**Forever and Again:**

**Necessary Conditions for “Quantum Immortality” and its Practical Implications**

Alexey Turchin

Science for Life Extension Foundation,

Digital Immortality Now

alexeiturchin@gmail.com

**Abstract**

This article explores theoretical conditions necessary for “quantum immortality” (QI) as well as its possible practical implications. It is demonstrated that the QI is a particular case of “multiverse immortality” (MI) which is based on two main assumptions: the very large size of the Universe (not necessary because of quantum effects), and the copy-friendly theory of personal identity. It is shown that a popular objection about the lowering of the world-share (measure) of an observer in the case of QI doesn’t work, as the world-share decline could be compensated by the merging timelines for the simpler minds, and also some types of personal preferences are not dependent on such changes. Despite large uncertainty about MI’s validity, it still has appreciable practical consequences in some important outcomes like suicide and aging. The article demonstrates that MI could be used to significantly increase the expected subjective probability of success of risky life extension technologies, like cryonics, but makes euthanasia impractical, because of the risks of eternal suffering. Euthanasia should be replaced with cryothanasia, i.e. cryopreservation after voluntary death. Another possible application of MI is as a last chance to survive a global catastrophe. MI could be considered a plan D of reaching immortality, where plan A consists of the survival until the creation of the beneficial AI via fighting aging, plan B is cryonics, and plan C is digital immortality.

**Highlights:**

* Quantum immortality (QI) is a particular case of multiverse immortality (MI).
* The validity of MI depends on the size of the Universe and the nature of personal identity, but still uncertain.
* The counterargument of “world-share decline” depends on the type of preferences and merging timelines.
* MI greatly increases the subjective chances of the success of cryonics, makes euthanasia impossible, but favors cryothanasia.
* MI is our last line of defense against existential risks.

**Disclaimer**: Suicide will never be useful in terms of quantum immortality: it will probably result in a non-deadly serious injury, permanent brain damage, and infinite suffering. The idea of quantum immortality is not an argument for suicide, but is one of the strongest arguments against it, as suicide will never work to end suffering. If you are still interested in suicide, seek professional psychological help.

**Memetic hazard** for people who tend to react emotionally to thought experiments!

**Abbreviation**:

DA – Doomsday argument

QM – Quantum mechanics

QS – Quantum suicide thought experiment

QI – Quantum immortality

MI – Multiverse immortality: umbrella term for QI and big world immortality

MWI – Many-worlds interpretation of quantum mechanics

[1. Introduction 3](#_Toc533086215)

[2 The nature and formalism of the multiverse immortality 4](#_Toc533086216)

[2.1 History of the idea 4](#_Toc533086217)

[2.2 The Quantum suicide thought experiment and infinite survival timelines 5](#_Toc533086218)

[2.3 Formalism of multiverse immortality 5](#_Toc533086219)

[2.5 The logical indefinability of the observer’s death 6](#_Toc533086220)

[2.6 Examples of the multiverse immortality 6](#_Toc533086221)

[2.7 Overview of the types of multiverse immortality 7](#_Toc533086222)

[3. Necessary conditions for multiverse immortality: the huge size of the Universe and a “copy-friendly” theory of personal identity 8](#_Toc533086223)

[3.1 Condition 1: Repetition of human minds in the Universe 9](#_Toc533086224)

[3.1.1 Tegmark’s levels of multiverse 9](#_Toc533086225)

[3.1.2 The finite size of human minds 10](#_Toc533086226)

[3.1.3 Sufficient diversity of the Universe 10](#_Toc533086227)

[3.1.4 “Actuality” of the existence of the very remote copies 10](#_Toc533086228)

[3.2 Condition 2: Appropriate theory of personal identity 11](#_Toc533086229)

[3.2.1 The correct theory of personal identity is still unknown 11](#_Toc533086230)

[3.2.2 The paradox of “Teleportation on Mars” and its similarity to multiverse immortality 11](#_Toc533086231)

[3.2.3 “Copy-friendly” and “unique-soul” approaches to the problem of the observer identity 11](#_Toc533086232)

[3.2.4 The merge of observer-identity and memory-identity 12](#_Toc533086233)

[3.3 Personal identity theories’ connection with various types of multiverse immortality 12](#_Toc533086234)

[3.4 A rough estimation of the probability that at least some form of MI is true 13](#_Toc533086235)

[4. Possible negative effects of MI 13](#_Toc533086236)

[4.1 The idea of MI is an informational hazard 13](#_Toc533086237)

[4.2 Impossibility of death may be bad 14](#_Toc533086238)

[4.3 Long-term inescapable suffering is possible 14](#_Toc533086239)

[4.4 Bad infinities and bad circles 15](#_Toc533086240)

[4.5 Marginalization of measure, strange worlds and infinite torture 15](#_Toc533086241)

[4.6 Dangers of mistakenly ignoring MI and of MI-euphoria 15](#_Toc533086242)

[5. MI as a decision theory problem 15](#_Toc533086243)

[5.1. MI and expected utility 15](#_Toc533086244)

[5.2 Merging timelines compensate the decline of measure 17](#_Toc533086245)

[6. Possible practical applications of the multiverse immortality 18](#_Toc533086246)

[6.1 Surviving until life extension technologies 18](#_Toc533086247)

[6.2 Multiverse immortality significantly increases the chances of the success of cryonics 18](#_Toc533086248)

[6.3 The impossibility of euthanasia and the need for cryothanasia 19](#_Toc533086249)

[6.4 Multiverse immortality increases the chances of the success of digital immortality, and of the acausal trade between multiverse’s branches 19](#_Toc533086250)

[6.5 “Universal problem solver” as an example of bad implementation of MI 20](#_Toc533086251)

[6.6 Collective survival in a global catastrophe 20](#_Toc533086252)

[Table 4. The outcome matrix for multiverse immortality and significant life choices 21](#_Toc533086253)

[6.7 Multiverse immortality in the age of superintelligence: will I become AI? 21](#_Toc533086254)

[Conclusion 21](#_Toc533086255)

[Appendix 1. Multiverse immortality and probabilities 26](#_Toc533086256)

[MI immortality formalism 26](#_Toc533086257)

[MI and probabilities 26](#_Toc533086258)

[Difference in the probabilities of outcomes in the multiverse and quantum immortality 26](#_Toc533086259)

[Marginalization: Strange outcomes will dominate after several QS experiments 27](#_Toc533086260)

[Many futures and different probability distribution at the moment of death 27](#_Toc533086261)

[“Many futures” imply that several types of immortality will happen 27](#_Toc533086262)

[“Believer against one’s will” thought experiment 28](#_Toc533086263)

[Multiverse-immortality for computer algorithms 28](#_Toc533086264)

[Appendix 2. Multiverse immortality and other similar probabilistic arguments 29](#_Toc533086265)

[Multiverse-immortality and simulation argument 29](#_Toc533086266)

[Multiverse-immortality and Doomsday argument 29](#_Toc533086267)

[Modal realism and back causation 29](#_Toc533086268)

[Multiverse-immortality reset and infinite timelines in the space of possible minds 29](#_Toc533086269)

[The same “me” in different worlds and superposition of interpretations 30](#_Toc533086270)

[Multiverse-immortality and qualia 30](#_Toc533086271)

[Appendix 3. Other objections to QI and wrong interpretations 30](#_Toc533086272)

[Aging and slow death 30](#_Toc533086273)

[Not-exactly-me copies and jumping over bad event via forgetting 31](#_Toc533086274)

[Resetting of timeclock, respawning and reference class problem 31](#_Toc533086275)

[General anti-immortality principle based on observed age 32](#_Toc533086276)

[QI is only artefact of thinking, not the property of the world 32](#_Toc533086277)

[Adding up to normality 32](#_Toc533086278)

[Appendix 4. Next step of QI theory: Dust theory and Boltzmann brains 32](#_Toc533086279)

[Multiverse-immortality and Boltzmann brains 32](#_Toc533086280)

[Multiverse-immortality and Dust-theory universe 34](#_Toc533086281)

[Dust theory and Flux-universe 34](#_Toc533086282)

[Miracles and glitches in the observer-moments chains 35](#_Toc533086283)

[Blog post and comments about the decline of measure and its compensation by merging timelines 35](#_Toc533086284)

[Comment 1. Three solutions of the conundrum: utility calculation with QI, measure manipulation magic or open individualism 37](#_Toc533086285)

[Comment 2: any definition of identity creates possibility of "pseudo-copies”. 37](#_Toc533086286)

# 1. Introduction

The main idea behind quantum immortality is that constant branching of the multiverse – according to the many-worlds interpretation (MWI) of quantum mechanics (QM) – ensures the existence of timelines in which an observer will survive any dangerous situation. The idea was suggested in a form of a quantum suicide (QS) thought experiment by Tegmark (Tegmark 1998) and others.

