

Theology and Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rtas20

Many Worlds and Moral Responsibility

Emily Qureshi-Hurst

To cite this article: Emily Qureshi-Hurst (2024) Many Worlds and Moral Responsibility, Theology and Science, 22:3, 456-473, DOI: 10.1080/14746700.2024.2359187

To link to this article: https://doi.org/10.1080/14746700.2024.2359187

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 24 Jul 2024.

Submit your article to this journal

Article views: 1574

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rtas20

THEOLOGY AND SCIENCE
2024, VOL. 22, NO. 3, 456–473
https://doi.org/10.1080/14746700.2024.2359187

OPEN ACCESS Check for updates

Many Worlds and Moral Responsibility

Emily Qureshi-Hurst

Pembroke College, University of Oxford, Oxford, UK

ABSTRACT

Moral responsibility is integral to most forms of theism. Yet its coherence as a concept is contingent upon the fulfilment of certain metaphysical conditions. In this paper I argue that (1) the continued existence of the self, and (2) freedom of will and action, are necessary conditions for moral responsibility, particularly within a theistic context. I then argue that the Everett Interpretation of Quantum Mechanics (or, the Many Worlds Interpretation) throws these assumptions into question in new and interesting ways.

KEYWORDS

Many worlds; Everett; Moral responsibility; Personal identity; Free will

Our understanding both of moral value and our responsibility to uphold that value is shaped by our experience of the world around us. Most ordinary people's moral intuitions about right action—whether some action was freely made, whether it accords with shared moral principles, and whether a person can be held responsible for it—have developed in the context of a(n assumed) singular universe. But what if the structure of reality was radically different from what intuition and ordinary experience tell us? According to the Everett interpretation of Quantum Mechanics (aka the Many Worlds Interpretation, henceforth MWI), reality is more complex and riddled with more multiplicity than quotidian existence could possibly suggest. If Everett was right about how to interpret the quantum formalism, then we are actually living in a world in which each of us has many alternate futures laid before us. All possible outcomes of all quantum events are actual, with each outcome becoming realised in a separate branch of the universal wavefunction. If we do away with a singular universe and a unique enduring self, what should become of the ordinary moral practices that emerged in the context of that (assumed) singular universe?

I argue in this paper that MWI raises specific ethical problems that other interpretations may not face. In fact, I will argue that MWI is not straightforwardly able to satisfy two conditions typically deemed necessary for our theistic notion of moral responsibility to apply. As such, more work is required to shore-up the concept of moral responsibility if it is to remain credible in light of Everettian QM. Although these issues have been discussed before, Everettianism raises them in new ways and may well block some of the standard responses to these problems (such as denying the possibility of fission events or denying that the world really is deterministic). MWI is rapidly rising in popularity. Insofar as

CONTACT Emily Qureshi-Hurst emily.qureshihurst@gmail.com

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

moral responsibility is integral to theism, establishing its coherence within an Everettian framework is essential. My aim, then, is to show how MWI causes acute problems for the theistic worldview in ways that other, singular, universe models do not.

Very generally, when we say an individual is morally responsible for a certain action, we are saying that they possess certain capacities (including volition, control over one's actions, reasonable foresight as to the consequences of a given action, etc …) and have willingly exercised these capacities to perform a particular action. The individual is then praiseworthy if the action is deemed right or good and blameworthy if the action is deemed wrong or bad. This is an important feature of the theistic worldview which is committed to the idea that God passes judgement—praise and blame—on individuals for their actions. For example, "God will bring every deed into judgment, including every hidden thing, whether it is good or evil" (Ecclesiastes 12:14). Importantly, God only passes judgement on beings who fall under the class of "moral agents," defined by the leading living philosopher of religion Richard Swinburne as beings with "moral awareness and limited free will, power, and knowledge."[1] He calls such beings "humanly free agents," believing that only humans possess the kinds of morally relevant capacities with which we are concerned.

Exactly which capacities might we be talking about here? There is a healthy amount of disagreement amongst philosophers with regards to this question. I want to suggest the following two necessary (but perhaps not sufficient) conditions that must be satisfied for an individual to be held (justly) responsible for their actions.[2]

Some person x is morally responsible for a given act a if:

(1) x is numerically identical with the individual who committed a.
(2) x freely chose to act as they did.

(1) is, in my view, obviously necessary. It is wrong to hold someone morally responsible for the actions of someone else.[3] (2), on the other hand, has been the subject of greater controversy in philosophical circles. Nevertheless, human freedom is an important aspect of theological anthropology, as emphasised by Swinburne:

> God has good reason to cause the existence of … humanly free agents who would need to have bodies through which to have enjoyable sensations, form largely true beliefs about the world, and form their own purposes in the light of these beliefs, which would make a difference to the world.[4]

Swinburne argues that freedom and embodied personhood are essential for us to be able to choose to love God, to choose to do good things for the world in which we live, and also to have the capacity to reject God and perform morally bad actions. Swinburne is representative of the broader Christian position here. In general, Christian theism is committed to the existence of *free* agents who can live temporally extended, meaningful, lives and make decisions about how those lives play out. They can then be held responsible for the actions they take. As we shall see, Everettian QM throws this idea into question.

