. So it is the probabilities for histories that matter, and the statistics that they encode, for that is all we can observe.
> (1998, p. 389)

As was noted earlier, the empirical content of quantum theory resides in the *objective chances* that can be attributed to outcomes, and this must still be the case under MW. Saunders points out that objective chances, particularly those based in a deterministic theory, are nonetheless said to be attached to a kind of *perspective.* The fact that objective chances are rightly assigned to coin-flips, weather models, Brownian motion, etc., is usually explained by citing the limited scale and scope of human experience. In the universe described by MW, branches are (for all practical purposes) causally isolated from each other; so that there is no sense in which our common experience or perspective could involve facts from worlds on distinct branches. Saunders takes this to imply that MW does not assign an objective chance to an event or history that crosses such boundaries; and, consequently, MW does not compel one to have the corresponding expectations. Hence, it would be improper to expect to suddenly find myself still in bed, with all the effects of my actions so far this morning (including my memories) totally absent, because that state of affairs breaks with the present perspective and does not contribute to an intelligible experience. This argument is in line with the approach that originated in Everett's doctoral thesis. Everett stressed that the nature of observation itself must be examined within the formalism in order to appreciate the theory's empirical content. His proposal was to read the predictions of the theory off of the entangled state of the relevant measurement record. It is in *this* way that the predictions of MW are meant to respect experiences. Other branches are not excluded because of the limitations of one's own experience *per se*; rather, other branches are excluded because they are not involved in *any relevant* experience.

Once again, the chaos represented by $|\mathbf{\Omega}\rangle$ is not a situation in which an experience is being had—human or otherwise. There is also no measurement record, apparatus, weapon, observer, nor any lasting record of their former existence. $|\mathbf{\Omega}\rangle$ does not describe anything of significance. It is not a situation involving anything that any of the participants in quantum suicide could care for, perceive, or otherwise observe. According to the same principle by which other branches are excluded, we conclude that *MW does not attach an objective chance or probability to* $|\mathbf{\Omega}\rangle$. Consequently, $|\mathbf{\Omega}\rangle$ does not represent anything that the participants in quantum suicide are compelled to expect.

There is another way to make this point that deals more directly with the physics behind the metaphysics of MW. In the most prominent and influential versions, worlds are understood as *emergent structures* whose emergence is due to a phenomenon known as *environmental decoherence*. In the study of environmental decoherence, measurement processes are analyzed in terms of the interaction of three systems: the object system, the apparatus (and/or observer), and the environment. The interaction between the object system and apparatus proceeds much as we have already described, causing some degree of freedom in the apparatus to become entangled with the state of the object system. But interactions with the ambient environment are now also considered. The environment will interact primarily with the macroscopic apparatus; e.g., through incident particles and radiation. In the simplest cases, the apparatus-environment interaction is constant and it dominates other interactions. The principal idea in the theory of environmental decoherence is that interactions with the environment select for observables that commute with the interaction Hamiltonian. Such observables are constants of motion under that interaction, so that those values are stable under the influence of the environment. This allows correlations among selected observables to persist, while other correlations are dispersed. Moreover, the constant entanglement of selected observables with the environment ensures that interference effects are negligible. The result is that the state of the apparatus can be decomposed in an *environment-selected eigenbasis* such that the components of the decomposition are dynamically isolated from each other and their respective evolutions are practically independent of the environment. David Wallace, whose position may be taken as representative of decoherence-based approaches, contends that these components qualify as *macroscopic worlds* precisely because they satisfy the following principle:

Dennett's criterion: A macro-object is a pattern, and the existence of a pattern as a real thing depends on the usefulness—in particular, the explanatory power and predictive reliability—of theories which admit that pattern in their ontology. (2003a, p. 93)

The universal quantum state (or some portion of it) is said to describe several distinct worlds because treating it as the state of an object-apparatus-…-environment system and keeping track of the components in the environment-selected eigenbasis dramatically improves one's ability to make predictions and otherwise use the theory. Without a *higher-order ontology* it would be immensely difficult to say anything about the universal quantum state under realistic conditions.

Wallace's principle implicitly refers to a kind of perspective, just as Saunders' analysis of objective chance does. The test of a higher-order ontology is that it is useful *to beings like us*, with our limitations. It may be hard to imagine a being without such limitations, but the principle nonetheless makes the ontology of the theory contingent on how it is going to be used. Returning to our example, it is presumably because environmental decoherence is at work that we can even talk about an object-apparatus-weapon system. In some small portion of the universal quantum state, a pattern has developed that can be modeled as an object-apparatus-weapon-environment system in which the environment selects a preferred basis. This pattern gives us a handle on that small portion of the universal state and allows us to track its development. More importantly, the many stable correlations permit us to talk about observers that are able to track their own development within this pattern. At least this much must be the case if some *real thing* is going to be able to *use* the theory. However, detonation of the doomsday device would disrupt this pattern. The weapon destroys the correlations among macroscopic objects; whats more, it scrambles the boundary between those objects and the environment, so that the formerly preferred observables are no longer useful. $|\mathbf{\Omega}\rangle$ is not part of the *object-apparatus-…-environment pattern*—it is certainly of no more use to the participants in quantum suicide than a typical component of the actual universal quantum state. Thus, from the hypothetical perspective of the participants in quantum suicide, $|\mathbf{\Omega}\rangle$ fails Dennett's criterion. Individuals who are in a position to perform quantum suicide have no reason to include a world corresponding to $|\mathbf{\Omega}\rangle$ in their ontology. Given this understanding of the metaphysic of MW, there is no contradiction is saying that *the doomsday device remains inactive in all worlds* when reading a prediction from the quantum state. If it is our higher-order ontology that ultimately matters, then participants in quantum suicide should fully expect to survive unharmed.