Most scientists who have written about the idea of quantum immortality have felt an obligation to disprove it (Randall 2004; Mallah 2009; Almond 2011c; Aranyosi 2012), in the same way as most writers try to disprove the Doomsday argument (Bostrom, 1999), but do not analyze the implications of the small probability that QI is true. As a result, the possible practical consequences of the QI are underexplored. But some find QI or its derivates plausible (Loew 2017).

QI is closely related to the effective altruism circle of problems. If quantum immortality is real, it means there is a possibility of infinite future suffering for any sentient being (Aranyosi 2012), as the person will not be able to die, but will age. So QI is an *s-risk* (Daniel 2017). However, in this article it will be demonstrated that there is a way to prevent this s-risk for people who are currently alive by increasing share of their future positive infinite timelines via signing up to crionics1.

An overview of the history of the idea, its formalism, and its relation to the idea of death is presented in Section 2. Section 3 is devoted to the difference between “quantum immortality” and “multiverse immortality” and the conditions necessary for them to be true are analyzed. Section 4 goes into possible bad consequences of “natural” QI. In Section 5, QI is explored from the point of view of decision theory and the “measure decline” objection is analyzed. Section 6 provides an overview of the practical applications of QI, including overcoming the negative effects of “natural QI” via cryothanasia (i.e. cryopreservation after voluntary death).

# 2 The nature and formalism of the multiverse immortality

##

## 2.1 History of the idea

The first stage of the conception that infinite size of the universe implies some form of immortality was the idea of so-called “eternal return,” that is, the repeated appearance of the same observers. The earliest versions of the eternal return theory appeared in ancient philosophy (Eliade 1949) and still attract interest today (Bergström 2012). Empedocles of Akragas reportedly believed in the cyclic return of everything based on the recombination of four elements, ensuring personal immortality (Brown 1984).

In more recent thought, the idea that an infinite universe implies some form of immortality came to Romantic poet Heine in the first half of the 19th century, who wrote:

Time is infinite, but the things in time, the concrete bodies, are finite. They may indeed disperse into the smallest particles; but these particles, the atoms, have their determinate numbers, and the numbers of the configurations which, all of themselves, are formed out of them is also determinate. Now, however long a time may pass, according to the eternal laws governing the combinations of this eternal play of repetition, all configurations which have previously existed on this earth must yet meet, attract, repulse, kiss, and corrupt each other again (Kaufmann 2013).

The idea of “cyclic immortality” was formalized by French socialist Blanqui in the book “Immortality Through the Stars” (Blanqui 1872). Nietzsche came to the same idea in 1881 when he suggested his famous theory of “*eternal return*” (Nietzsche 1883).

However, Nietzsche didn’t understand (or ignored) the obvious consequences of the eternal return: *that there will be always a world, similar to our world until the moment of the observer’s death, but different in the moment of the death in such a way that the observer will not die*. This basically means immortality from the subjective view of any observer. This is the idea of *multiverse immortality* (of which *quantum* *immortality* is an important specific case) in a nutshell, and it is a natural development of the idea of the “eternal return.”

In the middle of 20th century H.Everett was probably the first to come to the idea that a constantly branching universe implies immortality from the subjective point of view, based on his discovery of the many-worlds interpretation of quantum mechanics, but never stated this publicly (Shikhovtsev 2003). Only at the end of the 20s century the idea – that the existence of the multiverse implies some form of immortality – become known as “quantum immortality” after the “quantum suicide” thought experiment proposed by Squires, Moravec, and Marchal and popularized by Tegmark (Marchal 1991; Moravec 1988; Squires 1994; Tegmark 1998).

The preservation of information implicit in quantum mechanics and known as the “no-hide theorem” (Braunstein and Pati 2007) is not connected with the idea of “quantum immortality” as it was originally presented by Tegmark and others. In the original QI theory, human survival happens only in one of the branches of quantum multiverse, and preservation of the information in other branches is not necessary. However, such quantum preservation could be used by another approach to the resurrection of the dead, i.e. “quantum archeology,” which is out of the scope of this paper, but has been previously addressed by the author (Turchin and Chernyakov 2018) and by Jones (Jones 2017).

Previous work has sought to address the validity and practical features of QI. There are only a few vocal proponents of the QI. Higgo (Higgo 1998) tried to present the idea of quantum immortality as a valid theory of immortality, but his analysis is not deep. Wilson explores consequences of quantum immortality in his short story “Divided by Infinity,” where the protagonist survives multiple suicide attempts and even a global catastrophe and is resurrected by aliens (Wilson 1998).

Randall criticizes QI from the point of view that its most likely outcomes are neither “technological resurrection” nor “eternal decrepitude”, but chaotic random observers similar to Boltzmann brains. He does, though, note the usefulness of QI for cryonics (Randall 2004). Bevers explores how QI relates to the Born rule and apparently breaks it (Bevers 2011). Cirkovic correctly mentions that the theory of personal identity is critical for QI (Cirkovic 2004), and proceeds to prove that the subjective credence should be equal to the objective probabilities in the case of QI. Almond extensively criticizes the idea that the quantum suicide thought experiment could be used as a proof of the many-worlds interpretation of quantum mechanics (Almond, 2011). He also suggests that quantum suicide could be a “universal problem solver,” and that the concept of civilizational-level quantum suicide could be used to explain the Fermi paradox (Almond, 2008).

“Dust theory” (Egan 2009) seems to be the next step in the developing the idea of QI, similar in significance to the step from the “eternal return” in the 19th century to the “quantum suicide” thought experiment. In dust theory, something similar to quantum immortality happens in every observer-moment; such moments could exist completely separately from each other, as random patterns in dust, or as Boltzmann brains, but could look subjectively connected “from inside” based on their similarity. An important conclusion of such a theory is that there is no need for an external reality, and that each observer-moment could have many pasts and many futures. The mathematical theory of such a world was recently suggested by (Mueller 2017). The validity of such theories is out of the scope of this paper; in this work, the actual existence of an objective world governed by observer-independent laws of physics is assumed.

Loew suggested a similar to the dust theory idea of “Boltzmannian immortality”, which is basically a reincarnation of the eternal return which takes into account Boltzmann brains randomly appearing from vacuum if an empty universe lasts for an infinitely long period (Loew 2017).

## 2.2 The Quantum suicide thought experiment and infinite survival timelines

The main idea of the quantum suicide thought experiment is that a quantum event like radioactive decay of an atom will trigger a powerful bomb near me, an observer, with a probability2 of 0.5. As a result, two future timelines are possible, depending on whether or not the bomb explodes. If Everett’s many-worlds interpretation of QM is true, both timelines will be real. In one, I will exist, and in the other, I will be instantly destroyed. Thus, I will observe only the timeline in which the bomb didn’t explode. This experiment seems similar to the thought experiment of Schrödinger’s cat, but the tantalizing idea that the observer could be in a superposition of dead and alive states is missing here.

Thus, QS means that an observer will always observe a winning result in “Russian roulette,” which may seem as if they have surprising skills of survival. The main question is the equality of the inevitable observation of not dying in such an experiment to subjective immortality: in the many-worlds interpretation of QM there will always be a timeline in which the bomb didn’t explode after any number of the attempts – but is this condition enough to conclude that the observer is immortal?

QS thought experiment has also been analyzed as a possible proof for MWI theory: how many failed attempts of survival is enough for an observer to conclude that the MWI is true? Almond and others have invested significant time in proving that no number of instances of survival in QS is a proof of MWI. However, as Bostrom notes, if observations are discarded based on the assumption that in an infinite universe there exists an observer for any possible observation, science is not possible (Bostrom 2002), as observation loses its predictive power.