A Whistle-Stop Tour of Everettian Quantum Mechanics

The Everett interpretation of Quantum Mechanics (sometimes also called "Many Worlds," though never by Everett himself) emerged in a doctoral dissertation written

by Hugh Everett III in the 1950s. The self-proclaimed aim of Everett's dissertation was "to regard pure wave mechanics as a complete theory." In so doing, "it postulates that a wave function that obeys a linear wave equation everywhere and at all times supplies a complete mathematical model for every isolated physical system without exception."[5] What follows is both simple and radical: fundamentally, reality is comprised only of the wavefunction of the universe, and this universal wavefunction evolves deterministically in accordance with the Schrödinger equation. In other words, what follows is a commitment to wavefunction monism. Moreover, at a quantum event the wavefunction splits into branches that contain each possible outcome of the event in question and cease interacting with each other.[6]

Simon Saunders suggests that, on the Everett interpretation, reality is best described by a "block universe" model that contains many alternative "pasts" and "futures" obtaining in non-interacting, quasiclassical, domains.[7] This point is echoed by Paul Tappenden.[8] On this view, our observational perspective (viz. our particular "branch") determines only which outcome we observe—all other possible outcomes are equally (albeit non-accessibly) real. It has been argued in recent years that the dynamical process known as decoherence causes these branches to cease interacting with each other, leading to the apparent existence of many worlds that appear to behave classically when viewed from within the perspective each. Effectively, the quantum state of our universe can be represented as containing superposed quasiclassical worlds.[9]

There is some disagreement amongst Everettians about whether the worlds posited by the theory are emergent or fundamental. These views are sometimes called the Conservative View and the Radical View respectively. The Radical View holds that at a quantum event the universe splits into two or more worlds each of which contain one possible outcome of the event in question.[10] The Conservative View, defended in recent years by David Wallace and Simon Saunders, claims that the worlds are emergent, macroscopic, phenomena that nonetheless ultimately depend on the same single, evolving, quantum state. On this view, world-branches are not a fundamental part of physical reality. Instead, these worlds emerge from the configuration of a set of sub-atomic particles. So, according to MWI, quantum events cause the wavefunction of the universe to branch, and these branches contain distinct worlds much like our own. Wallace addresses the reasonable question that follows from this claim, namely "how many branches are there?":

> Branching is caused by any process [or, quantum event] which magnifies microscopic superpositions up to the level where decoherence kicks in, and there are basically three such processes:
>
> 1. Deliberate human experiments: Schrödinger's cat, the two-slit experiment, Geiger counters, and the like.
>
> 2. "Natural quantum measurements," such as occur when radiation causes cell mutation.
>
> 3. Classically chaotic processes, which cause small variations in initial conditions to grow exponentially, and so which cause quantum states which are initially spread over small regions in phase space to spread over macroscopically large ones.[11]

Whilst the first is rather rare, the second two are ubiquitous. So much so, that the question "how many branches are there?" cannot generate a clear answer. We cannot point to

an exact moment at which a branch emerges and ceases to interact with any of the others, leading Wallace to liken the question "how many branches are there" to the question "how many experiences did you have yesterday." It makes sense to say that you had many, just not how many. Our best answer is *a lot*.[12]

Critics may ask: why we should posit a non-denumerable infinity of worlds to explain the particular features of the one world in which we find ourselves? Does such a move not make a mockery of Occam's razor? Everettians respond that MWI is no more than quantum mechanics itself, taken literally as a direct description of the physical world.[13] It may seem metaphysically weighty, but by being the simplest interpretation of the formalism, it is mathematically lightweight. Indeed, according to its proponents, Everettianism alone possesses the hermeneutical virtue of taking QM seriously. See Tim Maudlin's review of Wallace's book for a critique of this claim.[14]

The above details the bare bones of the theory, with brief mention of the hermeneutical uncertainties that persist therein. Whether the Radical or Conservative view is correct is orthogonal to the present discussion. For our present purposes, I will proceed by analysing the Radical version of Everettian QM. The reason for doing so is simple—if the problems identified can be overcome within the more radical reading of the Everett interpretation, then the problem is likely to also be solved (or solvable) in the less radical reading. According to the Radical version of Everettian QM, the universe splits or branches. Because of this, reality is comprised of ever more worlds or world-branches containing many individuals who share a past with the version of us in this world but not a future.

The Many Worlds Interpretation, thus understood, is worth taking very seriously. Science and Religion discourse has been woefully limited in its treatment of Quantum Mechanics, engaging almost exclusively with the Copenhagen interpretation. It is time to turn our attention to other interpretations, particularly as those interpretations grow in credibility as is the case with Everettian QM. By using the metaphysical picture provided by this interpretation of QM as our set and setting, we are able to examine whether, and to what extent, traditional theistic ideas can be accommodated within this framework. We can now turn to this paper's central question: *can we recover the theistic notion of moral responsibility in an Everettian universe?*

Requirements for Responsibility

Holding individuals morally responsible, namely making a judgement that some individual is praiseworthy or blameworthy for their actions, is a ubiquitous human practice. It is also a cornerstone of theism. We feel gratitude, awe, and respect for those who go above and beyond to do good; we resent those whose misdeeds we view as contemptable. In other words, we consider ourselves and others to be praiseworthy for good actions and blameworthy for bad actions if and to the extent that we are morally responsible for those actions. An individual who commits wilful murder is sent to prison; an individual who risks their life to save their fellow troops in battle is given a medal. This is echoed in Scripture. For example "we must all appear before the judgment seat of Christ, so that each one may receive what is due for what he has done in the body, whether good or evil." (2 Corinthians 5:10). The parable of the sheep and the goats in the Gospel of Matthew

also speaks of sorting individuals with respect to their moral behaviour. Thus, whether Everettian QM can provide such foundations is a highly important question.

Earlier, I identified two necessary conditions for moral responsibility. An individual *x* is morally responsible for a given action *a* if:

(1) *x* is numerically identical with the individual who committed *a*.
(2) *x* freely chose to act as they did.

Together these conditions ensure that the individual *x* is held morally responsible for some act *a* if they are the individual who is *causally* responsible (i.e. they are appropriately situated in the chain of causes that led to the given event) and their causal responsibility can be appropriately attributed to their free decision to perform a certain action.