The last thing to note is that similar arguments can be made regarding other forms of quantum suicide. Our prototypical version has special properties that greatly simplify the situation, but the general form of the argument is the same. Due to the peculiarities of MW, its predictions must presuppose some perspective or relevant set of experiences; that is, a prediction must be read off of some form of measurement record. Additionally, there is no set of facts that the predictions from different perspectives must necessarily agree on—including the perspectives of observers on the same branch (consider a Wigner's friend type situation.) Even the metaphysic of worlds must take a particular set of apparata and observers into account. If the predictions of MW are not made relative to each observer (measurement record), then the very intelligibility of MW is threatened. Quantum suicide is a circumstance in which the functioning of a system *qua* observer is tied to a particular value of the quantity being observed. That value will be recorded by the observer, and only that value will be recorded. The claim is that the prediction that MW makes regarding *the perspective embodied by that observer* is contrary to the usual Born probabilities.

4 Conclusions

What I hope to have shown is that the task of demonstrating MW's empirical adequacy is at odds with the task of maintaining the intelligibility of MW's distinctive metaphysic. Of course, there is no question of which should be given priority. It would be self-defeating for many-worlds theorists to give up on vindicating Born's rule. The connection between quantum theory and the evidence that we have for it is a matter of statistical inference. Predictions must be drawn from unitary quantum mechanics in a manner that, in effect, corresponds to reducing and renormalizing the universal quantum state before applying Born's rule. This procedure has no obvious connection to actual physics (or metaphysics) under MW. We have argued that some of the considerations that enter into the justification of the modified Born's rule also entail that the predictions of MW regarding quantum suicide differ from those of the standard formulation—and, consequently, the modified Born's rule. This might be understood as an argument for *further* modifying Born's rule, but the conclusion is (more plausibly) a reason to reject the idea that the metaphysic of MW compels one to have nontrivial expectations regarding a measurement outcome.

This is particularly problematic for many-worlds proponents who have adopted the decision-theoretic approach. They have presupposed a metaphysic of many worlds and sought an analysis of objective probability within such a universe. All the more problematic is the fact that principles like QS are *consistent* with the assumptions of the various decision-theoretic arguments. We may grant those assumptions and still come to the same conclusion. Deutsch's, Wallace's, and Greaves' respective axioms and Saunders' operational assumptions constrain the ordering that one places on the set of all possible *unambiguous* quantum actions, including measurements; but none of them constrain the way in which such an ordering is extended over *ambiguous* actions (i.e., actions in which there is genuine ignorance regarding the outcome.) Moreover, QS entails that preferences among unambiguous actions should not be understood in terms of varying degrees of uncertainty. Such preferences *can* be accounted for as differences in the *utility* attached to each resulting collection of worlds, but this is no help to the decision-theoretic approach. There are very few constraints that can be placed on the utility function of a rational actor; that is, an actor is generally not compelled (under pain of *irrationality*) to value things in one way rather than another. Therefore, the quantum suicide thought experiment ultimately calls into question the assumptions of the decision-theoretic approach.

References

[1] Greaves, H. (2004). “Understanding Deutsch’s probability in a deterministic multiverse,” *Studies in History and Philosophy of Modern Physics*, 35, 423-56.

[2] Greaves, H. (2007). “On the Everettian epistemic problem,” *Studies in History and Philosophy of Modern Physics*, 38(1), 120-52.

[3] Greaves, H., and Wayne Myrvold. “Everett and evidence,” in [12], 264-304.

[4] Lewis, D. (2004). How many lives has Schrödinger’s Cat? *Australasian Journal of Philosophy*, 82(1), 3-22.

[5] Lewis, P. (2000). What is it like to be Schrödinger’s cat? *Analysis*, 60(1), 22-9.

[6] Papineau, D. (1996). Many minds are no worse than one. *The British Journal for the Philosophy of Science*, 47(2), 233-41.

[7] Papineau, D. (2003). Why you don’t want to get in the box with Schrödinger’s cat. *Analysis*, 63(1), 51-8.

[8] Papineau, D. (2004). David Lewis and Schrödinger’s cat. *Australasian Journal of Philosophy*, 82(1), 153-69.

[9] Saunders, S. (1998). Time, quantum mechanics, and probability. *Synthese*, 114, 373-404.

[10] Saunders, S. (2003). Derivation of the Born rule from operational assumptions. *Proceedings of the Royal Society A*, 460, 1771-88.

[11] Saunders, S. and David Wallace. (2008). Branching and Uncertainty. *The British Journal for the Philosophy of Science*, 59, 293-305.

[12] Saunders, S., Jonathan Barrett, Adrian Kent, and David Wallace, eds. (2010). *Many worlds? Everett, quantum theory, and reality.* Oxford: Oxford University Press.

[13] Tappenden, P. (2004). The ins and outs of Schrödinger's cat box: A response to Papineau. *Analysis*, 64(2), 157-64.

[14] Tappenden, P. (2011). Evidence and uncertainty in Everett's multiverse. *The British Journal for the Philosophy of Science*, 62, 99-123.

[15] Tegmark, M. (1998). The interpretation of quantum mechanics: Many worlds or many words? *Fortschritte Der Physik*, 46, 855-62.

[16] Wallace, D. (2003a). Everett and structure. *Studies in History and Philosophy of Modern Physics*, 34, 87-105.

[17] Wallace, D. (2003b). Everettian rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studied in History and Philosophy of Modern Physics*, 34(3), 415-39.

[18] Wallace, D. (2007). Quantum probability from subjective likelihood: Improving on Deutsch's proof of the probability rule. *Studies in History and Philosophy of Modern Physics*, 38(2), 311-32.

[19] Wallace, D. (2010). "How to Prove the Born Rule," in [12], 227-263.