While QS requires exotic devices such as a bomb and a quantum random generator, there is no reason to think that any ordinary cause of death is different: dying will be more complex, but eventually there will be two main types of the future timelines: one in which I died and one in which I survived. However, slower dying implies some more complex dynamic, like an increased probability of survival in an injured state including suffering, which will be addressed later.

In a nutshell, QS means some form of “quantum immortality”, i.e. for any observer, there is (at least one) infinite future timeline in which that observer will continue to exist. However, such result could happen not only in a quantum world, but even in a completely classical world, if some conditions about the world’s size and the observer’s internal structure are applied, as will be shown in the next section.

## 2.3 Formalism of multiverse immortality

One useful instrument when speaking about observers is the idea of the “observer-moments,” that is, the shortest periods of subjective experience, probably corresponding to one step of internal human clock, which has a duration of a few tens of milliseconds. For simplicity’s sake, it is assumed here that any observer-moment is characterized by the name *I* and a time-stamp *t*, and the observer-moment is thus notated as *O(I, t)*. Then, immortality could be defined as the following statement: *for any observer-moment O(I,t) necessary exists (at least one) observer-moment O(I, t+1). There is no any last observer-moment O(I, tlast), so there is no death.*

Note that there is a hidden claim that the existence of *O(I,t+1)* is *sufficient* for immortality. This exact claim typically raises the most objections: while some may agree that somewhere in the multiverse a copy of me which survived my death exists, they claim that such existence is not immortality *qua* immortality because of a different identity, lower “measure”, or some decision-theoretic considerations. Some of these objections will be discussed in the sections that follow*.*

This definition of immortality is general and – hypothetically – could be satisfied in many ways (e.g. God created the world this way). What is claimed in the multiverse immortality theory is that the infinitely large size of the universe is sufficient to satisfy the immortality definition (given a few other assumptions).

## 2.5 The logical indefinability of the observer’s death

There is a well-known conjecture: “All humans are mortal, I am a human, so I will die.” Though it is well-known, it has several flaws. One of them is the “fallacy of four terms” (Copi and Cohen 1990), as the word “death” has different meanings in the premises and in the conclusion. In the premises, “death” means “death-for-others”, i.e. *observation of death by outside observers*, or cessation of the living activity of the body. In the conclusion of the syllogism, the term “death-for-me” is used, which means the *death of the observer from the observer’s point of view*, typically represented by the idea of “nothingness after death” or cessation of the process of observation. However, by definition, no one can experience and report anything about nothingness after death, so it is not possible to prove that it will actually happen.

Multiverse immortality means that death – from the observer’s point of view – is impossible. In other words, for any state of any observer with time-stamp *t0*, there is a possible next state *t0+1*, and, as everything possible exists, such an observer exists. Thus, a last moment of experience is impossible, and death – as the cessation of the process of experiencing – is impossible. Almond attacks this idea and notes that if MWI is not true, the death of observer is a consistent explanation or reality, but if it is assumed that the MWI is true, death in some branches immediately becomes an inconsistent explanation (Almond 2011d). According to his work, this jump between varying descriptions of death is not logical and undermines the idea of QI. In other words, if it is agreed that death is theoretically possible, because of a small world, it should also be agreed that death is possible in the branching world, as the death of some branches. This is not a question of fact, but one of interpretation: should the existence of person’s copy be regarded as immortality or not?

The answer to the problem raised by Almond is connected with the nature of personal identity, which will be analyzed in the next section. If identity is connected with the measure of existence, then there is some probability that lowering this measure is equal to death. If identity depends only on information, then a change in the number of copies is irrelevant. The nature of death depends on the nature of personal identity, and death could be described as “the end of identity”. Another counterargument is that if I am not currently on Mars, that doesn’t mean that I am dead on Mars. In any given moment, there are a lot of branches which are not happening with me, but it doesn’t mean that I die every moment – and if the non-existence of a branch is considered a death, I am dying every second in millions of ways.

## 2.6 Examples of the multiverse immortality

There is nothing strange in the idea of the multiverse immortality, when simpler objects and beings are considered:

**Numbers are immortal.** All numbers appear over and over again in multiple worlds, and it could be said that any given number is “immortal”. The same is true for number series: the number 27 will appear again and again, but we are not surprised by this, as the number is small compared to the size of our world, where we can find, for example, 27 ships and 27 sheep. However, the repetition of longer numbers in the world becomes more and more surprising – only because the world seems to be too small for such repetition. In the infinitely large and random world, any number will repeat.

**Simple molecules are immortal. A** molecule of water will appear again and again, because it is very simple, and consists of abundant components. But more complex compounds are rare. Some short organic molecules are likely to repeat rather often, while a molecule with around 100 randomly connected atoms will be so rare that it might appear once in our Hubble volume. In other words, the simpler the object, and the more stochastic is its environment, the more likely it will have something like multiverse immortality.

**Biological viruses**. Even some living beings could enjoy this type of immortality in an observable way. Biological viruses of one species are all functionally equivalent to each other (barring mutation) and the death of one copy is not the death of the virus.

## 2.7 Overview of the types of multiverse immortality

“Multiverse immortality” will be used as a general term, which unites three subtypes. However, these subtypes may have no observational differences: eternal return will look like quantum immortality if someone’s life history completely repeats in the future, until the moment of death, but, in this moment, becomes different – and the only difference is that the person will not die. All three types are illustrated on Figures 1-3.

A useful metaphor to understand multiverse immortality is Borges’ thought experiment of the library of Babel (Borges 2000). In this library, an infinite number of copies of all possible books exist. If one copy of a book disappears, the book will still exist within the library, may be different in just a few typos.

It should be noted that a different meaning of the term “quantum immortality” is sometimes discussed. In it, QI is some form of near-death experience (NDE) after which a person jumps back in his life and relives the moment of a dangerous situation without dying (typically, a car accident)3. While some people experience this form of NDE as a hallucination, it is not the original idea of QI, where there are no jumps back in time and which cannot be reported to outside observers. QI does not require a “quantum mind” or other quantum woo: it is based only on the many-worlds interpretation of quantum mechanics. It does not require any quantum effects except the existence of observer copies in other branches of a multiverse.

*Table 1. Types of the multiverse immortality*

|  |  |  |
| --- | --- | --- |
|  | **Subtype** | **Underlying properties of the multiverse**  |
| **Multiverse immortality** | Eternal return with alternative ending | Infinite in time |
| Big world immortality | Infinite in space or parallel worlds |
| Quantum immortality | Many-worlds interpretation of QM |



*Figure 1. Eternal return with an alternative ending at the moment of death.*



*Figure 2. Big world immortality. There is another copy of an observer in a causally disconnected region of the universe, and this observer-copy does not die at the moment of the observer’s death.*

*Figure 3. Quantum immortality. There is always a branch of the multiverse in which the observer does not die. Smaller circles represent “measure decline;” the green line indicates the timeline of survival.*

# 3. Necessary conditions for multiverse immortality: the huge size of the Universe and a “copy-friendly” theory of personal identity

Two conditions are essential for multiverse immortality: that an infinite number of copies of an observer with alternative life histories exists, and that the replacement of an observer with its exact copy is a sufficient condition for immortality.

The first condition is factual: it is a claim that the universe is built in such a way that for any situation in which an observer has died, there exists another copy of the same observer which lived through that moment. Note that this condition is realized by different physical mechanisms of quantum immortality and of big world immortality. In the first case, it is realized via branching of the world wave function, and in the second case, the size of the universe is so large that such copies will appear in it because of combinatoric effects. While the mechanisms are different, the result is the same: infinitely many copies ensure the existence of endless chains of observer-moment4.

The second condition is not a factual claim, but a decision-theoretic one; it states that an individual should act as if she is immortal after she learns the first factual condition (and assuming some type of subject-centered preferences). Both conditions will be explored in detail in this section.