Although causal and moral responsibility are frequently intertwined, they ought not to be conflated. This will become evident in the following set of examples. Consider initially two versions of a single event: my elderly grandmother falling to the floor and injuring herself at a family birthday party. First, I slip on a banana skin and fall into my grandmother, knocking her over. Second, I deliberately push my grandmother over because she ate the last piece of birthday cake. In each case, I am causally responsible for my grandmother falling to the ground because I am numerically identical with (viz. the exact same person as) the individual whose action resulted in my grandmother falling to the ground. Thus, causal responsibility is mine in each case. Only in the second case, however, am I morally responsible. If I slip and fall on someone, I am the cause of their injuries, but only accidentally so. Something additional needs to be included to make me *morally* responsible. That additional factor is volition, or free and deliberate action in accordance with one's will. I need to freely decide to push her, and then act in accordance with my free decision to bring *a* about. Moreover, it must be *me* who pushes her, not someone who looks very much like me (e.g. my identical twin) in order for the responsibility to be *mine*.[15]

Both (1) and (2) are essential for moral responsibility. Indeed, we might say that (2) is the criterion that distinguishes between causal and moral responsibility. This was indicated in the difference between the first and second version of events. If I am not free to decide on some action, and to act in such a way that I bring it about, then I cannot be morally responsible it. This holds for the version of events in which I slip on a banana skin and fall on my grandmother because I did not freely or intentionally choose to knock her over. Instead, I did so by accident. It would also be the case if I pushed her over because someone had a gun to my head, threatening me in a way I could not refuse. So, to be responsible, one needs to be numerically identical with the individual who committed the act and needs to have acted freely. We are now able to set nefarious family birthday parties aside and explore whether conditions (1) and (2) can be met in an Everettian world.

(1) The Problem of Persistence

Whether I am the same person throughout the various stages of my life is an age-old philosophical question, holding significance in a universe ruled by multiplicity *or* singularity. Nonetheless, the problem is brought into particularly stark relief when considering

the nature of moral responsibility in MWI. In the following subsections, I will explore whether condition (1) for moral responsibility can be satisfied in MWI.

(a) How do we persist through time?

It is an almost universal intuition that we continue to exist through time, i.e. that I am the same person as the five-year-old child I remember being and I will be the same person when I am in my eighties. Even if my personality changes, I have not ceased to be *me*. Christians may think of this in terms of a soul—I am the same individual as my childhood self because, although I have undergone much change, I still have the same soul. Philosophers have long reflected on this issue, and this discourse is illuminating in the context of the problem at hand. As well as its relevance for moral responsibility, personal identity is also important for other theistic reasons. Theism requires that individuals develop personal relationships with God across a lifetime—to maintain and cultivate relationships across time, I must be the same person throughout my life.

There are numerous proposals for the persistence conditions of transtemporal identity (in other words, necessary conditions for one's continued existence across time). Broadly, they can be split into (a) psychological continuity views and (b) bodily continuity views. The former holds that in order for an individual to continue existing, they must be psychologically continuous (either directly or indirectly) with some past person. This may include having direct memories of being some past person, or indirect memories of being previous persons who remember being previous persons. Thus, I may be deemed psychologically continuous, and thus identical with, a five-year-old either if I remember being that child or if I remember being, say, a twelve-year-old and that twelve-year-old remembers being the five-year-old.

Nevertheless, problems arise with using memory as a persistence condition, as memories are notoriously fallible. We do not remember a great many of the things that happen to us. I don't remember what I had for lunch last Tuesday, but it seems far-fetched to say, on that basis, that I am not the same person as whoever ate that lunch last Tuesday. Other psychological criteria that hold more promise refer to continuity of first-person experience. I am the same person as the person who ate that lunch last Tuesday if I have a continuous stream of first-person experience that connects me now to me then. Although one might assume this is problematised by sleep, it can be augmented to accommodate periods of unconsciousness. Psychological continuity views preserve the intuition that we would survive if our consciousness were uploaded to a computer or if we woke up in a different body, but seem incompatible with the claim that I was a foetus or could ever enter into a persistent vegetative state, as neither of these involve consciousness in any recognisable sense.

Bodily continuity views propose physical persistence conditions. Such views claim that an individual continues to exist if they share bodily continuity with a past person; this can be traced back to each stage of their development (including pre-natal). Bodily continuity views are typically founded on the idea that we are fundamentally physical, embodied, human animals whose identity is inextricably linked to our bodies. As such, these views easily accommodate the intuition that each of us was once a foetus, and that your elderly grandmother who has succumbed to dementia still remains your grandmother (even if she has undergone many significant changes in personality).

Nonetheless, bodily continuity views cannot make sense of the idea that one's consciousness could ever be uploaded onto a machine. Depending on your own predilections, this could be either an asset or a pitfall.

A helpful way of ascertaining which view better accords with our intuitions, or which is best able to stand up to scrutiny, is to conduct thought experiments that consider various situations in which identity might be lost or preserved. A seminal set of thought experiments of particular relevance to the issue at hand are Derek Parfit's "fission cases." Fission cases were originally inspired by cell fission, a process whereby a cell splits into two or more identical copies of itself. These thought experiments are intended to examine whether personal identity could be preserved in such a scenario. Although fission cases have typically been viewed as helpful but unrealisable intellectual exercises, MWI raises the very real possibility that human fission may have been part of reality all along.

(b) Fission Cases

Discourse on fission cases typically begins by imagining an individual splitting into two persons who seem to be exact copies (both physically and psychologically) of the pre-fission individual. More formally, the problem can be set out in the following way: let *a* be a being who will at some future time *t* undergo fission, and let *l* and *r* be those two persons who exist post-fission (sometimes referred to as *lefty* and *righty*). The question is: after *t,* is *a* identical to *lefty, righty,* both, or neither?

According to Leibniz' law, if two objects are numerically identical then they must have the exact same properties. Identity is a transitive relation, meaning if *a* is identical with *lefty* and *lefty* is identical with *righty,* then *a* must be identical with *righty.* Yet in this case, *lefty* and *righty* are clearly not numerically identical with each other. They have a multiplicity of different properties, including having distinct streams of first-person experience, occupying different locations in space, and being able to move independently of each other. As such, *lefty* and *righty* do not share all the same properties, and so are different persons. On this basis, *a* can be identical to neither of them. Parfit echoes this sentiment, writing that "identity is a one-one relation," and so collapses into incoherence when branching occurs.[16] According to Leibniz' law, then, fission brings about the end of *a*'s existence.

This conclusion is unproblematic for most bodily continuity theorists, particularly if *a*'s original body is gone. If *a*'s original body provides the biological material for the second two bodies, perhaps coming to this conclusion is a little more complicated. Yet in all cases in which *a*'s original body is destroyed, *a* is dead. The psychological continuity theorist, however, faces bigger problems. They must conclude that both *lefty* and *righty* are numerically identical with *a* as each shares psychological continuity with *a*, despite the fact that *lefty* and *righty* are not identical with each other. Various solutions have been offered.

(c) Who Am I in Many Worlds?

The first solution, the multiple occupancy view, developed out of ideas first presented by David Lewis.[17] Wallace and Saunders later developed a semantic argument about how

this could play out in MWI.[18] On the multiple occupancy view, an individual can survive branching because before the branching event their body actually contained two (or more) distinct persons or person stages whose futures deviate at the time of branching. Lewis claimed that persons are made up of an aggregate of such temporal parts or stages, and if an individual will at some future time undergo branching into two copies, then two people have always existed. One can attribute two sets of thoughts to the pre-split person stages, as there are actually two persons who later branch.

Derek Parfit rejects this view on the basis that it "greatly distorts the concept of a person."[19] The idea that our bodies contain the minds of more than one person at any given time profoundly clashes with our first-person experience of ourselves as singular and unique. This is made even more stark by the incomprehensible multiplicities involved in Everettian-type fission. On this issue, physicist Sean Carroll writes:

> We don't know how often branching happens, or even whether that's a sensible question to ask … but we do know that there is a lot of branching going on; it happens every time a quantum system in a superposition becomes entangled with the environment. In a typical human body, about 5,000 atoms undergo radioactive decay every second. If decay branches the wavefunction in two, that's 2^{5000} new branches every second.[20]

And this is only a fraction of it—the quantum events that occur within us are not limited to radioactive decay. Indeed, instances of branching may well be approaching an infinite number of times per second. If branching occurs as frequently as this, then according to the multiple occupancy view, I contain a *non-denumerable infinity* of persons / distinct person stages at any one time. This is metaphysically curious to say the least. It is also impossible to reconcile with the idea that human beings are bodies with a single, unique, soul.

I also cease to be able to say who I am at any given moment. Presuming I am a single, unique, individual, I may reasonably ask which of the multiple occupants of my body is actually *me*? On this question, Paul Tappenden writes: "an utterance of 'I' in Lewisian contexts of multiple utterance cannot straightforwardly be assumed to refer to the utterer. The semantic rule ['the word "I" refers to the speaker in any sentence in which it occurs'] cannot be adopted without further ado."[21] The multiple occupancy view cannot accommodate the ordinary notion of an individual referring to themselves. The word "I," then, ceases to have self-referential certainty. This is a significant problem for the multiple occupancy view. I suggest that we would need an extremely high threshold of verification to accept such a profoundly counter-intuitive conclusion. At the present time, the evidence simply isn't there.

An alternative response to the problem of fission is the so-called no-branching view. It holds that identity is an equivalence relation, meaning it is reflexive, symmetrical, and transitive, so it must be only a one-one relation between something and itself. Thus, it introduces a *no branching* condition into the psychological continuity view. In so doing, it manages to circumvent the above concerns regarding overpopulation and self-referential uncertainty by ruling out the possibility that multiple people can be identical with past persons and with each other. On this view, an individual can survive a life-changing event like fission only if no branching (of the individual person) occurs. In effect, *a* continues to exist only in cases in which there is no other identical individual with an equal claim to being *a*. The problem is, on this view I cannot survive the

branching that occurs in MWI. Any post-split persons are not *a*; they are instead new persons who emerge like a phoenix from *a*'s ashes.

This may work well for ordinary fission cases which would be one-off and extremely rare, as it rules out two people having an equal claim to *a*'s job, money, partner, house, etc. Yet if MWI is the case then fission is ubiquitous, thus we are forced to accept the deeply counter-intuitive conclusion that the five-year-old child I remember being is neither me nor still alive. Though I am psychologically continuous with the child and have watched that child's body grow into my own over the twenty-something years that divide us, she has not managed to survive the many branching events that have occurred. Furthermore, the no-branching view holds that each person with whom I am psychologically continuous only exists for a very short time (according to Carroll, it may be no more than a fraction of a second). On this view, I am not the same philosopher who began writing this paper, even though I remember thinking through the ideas, deciding upon a structure, and have had the uninterrupted experience of typing out each word.

A final option worth considering is offered by Parfit, who argues that philosophers are mistaken in thinking that personal identity is what is most important in fission cases. We should not ask "am I identical with some past or future person," but "*have I survived?*" Whilst numerical identity is, by definition, an all-or-nothing phenomenon, most of what we tend to think of as being involved in survival admit of degrees.[22] Thus, Parfit argues that after *t*, *a* is identical neither with *lefty* nor *righty*, but that this does not matter. What really matters is *a*'s survival, even if the person that survives the fission event (or any other relevant life changing experience) is not numerically identical with *a*, and "we ought to regard division as being about as good as ordinary survival."[23]

Carroll tries to make sense of this in the context of MWI, arguing that the idea that each of us is the exact same person from birth to death was never more than a useful approximation anyway. Everettian QM, he argues, encourages us to embrace the idea that "the lifespan of a person should be thought of as a branching tree, with multiple individuals at any one time, rather than a single trajectory—much like a splitting amoeba."[24] So the human person ought to be thought of like a phylogenetic tree, namely a set of descendants branching off from each other *ad infinitum* but sharing a common ancestor with whom all future persons are psychologically continuous.