## 3.1 Condition 1: Repetition of human minds in the Universe

### 3.1.1 Tegmark’s levels of multiverse

Max Tegmark identified four levels of the multiverse sufficient to create a world large enough to include all possible observers’ minds, at least those the size of a human mind (Tegmark, 2017):

*“Level 1. Cosmological Inflation:* Our bubble universe is very large. Supported by observations of anisotropy. One estimate is 10^{10^{10^{122}}} megaparsecs (Page 2007).

*Level 2. Chaotic inflation*: quantum foam during the Big Bang creates many bubbles with differing laws of physics. Supported by ideas that fine-tuning of physical laws is explained by the observer selection effect, also known as anthropic principle.

*Level 3. Many worlds interpretation of quantum mechanics*. The most simple of the QM explanations.

*Level 4. Mathematical universe*: all possible mathematical structures exist. The most simple explanation of the fact that everything exists at all.”

The first level is tentatively supported by data from contemporary cosmology, first of all, cosmological inflation, as explained by (Tegmark, 2017). So, big world immortality in a very large universe receives even stronger scientific support than QI, which is based on the (currently) untestable MWI. However, all this evidence is not direct and is merely an extrapolation of our best theories of cosmology. Some other cosmological theories also allow multiverse immortality, e.g. a cyclic universe, a finite universe that exists for an infinitely long time via a series of Big Bangs and Big Crunches (Steinhardt and Turok 2002), and a fecund universe, which replicates via black holes (Smolin 1992).

The anthropic principle in a fine-tuned universe is also regarded as evidence for a very large universe with a variety of possible physical constants (Bradley 2009), as numerous “attempts” are required to create the right combination. By some estimations, as many as 10500 instances may be required just to arrive at the optimal combination of physical constants5. There are many cosmological theories, but it is not easy to create a viable cosmological theory which results in just one small universe that is fine-tuned for life; in some sense, it is computationally simpler to create a process which creates all possible universes.

The more general argument for the infinite universe is philosophical. There is a process which makes the universe appear from “nothing,” as could be concluded based on the fact of the universe’s existence. For a non-infinite universe, this process must stop at some point, preventing new universes from appearing. But stopping this process would require a force acting on nothingness, i.e. beyond the universe. However, the universe is, by definition, the only thing which exists; there can be nothing beyond it. Thus, the process that created the universe cannot be exhausted. Approximately the same logic was presented in Kant’s antinomy about the finite or infinite size of the universe (Kant 1781), where both alternatives were paradoxical. There is a possible solution, though: that the universe is neither finite nor infinite, but is instead constantly growing.

One attractive property of levels 1 and 2 is that they are only currently causally disconnected, but arose from the same causal process of the same Big Bang, and thus could become connected again at some time infinitely far in the future. So, copies from levels 1 and 2 surely have the same ontological status of actual existence (actuality, see Section 3.1.4) as any other object in the observable universe, including even your next-door neighbor.

Some scientists do not subscribe to theories of a multiverse (e.g. Hossenfelder 2018). The main argument of these critics is that as other universes are non-observable, it is difficult to observe them, and thus they are beyond real science – but this argument is obviously circular. It is, of course, difficult to prove that the Universe is infinite, but to prove that it is finite is *impossible*, so MI cannot be disproved. Interestingly, for multiverse immortality, it is not actually necessary that “everything possible exists” or even that the universe is of infinite size; it is only necessary that the universe is very large.

Another possible objection to the big world is the idea that only “me-now” actually exists – similar to the “moving spotlight theory” of now (Deasy 2015), “actual now” (Bitbol 1994) and “ontological privilege” of now (Frischhut and Skiles 2013) – and everything unobservable doesn’t exist. But even this will not kill the theory of multiverse immortality, as in that case, consciousness becomes an ontological necessity which helps to create now-moment and thus can’t be turned off.

### 3.1.2 The finite size of human minds

Not only does the universe need to be huge, but human minds need to be finite in size for multiverse immortality to exist. As the human brain consists of a finite number of atoms (around 5x1026), it seems to be finite. Tegmark shows that quantum effects do not prevent the number of states from being finite, as any quantum system in the finite space could only be in the finite number of internal states (Bocchieri and Loinger 1957; Tegmark 2017). Moreover, the data comprising personal conscious memory was estimated to be around 2.5 GB (Carrigan Jr 2006), meaning that only 1010^10 different human minds are possible. Since most human minds will have immeasurably small differences, the actual number is even smaller.

If one looks at the observer-moment’s size, that is, the current size of human experience, it is dominated by the visual field, which may be roughly estimated by the size of the signal which the eye sends to the brain. It uses one million nerves in the fibers of the optic nerves, each probably presenting one pixel. By that measure, the actual size of an observer-moment can be estimated as around 1 MB and given the presence of much non-significant peripheral noise, it is probably even smaller. Thus, there are 101000000 separate visual observer-moments, most of which are indistinguishable noise.

Surely, the most important part of an observer-moment is not what a person sees, but what she feels and thinks about it, but these parts are probably less complex and thus require less storage than the visual field. For example, assuming that the size of the verbal observer-moment is one sentence, and such a sentence can’t have more than 20 words, most of which are simple and frequent (and thus selected from just 1000 of the most common words) it could be concluded that there are 1060 possible sentences, less than the number of particles in the visual universe. By these measures, the size of the human mind is, indeed, finite.

### 3.1.3 Sufficient diversity of the Universe

An infinite universe and a finite size for human minds are not enough to prove multiverse immortality. The universe must be diverse enough for any possible mind’s states to appear because more than just repetition of events is needed: an escape from even the worst possible situations is required. Imagine that a spacecraft is falling into the Sun. At first glance, it seems that there is no chance for the pilot to survive. However, in a sufficiently diverse universe, an alien spacecraft could rescue her, or a rogue black hole moving at near-light speed could remove the Sun – or, most likely, the scenario is just a dream.

### 3.1.4 “Actuality” of the existence of the very remote copies

One important condition is that the whole big universe must *actually* exist (Menzel 2017). In short, “actuality” is the idea that some remote thing exists in the same way that I exist now. In a sense, “actuality” is actual “existence”, which implies that there is no “second-rate” form of existence, i.e. existence cannot be discounted. The difference becomes clear if one looks at the modal status of events in the past. Depending on the theory of time, one would say that past events are either actual (in a timeless universe), or not-actual, i.e. they did exist but do not exist now, so they are more than just possible events, but less than currently extant events.

There is an open question in philosophical ontology regarding whether non-causally connected regions can be claimed to actually exist. However, in our case, in some theories, they are connected, as in the cases of the MW-interpretation of QM and of the eternal universe; or they were connected; or they will be connected. For the validity of multiverse immortality, an observer’s remote copies must actually exist, or they should at least become actual in the future: if they are not actual, they do not exist. There are 3 solutions to the actuality problem:

1. Modal realism, that is, everything possible exists.
2. Actuality is somehow connected with the “now” moment and is “irradiated” by it.
3. “Relative actuality”, wherein actuality appears only if two objects are causally connected. In this case, unobservable regions of the universe are not actual.

In the case of quantum immortality, actuality is provided by the fact that the next observer-moment exists in the future and is causally connected with me-now. But if one starts to think about copies in other Big Bangs, it is easy to ask: if they are so remote, and so disconnected from our universe, how could they enable a person’s immortality? This question is an intuitive expression of doubts regarding the actuality of remote copies. Intuitively, it seems that the more potential there is for the causal connection between copies, the better it is for multiverse immortality, but it is not clear how well-grounded this intuition may be. There is even a point of view that if all moments of time actually exist, there should be no fear of death, as each moment will continue to exist (Deng 2015).

## 3.2 Condition 2: Appropriate theory of personal identity

In order for multiverse immortality to work, a person should treat the existence of a copy of herself as a sufficient condition of her own immortality. This problem is known as the “identity problem”, and its main question may be summed up as “Is information identity of minds is enough for personal identity?”

### 3.2.1 The correct theory of personal identity is still unknown

The theory of personal identity is notoriously difficult, despite seeming simple, and attempts to describe it will quickly produce controversies and spill off-topic. The basis of human identity is not known; for one thing, the nature of consciousness and qualia are not known (Chalmers 1996), and thus the nature of the human mind and of personal identity cannot be determined.