According to Parfit, although identity is not preserved in fission, everything that matters for morality is. The solution, then, is not to show how condition (1) can hold in MWI, but to deny that condition (1) is in fact necessary for moral responsibility. Instead, what matters is psychological connectedness. If this is the case, then perhaps moral responsibility can be recovered in the face of Everett's infinite fission events. It appears to us, from within the perspective of our individual world-branch, that there is only one unique being (me) who is psychologically continuous with past me, and so I experience my own survival throughout the day-to-day branching events of my life. Thus, for all practical purposes, we may be able to conclude that persons survive branching and can continue to be deemed morally responsible for their behaviour. It remains the case, however, that numerical identity cannot be preserved in MWI as fission constitutes a violation of Leibniz's law.

If survival is the best that we can do in MWI, we must ask whether survival is enough to satisfy condition (1) and recover moral responsibility. I argue that it does not seem

robust enough to give us much of what identity means in a theological context. As Psalm 139:13 says of God, "you created my inmost being; you knit me together in my mother's womb." Christianity holds that there is something essential about your "inmost being" that was created by God and stays with you throughout your life. Without numerical identity, this loses some of its significance. Moreover, for those who believe in the existence of an immortal soul (which, by virtue of its immortality, presumably does not get destroyed) it is not really clear what it might mean to survive branching but cease to be identical with your past self. It seems to me that even if certain secular solutions work for the problem of fission, these may not work in the context of theism. More work is required to establish a theologically robust model of personal identity that can withstand the threat of Everettianism. Now that we have considered condition (1), we can move on to consideration of condition (2).

(2) Freedom and Determinism
(a) Determinism and Indeterminism

The scientific paradigm provided by Newtonian mechanics appeared to entail causal determinism. Causal determinism can be defined counterfactually: given the state of the universe immediately prior to some event e + the laws of nature, no event other than e could have occurred. In modal terms, there is no possible world in which both the laws of nature and the past are identical to our universe but where the future is different. The quantum revolution's departure from the Newtonian paradigm led the scientific community toward the opposing view: indeterminism.

The Schrödinger equation allows us to describe the wave function of a superposed quantum object as it evolves deterministically through time. Essentially, an unobserved quantum object exists in a superposition, meaning it occupies all possible states. According to textbook QM, performing a measurement fundamentally alters the system, causing the wave function to collapse into a particular state and actualise one of a range of possible outcomes. The outcome of such a quantum event can only be predicted within a probability range that is spread out over the superposition, with various states being more likely than others but none being certain. Textbook QM, sometimes referred to as the Copenhagen interpretation, claims that this indeterminacy is ontological, meaning it reflects some mind-independent feature of reality and not an epistemic gap that will be filled with future scientific development. Thus, the Copenhagen interpretation paints an indeterministic picture of the quantum world.

The Everett interpretation, however, claims that the universe can be described *in its totality* by the Schrödinger equation, according to which the wavefunction (in this case, of the universe) evolves deterministically. In MWI, there is no collapse mechanism that forces a measured quantum object to occupy a single state. Instead, all possible outcomes actually occur, albeit in various non-interacting branches of the universal wavefunction. Two things can be gleaned from this. First, from a God's-eye perspective, the universe is deterministic: the outcome of an experiment can be predicted with certainty because its outcome is fixed by deterministic laws. Second, however, is that from *within* the quasi-classical non-interacting branches, the universe appears *in*deterministic. The outcome of a given experiment in your branch cannot be predicted with any certainty

greater than can be provided by the probabilistic equations of textbook QM. So the metaphysical picture Everett paints is a complex one. The universe is deterministic, but from within the perspective of a single branch it appears indeterministic.

Tim Maudlin explores this in his (rather critical) review of Wallace's 2012 book:

> Since [the Everett interpretation] is deterministic, one knows beforehand what will happen in a given situation: if you make an "x-spin measurement" on a y-spin up electron, for example, with certainty the universe will "split" into equally weighted branches, one for each "outcome." And if different things happen on these different branches—if, for example, on one branch your successor gets a dollar and on the other your successor loses a dollar—then there is no interesting sense in which you can "bet" on the outcome. If only one alternative will occur and you don't know which, then you can bet, and you will either win or lose the bet. You might, for example, be willing to pay a dollar for such a bet. But in an Everettian universe, you just are not betting in this sense: you are rather purchasing gains or losses for your successors, which they will certainly receive.[25]

According to Maudlin, if you place a bet on a quantum experiment, you are merely buying gains or losses for future versions of you. Although this is an odd form of determinism, it is determinism nonetheless. What are the implications of this for free will and moral responsibility? In what follows I will explore a compatibilist and then an incompatiblist answer.

(b) Freedom and the Principle of Alternate Possibilities

One way we might articulate the challenge to freedom (and thus moral responsibility) from causal determinism is via what has come to be known as the Consequence Argument. Peter Van Inwagen puts the problem this way:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.[26]

If causal determinism is true, then (laws of nature) + (the past state of the universe) entails one future. But both the laws of nature and the past state of the universe are resolutely out of our control. In fact, each of them was set millennia before our birth. If all our actions amount to are the immovable consequences of the laws of nature and the distant past, then we are not free. If we are not free, then according to the conditions set out above, we cannot be held morally responsible for our actions. This claim is integral to Christian theism, as Swinburne emphasises via his sustained commitment to libertarian freedom and its inherent connection with moral responsibility. Indeed he argues that having libertarian free will and holding moral beliefs (i.e. beliefs about what is right and wrong that guide our actions) are sufficient conditions for moral responsibility. Thus agents are responsible for their free, intentional, well-informed actions.[27]

In order to assess whether Van Inwagen is right to conclude that we cannot be free in a causally deterministic universe we must unpack the concept of freedom. What do we mean by freedom of both will and action? Typically, at least in the context of moral responsibility, we mean some form of *control* over both our actions and the desires that lead us to act. When identifying freedom as a necessary condition for responsibility

I (and many others who have felt similarly) am ruling out actions that are a result of force, coercion, deception, hallucination, or compulsion, and including all actions that are a result of choice, deliberation, volition, and intention.