### 3.2.2 The paradox of “Teleportation on Mars” and its similarity to multiverse immortality

In the Mars teleportation paradox (Parfit 1984), a person is scanned, the copy is sent to Mars and recreated there, and the original is instantly destroyed on Earth. This paradox is similar to multiverse immortality, with just one difference: no data are actually sent to a remote celestial body, as the copy is already there. If a person agrees to Mars teleportation, she also should agree that existence of its remote copies implies some form of immortality. However, there is a caveat in the case of MI, as the number of surviving copies is constantly declining (as discussed in Section 5.1).

This paradox is not a factual problem – all possible facts are known and there is nothing which could be measured – but a decision-theoretical problem: a person must make a decision regarding whether or not to use the teleportation system. This decision depends on the goal system of the person and the way in which he treats his personal identity. If the person must go to Mars, to, say, save her children, she should use the transportation system. However, if it is just a recreational trip, the risk of losing something important about one’s own existence are too high to make it worthwhile.

### 3.2.3 “Copy-friendly” and “unique-soul” approaches to the problem of the observer identity

There are three main views on the nature of *personal identity*:

1. Copy-friendly view: Postulates that informational identity is enough for personal identity, and any exact copy of me is me.
2. “Privileged original” point of view: Postulates the existence of some unique substrate of identity that cannot be measured from outside but which clearly distinguish the original from the copies. The typical candidate for this substrate is either a “soul,” or a causal continuity of the human consciousness between subsequent observer-moments.
3. Open individualism: Denies the existence of the “identity” at all and postulates that all sentient beings are the same at their core.

Obviously, if some form of “soul” exists, quantum immortality will not work, as death is not real non-existence. Similarly, open individualism kills the idea of a separate, unique human, so the existence of any other mind is enough for some form of immortality to exist. Thus, only two main theories about personal identity are relevant: it is either information-only or based on causal continuity.

Causal continuity as the nature of human identity is a rather popular view, but it faces some difficult problems (Wiley and Koene 2015). For example, if a person experiences narcosis and his brain activity almost stops, should that be regarded as the death of personality? And what about the abrupt end of a dream? Causal continuity also doesn’t mean exclusivity of the next state of consciousness, as two new states could follow smoothly from the “parent” state. If quantum immortality is true, this is exactly what happens within it: minds smoothly branch into two (or more) minds with every new moment in time.

### 3.2.4 The merge of observer-identity and memory-identity

Identity could also be presented as two types of identity: identity as a chain of memories (which is a copy-friendly view, as any exact copy will have the same chain of memories), and identity as a chain of connected observer-moments (which is more dependent on the nature of consciousness). Moreover, both types of identity can be separated, as demonstrated in a thought experiment by Parfit, where the memories of one person are transplanted to another (Parfit 1984). These two types of identity could result in two types of QI: one as an amnestic observer, and one involving my memories in some other observer. The interaction between these two types of immortality is complex, but it may be assumed that they will eventually merge: the amnestic observer will regain her memories in some timeline, and thus it eventually will be just one type of immortality.

We, humans, experience such a merge of observer-identity and memory-identity every morning when we become awake after bizarre dreams and are remembering who we are. The same logic is applicable to the QI situation of very long and debilitating illness: there will be timelines where I will regain full memory of who I am, and, in some sense, I will jump over periods of minimal or bizarre consciousness.

Some suggested counterargument to QI of “impossibility of sleep”: QI-style logic implies that it is impossible to fail asleep, as in the moment of becoming asleep there will be timelines where I am still awake. However, for most humans, night dreaming starts immediately at the moment of becoming asleep, so the observations continue, but just don’t form memories. But in case of deep narcosis, the argument may be still valid with terrifying perspective of anesthesia awareness; but it also possible if the observer-states will coincide at the beginning the end of the operation, the observer will “jump” over it.

## 3.3 Personal identity theories’ connection with various types of multiverse immortality

Various combinations of the *personal identity theories* and the *big world theories* produce various answers about the possibility of MI. For example, if causal connection of minds is required for identity, AND the many-worlds interpretation of QM is true, then multiverse-immortality will work. In addition, if only informational identity is required AND only a cosmological inflation-based big world is true, multiverse-immortality will also work.

The relationship between the big world theories and identity theories is presented in Table 2. This table shows how multiverse immortality depends on these two variables. In it, “YES” means that multiverse immortality will work, while “NO” means that it will not work. Both variables are currently unknown to. Simply speaking, multiverse immortality will not work if the actual world is small or if personal identity is very fragile.

*Table 2. The relation between immortality, multiverse theories and personal identity theories*

|  |  |  |  |
| --- | --- | --- | --- |
| **Big world theory** *⇒***Personal identity theory**⇓ | MWI interpretation of quantum mechanics | Big world (cosmological inflation) | Only the visible universe actually exists |
| “I am only information”: copy-friendly view of personal identity | Yes | Yes | No |
| Causal continuity of consciousness is required for identity | Yes | No | No |
| Open individualism | Yes | Yes | Yes |

## 3.4 A rough estimation of the probability that at least some form of MI is true

Looking at Table 2, it follows that in only two-thirds of the cases do the combination of personal identity theory and big world theory point to the validity of multiverse immortality. Certainly, the numerical estimates are arbitrary, but they indicate that the uncertainty is very large. It is not easy to put an actual number to subjective credence of the validity of MI. There are serious arguments against any form of MI based on the “measure decline argument”, but a counterargument will be presented in Section 5.

This all means that a rational subject cannot be sure either of the validity of MI, nor of its falsity. Assume it should be assigned some median credence *P* in the theory, e.g. *P* = 0.5. This is not a zero result, because the consequences of MI being true or false are asymmetric. For example, a 0.5 chance of winning 1 million USD is not a zero result. As will be shown later, if MI is true, it implies a very bad outcome by default, which could turn into a very good one via relatively simple interventions. MI also causes some actions, like suicide, to have extremely negative utility.

# 4. Possible negative effects of MI

A naïve view of multiverse immortality is that it is some surprisingly good thing. However, there is no free lunch for immortality, and it will be shown in this section that natural MI could have negative outcomes for humanity. However, as will be demonstrated in Section 6, MI may be optimized to increase the probability of positive outcomes.

## 4.1 The idea of MI is an informational hazard

It seems that some people are motivated to object to MI based on ethical positions. The logic is that MI may lead an unstable person to suicide6, so it should be claimed that quantum suicide is false. However, the genie is out of the bottle; thus, it would be appropriate to explain that quantum immortality has nothing to do with “suicide” – and that this was an inappropriate name, one that might lead an unstable person to think that suicide will somehow make him immortal. If MI is true, any attempt at suicide will result in failure as well as injury, and only ever increase suffering; this should be explained to compensate for the potential damage of the idea of QS.

Similarly, the idea of the existence of the soul may result in a much stronger temptation toward suicide, as it at least promises another, better world, but I have never heard that the idea was hidden for fear of resulting in suicide attempts. Religions try to prevent suicide (and suicide is logical by their premises) by adding additional rules against it. MI itself does not promote suicide, and personal instability may be the main source of suicidal ideation.

If one really wants to test MI, there is no need to perform QS, but it is better to wait until the age of 110. Moreover, one may need not to undertake any experiment at all, as the mere fact that the life on Earth has survived in a dangerous universe – one full of gamma-ray bursts and large asteroids – for billions of years may be a case of multiverse immortality.

## 4.2 Impossibility of death may be bad

Multiverse immortality means that death is impossible. It does not mean that any given person will definitely have a linearly infinite lifespan, as timer resets and memory loss are possible. But it means that euthanasia is impossible, and one cannot choose death, no matter how much one may desire it. There is no escape from suffering and no way to go to “another” world. If MI is true, euthanasia seems much more dangerous to a person who already suffers, because it will perpetually fail and these failures will only increase her suffering. The risk of being buried alive also becomes more probable because of the multiverse immortality, as it is one of the forms of the strange survival, which objective probability is very small, but which could present a significant share of all the ways of unlikely survival.

## 4.3 Long-term inescapable suffering is possible

If death is impossible, someone could be locked into a very bad situation where she can’t die, but also can’t become healthy again. It is unlikely that such an improbable state of mind will exist for too long a period, like millennia, as when the probability of survival becomes very small, strange survival scenarios will dominate (called “low measure marginalization” by (Almond 2010). One such scenario might be aliens arriving with a cure for the illness, but more likely, the suffering person will find herself in a simulation or resurrected by superintelligence in our world, perhaps following the use of cryonics.

Aranyosi summarized the problem: “David Lewis’s point that there is a terrifying corollary to the argument, namely, that we should expect to live forever in a crippled, more and more damaged state, that barely sustains life. This is the prospect of eternal quantum torment” (Aranyosi 2012; Lewis 2004). The idea of outcomes infinitely worse than death for the whole of humanity was explored by Daniel (2017), who called them “s-risks”. If MI is true and there is no high-tech escape on the horizon, everyone will experience his own personal hell.

Aranyosi suggested a comforting corollary (Aranyosi 2012), based on the idea that multiverse immortality requires not remaining in the “alive state”, but remaining in the *conscious* state, and thus damage to the brain should not be very high. It means, according to Aranyosi, that being in the nearest vicinity of death is less probable than being in just “the vicinity of the vicinity”: the difference is akin to the difference between constant agony and short-term health improvement. However, it is well known that very chronic states of health exist which don’t affect consciousness are possible, e.g. cancer, whole-body paralysis, depression, and lock-in syndrome.

However, these bad outcomes become less probable for people living in the 21st century, as developments in medical technology increase the number of possible futures in which any disease can be cured, or where a person will be put in cryostasis, or wake up in the next level of a nested simulation. Aranyosi suggested several other reasons why eternal suffering is less probable:

1. *Early escape from a bad situation*: “According to my line of thought, you should rather expect to always luckily avoid life-threatening events in infinitely many such crossing attempts, by not being hit (too hard) by a car to begin with. That is so because according to my argument the branching of the world, relevant from the subjective perspective, takes place earlier than it does according to Lewis. According to him, it takes place just before the moment of death, according to my reasoning it takes place just before the moment of losing consciousness” (Aranyosi 2012, p.255).
2. *Limits of suffering.* “The more damage your brain suffers, the less you are able to suffer” (Aranyosi 2012, p.257).
3. *Inability to remember suffering.* “Emergence from coma or the vegetative state is followed by amnesia is not an eternal life of suffering, but rather one extremely brief moment of possibly painful self-awareness – call it the ‘Momentary Life’ scenario.” (Aranyosi 2012, p.257).

## 4.4 Bad infinities and bad circles

Multiverse immortality may cause one to be locked into a very stable but improbable world – much like the scenario in the episode “White Christmas” of the TV series “Black Mirror (Watkins 2014),” in which a character is locked into a simulation of a room for a subjective 30 million years.

Another bad option is a *circular chain of observer-moments*. Multiverse immortality does not require that the “next” moment will be in the actual future, especially in the timeless universe, where all moments are equally actual. Thus a “Groundhog Day” scenario becomes possible. The circle could be very short, like several seconds, in which a dying consciousness repeatedly returns to the same state as several seconds ago, and as it doesn’t have any future moments it resets to the last similar moment. Surely, this could happen only in a very narrow state of consciousness, where the internal clock and memory are damaged.

## 4.5 Marginalization of measure, strange worlds and infinite torture

Because of the “marginalization of measure” some very improbable and extremely bad worlds could receive a higher subjective probability. For example, a world in which future AI decided to torture human beings eternally has *a priori* low probability, but it could be one of the most probable ways to survive almost eternally in the human form7. This is one of the types of s-risks connected with superintelligence (Sotala and Gloor 2017). In other words, MI subjectively increases the s-risks connected with AI.

## 4.6 Dangers of mistakenly ignoring MI and of MI-euphoria

If MI is true these several extremely bad outcomes could be real possibilities for any observer. That is why trying to dismiss MI as a weird probabilistic trick may have an almost infinite negative cost. People who strongly emotionally believe in MI may choose more risky actions, such as not fastening seat belts in a car. But the same argument is even more applicable to religious people who believe in fate. Fatalism is associated with a higher level of accidents (Şimşekoğlu et al. 2013), and the highest world level is in Iran a country of particularly devout people. However, the properties of MI could be used to replace bad outcomes with good outcomes for people who are currently living.

# 5. MI as a decision theory problem

##

## 5.1. MI and expected utility

One of the main objections to the MI is the “measure decline” problem, that is, the decline of the number of copies of a given person in the multiverse. It may sound exotic, but can be presented in the form of utility expectations. In a personal communication, V. Kosoy suggested that the idea of QI should not affect our choices from the point of view of decision theory if expected utilities are correctly calculated.

Imagine a thought experiment, QS combined with a lottery (QS lottery). Bob is trying to earn money via a QS experiment (don’t do it, as most likely you will end up without money and seriously injured or dead). He asks his friend Alice to buy a lottery ticket and he takes a sleeping pill. If the ticket is not a winner, Alice kills Bob while he is asleep. Thus, Bob will wake up only in the worlds in which he won the lottery (and Alice didn’t kill him and run away with the winning ticket).

However, the question “Did Bob actually win anything in the game?” depends on how the calculation of expected utility is performed. For example, if Bob plays a normal lottery with a 0.1 chance of winning 1000 USD, then the expected utility of the game is 100 USD. According to the expected utility objection, this expected utility doesn’t change regardless of if the person assumes MWI or not. If Bob doesn’t assume MWI, he has a 0.1 chance to win 1000 USD, i.e. 100 USD of expected utility. If Bob assumes that MWI is true, Bob should expect that of 10 of his copies, one will win 1000 USD and nine copies will win nothing, so, given that Bob is *randomly distributed among his copies*, Bob again has 100 USD of expected utility.

However, in the case of QS the situation is different, as Bob can’t be distributed between non-existent copies. In other words, if it is assumed that Bob is randomly distributed between his actually extant copies at moment *t*+1, the QS game becomes a winning game (as only one copy exists); however, looking at the “worlds” without Bob, the game becomes unattractive. Again, the question otherwise known as the problem of the validity of quantum immortality arises: should a decline in the number of one’s copies be regarded as death?

This thought experiment could be elaborated depending on the nature of one’s preference system. For example, imagine a mother needs 1000 USD to pay for a cure for her child. In this case, it is obvious that if she plays a QS lottery, and wins 1000 in 1/10 of the worlds, she left her child without the cure (and a mother) in 9/10 of the worlds, and this is not what she wants. Such values could be called “world-caring”. Most of the evolutionary evolved values are world-caring, as evolution favors those who have descendants in the largest share of all possible worlds.

However, there is another type of values, which doesn’t depend on the number of worlds, but instead about the existence of at least one world with some property. An example might be illustrated by a person who cares only about what he might feel. In that case, he could ignore worlds, where he doesn’t feel anything because he doesn’t exist there.

These two types of preferences and corresponding solutions of QS expected utility correspond to the two most accepted decision theories: causal decision theory (which cares about worlds) and evidential decision theory (which cares about experiences) ( as explored by many, e.g. Soares and Levinstein 2017). Evidential decision theory seems to favor quantum immortality, as it recommends actions based only on existing experiences, and as the person could only experience not being killed in QS.

However, both of these theories are not perfect: they don’t account for the behavior of other agents which use the same theory, and, as a result, they recommend agents should defect against each other in prisoner’s dilemma-type situations. For example, even if QS lottery is a good strategy for one agent, it would be a complete disaster if everyone constantly implemented that strategy because the world would be soon empty, and money would lose its value.

To compensate for the shortcomings of existing theories, Wei Dai invented Updateless Decision Theory (UDT); similar theories have also been put forward by others. The formulation of UDT is: “UDT specifies that the optimal agent is the one with the best algorithm – the best mapping from observations to actions – across a probability distribution of all world-histories” (LessWrong Wiki 2018; Armstrong 2011). This means that if such a theory is implemented almost all agents will typically win. UDT argues against a QS lottery, as if everybody plays such a lottery, everybody will eventually lose; however, it argues for cryonics.