Fischer and Ravizza suggest that we may appropriately hold an agent morally responsible for a given action unless one of two negative conditions are satisfied, the "epistemic condition" and the "freedom-relevant condition," and that neither of these require determinism to be false.[28] Thus, we begin with a compatibilist approach to solving the problem of free will, determinism, and moral responsibility. The former condition rules out moral responsibility if the individual did not possess knowledge of the relevant facts at the time of acting, such as causing someone to go into anaphylaxis by cooking them a dish containing an unknown allergen. The latter condition rules out moral responsibility in cases where someone was forced (by threat, hypnotism, coercion, or similar) into acting. To be morally responsible, then, an agent must know the relevant facts surrounding their action and have acted with the right sort of beliefs and intentions. They must also have exercised the appropriate type of control over their action, i.e. they must be free in the relevant sense.

One of the ways this freedom has typically been understood is as the freedom to do otherwise. This is sometimes referred to as the Principle of Alternate Possibilities. More specifically, *x* chose *a* freely iff, at some time prior to the choosing of *a, x* could have chosen *b*. I chose to have toast for breakfast this morning. This choice was free, according to the Principle of Alternate Possibilities, iff I could have chosen to have cereal or fruit instead. The Principle of Alternate Possibilities (PAP) was taken as a given until the work of Harry Frankfurt in the mid twentieth century.[29] Frankfurt's aim was to undermine the credibility of PAP by constructing a set of cases in which PAP is violated, meaning the individual in question could not have done otherwise, and yet we would still consider them morally responsible.

A paradigm (albeit adapted) Frankfurt case is as follows: Jones is planning on assassinating his local mayor. He has deep ideological differences with the mayor and believes firmly that his city will be better off with this mayor gone. Black agrees with Jones but has doubts about Jones' tenacity and is concerned he will back out at the last minute. Unbeknownst to Jones, Black implants a device in Jones' brain that will monitor Jones' thoughts and will be able to tell if Jones considers changing his mind. At this point, Black would be able to override Jones' brain and make him carry out the assassination. Jones, however, has no crisis of confidence and successfully carries out the assassination. In this case, although he did not know it, Jones could not have done otherwise. Black would have forced him to carry out the assassination had he changed his mind, so PAP was violated. Nevertheless, we have strong intuitions that Jones is straightforwardly and entirely blameworthy for carrying out the assassination.

Although such a state of affairs is currently a technological impossibility, the thought experiment reveals inconsistencies within the claim that PAP is a necessary condition for free action and thus moral responsibility. Frankfurt disambiguates two interrelated notions involved in the application of PAP in moral discourse. A violation of PAP involves a metaphysical fact: *x* could not have done otherwise. But this metaphysical fact alone is irrelevant for moral responsibility. For moral responsibility we need the second notion: motivation. What matters are the reasons that motivated *x* to act as

they did. For PAP to be violated in such a way that moral responsibility ceases to apply, the metaphysical fact must be integral in *x*'s reasons for performing the action. In other words, *x* is only freed of moral responsibility if *x* acted *precisely because they could not do otherwise*, perhaps due to coercion or force of some kind. If they acted for their own reasons, it is irrelevant whether they could have acted otherwise had their motivations changed. Prior to Frankfurt's work, the metaphysical fact was believed to be all that was required, and indeed that is all that PAP contains. But PAP requires the metaphysical fact to have motivational force if it is to be relevant to whether *x* can be deemed moral responsible.

We can apply Frankfurt cases to determinism in the following way. While determinism may make an agent's action unavoidable, it does not follow that the agent acts as he does *only* because determinism is true: it may also be true that he acts as he does because he has reasons for doing so. What is really important for moral responsibility is not whether the individual could have done otherwise or not, it is whether they were appropriately motivated. As Frankfurt writes: "the principle of alternate possibilities should thus be replaced, in my opinion, by the following principle: a person is not morally responsible for what he has done if he did it only because he could not have done otherwise."[30]

Fischer and Ravizza build upon these ideas in an attempt to ascertain whether an agent can have morally relevant control over their actions even if they cannot not do otherwise. They identify two types of control: guidance control and regulative control. The former, guidance control, is the type of control a pilot has over an aeroplane with the autopilot switched on—the pilot can use the control system to direct the aeroplane along a given route, but they do not have the ability to override the system and, say, crash the plane. The latter, regulative control, involves both guidance control and something further, namely the ability to direct the aeroplane in whichever way the pilot chooses. A pilot has regulative control when they have the freedom to fly the plane in any direction and to any destination, either safely or dangerously. Guidance control does not require the ability to do otherwise, whereas regulative control does.

Fischer and Ravizza argue that guidance control alone can satisfy the freedom-relevant condition, making it a sufficient condition for moral responsibility. As guidance control does not require the ability to do otherwise, the problem raised by causal determinism is sidestepped.[31] Moral responsibility requires agency; for it to hold, the individual in question must have acted in such a way that the action is a result of *them* (viz. their choices and their actions that realise those choices). If an individual is coerced, manipulated, forced, or deceived into performing a certain action then Fischer and Ravizza say they do not have the appropriate guidance control over their action—some external force was guiding them. Similarly, in Frankfurt-style cases, the agent has guidance control insofar as they are guiding their behaviour toward a given destination for their own autonomous reasons.

How does guidance control satisfy the requirements for moral responsibility? Fischer and Ravizza urge us to look not at the *agents*, but at the *mechanisms* that lead to action. Though the agent could not do otherwise in Frankfurt-style cases, in cases in which only guidance control is present, and when the agent is in a causally deterministic system, the mechanism by which they decide to act is reasons-responsive in a morally relevant

manner. By reasons-responsiveness they mean that the mechanism of their choice is responsive to reasons that motivate them to act in morally relevant ways. In Frankfurt-style cases, for example, the individual who could not have acted otherwise nonetheless acts for entirely their own reasons and the external restrictions that would be imposed on their action were they to act otherwise are unbeknownst to them. If they had tried to act otherwise, triggering a Frankfurt-style device that would have forced them to carry out the action, then the Frankfurt-mechanism would not be reason-responsive (nor would it be a case of the individual having guidance control). So whilst the *agent* could not have done otherwise, the *mechanism* by which they chose to act and then acted could have been otherwise. What matters is whether it comes from within the agent or not.

On the issue of causal determinism, Fischer and Ravizza write:

> Why should a new scientific discovery of the sort envisaged—let us say a discovery about the precise form of the natural laws—require us to stop thinking of ourselves and certain others as persons, as subject to love, hatred, respect, friendship, indignation, and moral assessment? … Whether we are persons—whether we are morally responsible agents who can have deep personal relationships, friendships, and families—should not depend on whether causal determinism is true: our most basic views about ourselves—and our most fundamental kinds of relationships—should not be held hostage to the particular features of the best theories offered by current physics.[32]

In a deterministic world individuals can never act other than in the ways they do. Although PAP has shown by Frankfurt to be insufficient in conferring moral responsibility to agents, whether guidance control gives the agent sufficient sovereignty over their actions remains an open question, particularly in the context of MWI. If an individual could guide themselves toward branches in which they made certain decisions, then perhaps this would be a plausible route by which we could recover moral responsibility in MWI. Their guided, albeit determined, action still contains a sufficient amount of control to confer moral responsibility. But of course it does not work this way. First, decisions that lead to actions are not the kind of thing that cause the universal wavefunction to branch. Second, even if they did, it is not up to us which branch we end up in (and thus which decision we guide ourselves toward).

The autopilot example brings out this problem clearly: if an aeroplane crashes whilst on autopilot, the pilot is not considered fully responsible for the incident. Instead, the airline will examine the autopilot system to determine whether there is some internal fault that allowed the pilot to guide the aeroplane into an unsafe situation. Responsibility is assumed to lie with the process that determined the path taken, not the pilot who guided it along that path. Perhaps some readers will feel similarly about moral agents in a deterministic world—guidance control does not give the agent a morally relevant amount of autonomy over their actions. The problems with this solution echo familiar criticisms levelled at compatibilism. As Swinburne puts it, whilst the compatibilist claims "that there is the crucial difference between the case where an agent is prevented by external force from doing an action and the case where he or she is fully caused not to do it by some desire or brain state," the incompatiblist will always reply, "if the agent cannot help his or her intentions why should that difference matter?"[33] Perhaps an incompatibilist solution might be more convincing.

(c) Freedom *within* Branches

The challenge from causal determinism is that there is no possible world in which the laws of nature + the past can lead to anything other than the actual future. Thus, there is no possible world containing our laws of nature + our past in which *x* chooses anything other than *a*. In some ways this is true in MWI, insofar as every possible outcome of an event is actualised, but in other ways (namely, from within the perspectives of worlds or world-branches) it is false. The outcome of a quantum event is underdetermined by the theory—it is genuinely random and can only be predicted probabilistically. This is precisely what determinism denies. Rather than there being one possible future given the laws of nature + the past, there are many possible futures. In fact, there are many *actual* worlds, or branches of the wavefunction, in which the laws of nature + the past give a different future to the one we see in our world. Thus, whilst MWI is deterministic when viewed from a God's-eye perspective, from within our branch we observe (perspectival) indeterminism. Could this provide enough metaphysical wiggle room to give us freedom of a morally relevant kind?

In the previous section, I mentioned the possibility that freedom of the morally relevant kind could be accommodated within MWI if there was some sense in which agents could guide themselves into certain branches in which they make particular decisions. This is one potential incompatiblist solution. On this view, although all possibilities are actualised, which specific actualisation the agent occupies is in some sense up to them. Whilst this would, in my view, qualify as the kind of freedom that would confer moral responsibility to the agent in question, it is not the way things work within MWI. Our decisions to act are not the kind of things that branch the wave function (as far as we know), and even if they were, a moment's introspection reveals that we do not possess the power of guiding ourselves toward certain branches of the wavefunction over others.

A more general problem faces all such incompatiblist solutions: it may be the case that any indeterminism (perspectival or objective) only appears at the quantum level. It is by no means clear how indeterminism in the quantum realm could echo up to the level of human experience, *at least in such a way that it could be harnessed by human agency*. Jeffrey Koperski, for example, argues that according to the "amplification problem" quantum effects are unable to generate significant results in the macro world in which our day-to-day lives play out.[34] The amplification problem states, roughly, that what goes in in the quantum realm stays in the quantum realm. More specifically, there exist a multiplicity of, in the words of Nobel physicist Robert Laughlin, "quantum protectorates" that prevent quantum effects becoming apparent in the physical sphere still described by classical physics.[35] Koperski writes:

> In light of current physics, the amplification problem cannot be solved. Not only are amplification mechanisms hard to find, but the physics between scales puts obstacles in the way. In other words, nature is strongly predisposed to block the amplification of indeterministic quantum events.[36]

Even if indeterminism does rule the quantum realm, albeit only within the perspective of various branches, it remains unlikely to provide the kind of ontological openness required for morally relevant freedom. That being said, there are some instances in which quantum effects have macroscopic consequences. For example, quantum

fluctuations in the early (inflationary) universe which undoubtedly had significant effects on the universe as a whole. Nevertheless, it remains unclear as to how this could help. In order for this to be relevant to the issues of freedom and responsibility, these macroscopic effects would need to be harnessed by human agency, and it is far from clear how such a thing could ever possible. So, at the present time, it seems as though an incompatiblist solution is cut off at the knees.[37]

Can We Be Responsible in an Everettian World?

At the outset of this article I set out to ascertain whether moral responsibility can be accommodated within the Everett interpretation of QM. I stated why this concept is so integral to theism, referring to various Scriptural passages in which God is shown to judge the moral actions of human agents. Having shown why moral responsibility matters, I identified two necessary conditions for moral responsibility relating to identity and freedom. Each of these is an important feature of theological anthropology anyway, so examining how they play out in an Everettian context is valuable in its own right as well as for the coherence of moral responsibility. I argued that neither of those can be deemed to hold, at least straightforwardly, in Everett's Many Worlds Interpretation.