While many human goals are in fact oriented in the external world, there is one important exception. Many people have an extremely negative preference for non-existence, also known as the *fear of death*. MI claims that there is always some form of survival – perhaps not a desirable form – but survival only *in some share of all worlds*, which is called “*measure*” because in quantum case (as well as infinite) we can’t count worlds numerically but still need a way to distinguish between more abundant world and less abundant. The *measure* could be understood as “amount of existence” or the thickness of the Everett’s multiverse branch, but in normal cases, if the “measure” will be measured, it will be just a Born’s rule’s probability. In the multiverse, there is a difference between having a one percent chance of survival and surviving with a measure of one percent, but this difference typically disappears in the expected utility calculations. The difference may be equal to the difference of winning 1000 USD with 0.01 probability and having 10 USD. Both events have the same expected utility, but could have different personal meanings depending on details of the preferences.

If someone has a preference for existence, MI offers a big utility gain. Moreover, if everybody is freed of the paralyzing fear of death, it may have a global positive outcome. But normal MI is bad, as shown in the previous section, so normal MI, if widely accepted, will have a negative impact on the wellbeing of people who were currently alive. However, in the case of three alternatives, MI changes the calculations of expected utility.

## 5.2 Merging timelines compensate the decline of measure

As discussed in the previous subsection, one of the main arguments against QI is that a person’s “measure” (or share of all worlds) declines while branching continues. However, as the person’s mental state becomes simpler when she is closer to death, she is *de facto* merged with “other-she” who also has a simpler mind now, but who in the past was different from the observer.

For example, imagine a thought experiment where there is an observer Alice1 with memories (M1, M2… Mt) and another person, Alice2, with memories (N1, M2…Mt); they have a different first memory M1. Maybe Alice1’s first cup was green and Alice2’s was yellow. If, because of the onset of Alzheimer’s, they both forget the color of their first cup, they are now the same person with memories (M2…Mt), which basically means that Alice1’s “measure” doubled! In the case of the complexity of observer-moments, the result will be the same: the closer to death, the simpler the person’s observer-moment, and the more people she will merge with. Pereira suggested essentially the same argument as counter-evidence of the existence of very complex minds, calling it the Super-Strong Self-Sampling Assumption (Pereira 2017).

This simplification of the observer could compensate for the loss of measure because of MI. In our example, after the measure of Alices doubles, one of them could die, and the measure will again be reduced to 1. Losing 1 bit of complexity produces a doubling of the measure, so the simplification process could generate dramatic jumps in measure. For example, forgetting just 1 kilobit of personal information is equal to a 10300 jump of measure – enough to run the QS experiment with a 0.5 probability of death 1000 times.

If this logic about an increase in measure is assumed to be true, one could spend an evening hitting his head with a stone and thus losing more and more memories, and obtaining a higher and higher share of the universe. This is obviously absurd behavior for a human being, but could be a failure mode for an AI, which uses the utility equation with the measure multiplier to calculate the expected utility of future states. In other words, an attempt to exclude quantum immortality by multiplying one’s gains in measure changes may pave the way to even more absurd conclusions.

But this seems to work only in the case of big world immortality, where copies are not causally connected. Whether the timelines merge in the case of the multiverse interpretation of QM is not clear. One view is that in QM, a single event may have multiple histories (Feynman, Hibbs, and Styer 2010). In that case, merging seems to be possible.

Almond suggested (Almond 2010) another argument against QI: that the decline of the number of observers means the decline of the probability to be one of them, the same way, as in the multiverse Sleeping Beauty thought experiment she is less likely to be on Tuesday, as there is only one her copy on Tuesday, but two copies on Monday. However, in case of QI, the person always knows her time position (as t+1 in QI formalism) or Tuesday in the Sleeping Beauty case and there is no “dissolving” of the share of observers, at least until the observers start to forget his time position due to the brain damage.

5.3. Is MI only a gambler’s fallacy?

One of the main questions raised by many researchers of quantum immortality is the following: can repeated survivals in the quantum suicide experiment serve as evidence that MWI is true? Tegmark thinks that it is so (Tegmark 2017), but others suggested that such survival will be only something akin to the “gambler’s fallacy”, that is, even if a person has survived a few rounds of QS, it is not an evidence of any subsequent survival or of very large size of the universe. An example of such thinking was presented by Mallah: “Suppose there are 10 billion people, and 200 of them decide to try QS, so about 100 of those survive. The effective probability of a person being any one of those QS survivors is about 100 in 10 billion. This is true in either the single-world or MWI case, so seeing that you are a QS survivor does not provide evidence either for or against the MWI” (Mallah 2009, p.5).

However, the power of this counterevidence depends on the total number of people in the world. In the Mallah’s example above, it is not surprising (even for an outside viewer) if around 8 rounds (≈ log2200) of the QS lottery has survivors. However, there is only a 1 in 4096 chance in the classical world for survivors after 20 rounds, and thus their existence could be taken as evidence that actual size of the world is much larger, i.e. it could serve as evidence for the existence of some form of the multiverse. A non-violent form of such an experiment would be survival of an observer until around 140 years old, as for centenarians, median life expectancy is around 1 year, and the longest-lived person ever recorded lived only to 122.

# 6. Possible practical applications of the multiverse immortality

##

## 6.1 Surviving until life extension technologies

MI – from a subjective perspective – may assist in escaping death, but not injuries, and thus means an extended period of aging and dying, and a corresponding increase in suffering in the “ordinary,” low-tech world. Thus, there is nothing *a priori* good about MI. But MI could be useful if interpreted as a probability shift in the direction required for high-quality extended life and humanity avoids putting itself at needless risk by attempting “suicide experiments”. For example, it can be posited that a sufficiently long aging process will help humans who are alive now to survive until the appearance of powerful life extension technologies (De Grey and Rae 2007). These technologies will produce rejuvenation and might help us to live almost forever, even without the help of MI. The additional survival time needed to reach “longevity escape velocity” may be just several decades, and in some cases, might mean that a person does not die at the age of 50 years, but instead survives to 90. This is a rather small shift in life expectancy, which has an *a priori* probability of a few percent, and may be not very surprising for the person in question.

Trying to live for as long as possible provides an additional probability shift from multiverse-immortality, but even if that doesn’t exist, benefits still result. Thus, MI should affect decisions regarding life extension. However, MI does imply that it is better to choose medical treatments in a manner similar to Russian roulette; for example, surgery over chemo, as an outcome of immediate death outcome does not exist, but slow dying still exists.

## 6.2 Multiverse immortality significantly increases the chances of the success of cryonics

Multiverse immortality significantly increases the chances of cryonics succeeding at life extension. (“Cryonics” here should be regarded as a placeholder for any radical life extension or resurrection technology, there is nothing especially “quantum” in cryonics.) If MI is true, the biggest share of timelines where I survive until 2100 includes cryopreservation, as was mentioned by (Randall 2004). For example, if cryopreservation’s success chances are 0.1 per cent, and my chance to live until the year 2100 naturally is 1 in a million (excluding the effects of the new life extension technologies described above), that means that I have 1000 times higher chance to survive to 2100 because of cryonics, compared with my personal longevity. As MI means that some instance of me will survive to 2100, it is implied that my chances of cryopreservation success grow to 50 percent (which is our estimation that MI is possible), or 500 times. In other words, no matter how small the chance of cryonics extending life may be, MI increases them up to the level of its own probability of success.