I have hinted at several places that if a solution is to be found it may only work perspectivally. In other words, we have subjective justification for upholding our standard practice of holding people morally responsible for their actions because, within our world-branch and for all practical purposes, these practices still seem justifiable. But if we want more than subjective, perspectival justification for moral responsibility (i.e. if we want the preservation of numerical identity across a moral agent's lifetime and we want that agent to be free in a libertarian sense), then MWI is highly problematic. Whilst perspectival justification might be enough for human-to-human interactions, it is not theologically robust. If God is wholly good and wholly just, then holding us morally responsible must be more than a perspectival and subjective affair. Whether in this branch of the wavefunction or another, more work needs to be done to show that identity and freedom can be reconciled with Everettian QM. That way, Christian theology will be ready if the theory turns out to be true.

Notes

1. R. Swinburne, *Mind, Brain, and Free Will* (Oxford: Oxford University Press, 2013), chapter 6.
2. There may be other conditions that need to be satisfied before moral responsibility holds, which fall outside the scope of the present discussion.
3. With the possible exception of the crucifixion, which is often interpreted as an event in which God died for the sins of human agents.
4. Swinburne, *Mind, Brain, and Free Will.*
5. Hugh Everett III, "'Relative State' Formulation of Quantum Mechanics," *Reviews of Modern Physics* 29 (1957), 454–462, 316.
6. Exactly what counts as a quantum event is given by Wallace and will be explored shortly.
7. Simon Saunders, "Decoherence, Relative States, and Evolutionary Adaptation," *Foundations of Physics* 23 (1993), 1553–1585, 1560–1561.
8. Paul Tappenden, "Saunders and Wallace on Everett and Lewis," *The British Journal for the Philosophy of Science* 59:3 (2008), 307–314, §2.

9. See Saunders 1993 and Wallace 2010 for a detailed explanation of how this works.
10. Such a view can be found in B. S. M. De Witt, "Quantum Mechanics and Reality," *Physics Today* 23:9 (1970), 30–35.
11. David Wallace, "Decoherence and Ontology," in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. Simon Saunders, Jonathan Barrett, Adrian Kent, and David Wallace (Oxford: Oxford University Press, 2010), 68.
12. Ibid., 102.
13. David Wallace, *The Emergent Multiverse [Electronic Resource]: Quantum Theory According to the Everett Interpretation*. Oxford Scholarship Online (Oxford: Oxford University Press, 2012), 1.
14. Tim Maudlin, *Critical Study David Wallace, The Emergent Multiverse: Quantum Theory According to the Everett Interpretation* (Oxford University Press, 2012).
15. That being said, I can be both causally and morally responsible for some action even if I do not myself commit the action. Consider a third version of events: I tell my younger cousin that I will give him £20 if he goes and pushes our grandmother over. If he obeys, and our grandmother falls to the ground, then both he and I are morally responsible for her resulting injuries even though it is only he who pushed her. I am appropriately situated in the chain of causes as to render myself causally (albeit indirectly) responsible and I freely chose to place myself there. In each case, the person who causes grandmother to fall (accidentally, directly, indirectly) is *me*.
16. Derek Parfit, "Personal Identity," *Philosophical Review* 80 (1971), 3–27, 10.
17. David K. Lewis, "Survival and Identity," in *The Identities of Persons*, ed. A. Rorty (Berkeley, CA: University of California Press, 1976).
18. Simon Saunders and David Wallace, "Branching and Uncertainty," *The British Journal for the Philosophy of Science* 59:3 (2008), 293–305.
19. Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), 256.
20. Sean Carroll, *Something Deeply Hidden: Quantum Worlds and the Emergence of Spacetime* (London: Oneworld Publications, 2021), 119–120.
21. Paul Tappenden, "Saunders and Wallace on Everett and Lewis," *The British Journal for the Philosophy of Science* 59:3 (2008), 307–314, 312.
22. Parfit, *Personal Identity*, 11.
23. Parfit, *Reasons and Persons*, 261.
24. Carroll, *Something Deeply Hidden*, 139–140.
25. Maudlin, *Critical Study David Wallace, The Emergent Multiverse*, 801.
26. P. Van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), 16.
27. Swinburne, *Mind, Brain, and Free Will*, chapter 8.
28. John Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge Studies in Philosophy and Law). (Cambridge: Cambridge University Press, 1998), 13.
29. Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66:23 (1969), 829–839.
30. Frankfurt, *Alternate Possibilities and Moral Responsibility*, 838.
31. Fischer and Ravizza, *Responsibility and Control*, 34.
32. Ibid., 16.
33. Swinburne, *Mind, Brain, and Free Will*, chapter 8.
34. Koperski's work focuses specifically on divine action, but it is also useful for our related question of human action.
35. Robert B. Laughlin and David Pines,"The Theory of Everything," *Proceedings of the National Academy of Sciences of the United States of America* 97:1 (2000), 28–31, 9.
36. J. Koperski, "Divine Action and the Quantum Amplification Problem," *Theology and Science* 13:4 (2015): 379–394, 381.
37. There are, of course, some Christians that reject the plausibility of free will altogether, endorsing an incompatiblist view that holds determinism to be true for theological reasons. These views may hold that our actions are determined by God (as is sometimes

thought to be indicated by Romans 9: 19–21). One such example is Calvin's idea of predestination. These views are somewhat orthogonal to the present paper's concerns as I am following Richard Swinburne and others in claiming that free will *is* a highly important feature of theological anthropology. For those who reject this, however, some of the problems I raise here may be able to be circumvented.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Emily Qureshi-Hurst is a Junior Research Fellow at Pembroke College, University of Oxford. She has written on many issues on the intersection between theology, physics, and philosophy, and her new book "Salvation in the Block Universe: Time, Tillich, and Transformation" is out 31st August 2024.

ORCID

Emily Qureshi-Hurst http://orcid.org/0000-0003-3108-4908