This has decision-level consequences: if, initially, you think that cryonics is not worth trying, then now, after learning about MI, you just have to sign up for cryonics, and MI “will do” the rest. However, if you want to be sure that your friends will also survive, you will still have to invest in the global quality and probability of the success of cryotechnologies. Signing up for cryonics replaces the default outcome of the “bad immortality” of infinitely long aging, with a good outcome of resurrection. Signing up for cryonics is also good from the UDT point of view, as it means that other people, who have similar lines of thought as you will also sign up; if more people sign up, cryocompanies will have more money for research in improving the technology and lowering its price. Expected utility calculations for cryonics for plausible scenarios both with and without MI, taking the risks of eternal suffering into account, is presented in Table 3.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Normal aging**  | **Signing up for cryonics** | **Expected utility calculations** |
| **Different outcomes for 1 million possible future timelines** | 1 survives until 150 years old, suffering. | 1 survives until 150 years old, suffering | Eternal suffering = –1000000 |
| 0 cryo-resurrected | 10,000 cryo-resurrected | Price of signing = – 1 000Resurrection = +10 000 |
| 999,999 died before 150 | 990,999 dead before 150 | Non-existence = –1000 |
| **No MI** | 99.9999% chance of non-existence | 98.9999% chance of non-existence, 1% chance of cryoresurrection0.00001% chance of eternal suffering | No cryo: -1001Cryo: - 1000-1+100-990= - 1891 Signing up for cryonics has **lower** utility |
| **MI probability shift:** | 100% chance of long-term suffering | 99.99% chance of cryoresurrection,0.01% chance of eternal suffering | No cryo: -1000000Cryo: 9900 - 10 000= -100Signing up for cryonics has **higher** utility |
| **MI with credence P=0.5** | 50% chance of long-term suffering, 50% chance of non-existence | 49.95% chance of cryoresurrection0.005% chance of suffering50% chance of non-existence | No cryo: - 500 500Cryo: 4990-5000-500 = -510Signing up for cryonics has **higher** utility |

*Table 3. An example where signing up for cryonics significantly lowers the chances of the worst outcome, that is, eternal suffering. Some plausible estimates of probability of different outcomes and utility are used to demonstrate that if MI is valid, cryonics becomes best choice, but if MI is not valid, cryonics – under the assumptions of its utility and probability listed in right column – is not the best choice.*

## 6.3 The impossibility of euthanasia and the need for cryothanasia

Euthanasia assumes that the voluntary death of a terminally ill patient will have two practical results: the cessation of personal suffering, and alleviation of the burden on the patient’s relatives. The first seems to be the most important reason for euthanasia, but if MI is true, in his subjective timeline the patient will only experience the failure of the procedure, which most likely means that his suffering will continue, and will be even worse8. Also, her hope of stopping her suffering will be ruined, and she will witness the shock of her relatives. Even if our credence in MI is 0.5, it still implies a very high probability of significantly negative outcomes.

However, if one chooses a combination of euthanasia and cryopreservation, called cryothanasia (Istvan 2014; Minerva and Sandberg 2017), the most probable line of survival will be regaining consciousness in the future when a cure for the illness has been developed. Surprisingly, this is not a popular idea.

## 6.4 Multiverse immortality increases the chances of the success of digital immortality, and of the acausal trade between multiverse’s branches

Similar logic to that which favors cryopreservation in case of MI is also applicable to *digital immortality* (DI), i.e. the idea that humanity will be reconstructed by future AI based on their information traces (Bell and Gray 2000; Turchin 2018b). In the face of multiverse immortality, DI becomes one of the most probable outcomes out of all possible ways of survival. DI could be an even more likely outcome than cryonics, as many people could be reconstructed via DI even if they didn’t sign up for cryonics. The idea of the resurrection of the dead as a “common task” comes from Russian cosmists, and first of all Fedorov (Young 2012), who hoped that all people who lived in the past will be eventually resurrected – and this surely will happen is some timelines!

However, there is a problem of information loss and resulting non-perfect resurrections. One possible solution is to replace lost pieces of information with random information (Almond, 2006). In some branches of MWI these random data will be the same as my data, so an effect similar to MI will be used to reconstruct exactly me. However, what should be done with the copies that are not exactly similar? In fact, this is not a big problem, as these not-similar minds will be exact resurrections of other minds from different universes, and these universes, in turn, will resurrect me; this could be called cross-resurrection. As a result, the total measure of all resurrections will be the same and as high as if I were to be resurrected without the use of random noise9.

## 6.5 “Universal problem solver” as an example of bad implementation of MI

Imagine a bomb, connected to a computer, which will kill me only if a certain condition is not met. In that case, I will survive only in the worlds in which the condition occurs. This system is an example of a “universal problem solver”; another specific case is the QS lottery discussed above. This is a bad application of MI, as the bomb might not work, or might only partially work, leaving me injured but not dead; after a few attempts, such outcomes will dominate. But the main reason is that if this practice becomes widespread, I (and everybody else) will soon find myself in an empty world, where everybody kills themselves for some reason. This would be a global catastrophe leading to human extinction. It is unlikely that I would be first and only one who uses the idea, so the practice would be catastrophic for my world, and any personal gain would be outweighed by the potential for widespread damage.

Almond suggested “civilization-level QS”, in which the whole civilization plays Russian roulette, hoping to win needed outcome and escape the problem of survival of only a part of its members (Almond, 2008). But such a civilization will not escape from the curse of low-measure marginalization, as it could end up in a simulation run by an adverse supercivilization. This consideration is based on the simulation argument by (Bostrom 2003) which suggests that significant share of all civilizations are past simulations running by future supercivilization, and the idea that different supercivilizations may have different goals for running essentially the same simulation, for example, some of them may be interested in performing non-ethical experiments, and this could be called “adverse simulation.”

It is improbable that civilization-level QS might offer a solution for the Fermi paradox because it is unlikely that *all* civilizations will choose this path, as it is clearly very risky; thus, it fails “non-exclusivity” principle for any valid Fermi paradox solution (Brin 1983).

## 6.6 Collective survival in a global catastrophe

There are several ways how MI could affect the subjective probability of human extinction risks.

1. *Survival of closely connected groups of people.* There are several hypothetical situations in which the death of a person is connected by highly reliable causal connections with the death of a group of people. For example, the collapse of a submerged submarine will probably kill everybody on board. In such cases, personal survival will mean that the submarine collapse does not happen, and it will appear that MI enhances the survival of the ship. The author has previously explored nuclear submarines as possible survival refuges in the case of a global catastrophe, and this effect could help increase their survivability (Turchin and Green 2017). The same is true for a spaceship. If one sits near the potential target of a nuclear strike, it also could lessen the observed probability of the nuclear war, so this approach could be used as “protection” against other global risks.

2. *Observational selection effects in probability estimation of the past catastrophes*. Some have suggested that the fact that humanity has survived without nuclear war may be explained by MI (Kaufman 2013). However, it is unlikely that WW3 will result in guaranteed human extinction; there will most likely be *some* survivors, and so this explanation seems invalid. Others have even suggested that CERN’s collider could end the world via some disaster stemming from an experiment, and humanity survives (by luck) only in timelines in which the collider experiences some technical difficulties (e.g., (Sandberg 2008) who offers some nice Bayesian calculations and (Yudkowsky 2008)).

3. *Lone wolf survival*. Even if a global catastrophe (like a pandemic or asteroid impact) occurs, MI ensures that I will be the only survivor, as can’t be killed from my observational point of view, but any other human being could. Thus, such an event will not technically be human extinction, but it is unlikely that I will be able to restart the civilization alone. However, I could be saved by aliens, as in a short story by (Wilson 1998), a perfect illustration of low-measure marginalization.

4. *Civilizational survival*. Moreover, as Phil Torres mentioned (Phil Torres 2018), there will be always timelines, where the whole human civilization survives any existential risks, and in long term such timelines will dominate the number of timelines with only one survivor. This could be regarded as a Plan D of global risks prevention, after plan A: prevention, plan B: survival in refuges (Turchin 2018a), plan C: leaving traces about humanity (Turchin and Denkenberger 2018).

## Table 4. The outcome matrix for multiverse immortality and significant life choices

|  | **Normal aging** | **Suicide** | **Euthanasia** | **Life extension** **with the goal** **to survive** **until strong AI** | **Cryonics** | **Cryo-thanasia** | **Digital immortality** | **Merge with AI** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MI** **is true** | Risk of eternal sufferings | Very bad (injury) | Very bad | Very good | Very good | Very Good | Very Good | Very Good |
| **MI** **is false** | Neutral | Bad | Good | Neutral | Neutral | Good | Neutral | Neutral |

## 6.7 Multiverse immortality in the age of superintelligence: will I become AI?

As discussed above, if artificial superintelligence appears, it will able to reconstruct me based on my digital traces. MI increases the subjective probability of observation of the worlds in which AI will be interested in such reconstruction. I hope that the most AI which will practice reconstruction based on the digital immortality will be benevolent and will not cause infinite suffering to the resurrected people (Daniel 2017).

When looking at longer timelines, MI implies that the observer will exist for billions and even trillions of years. The most likely way this could happen is if the observer becomes AI or merges with some future superintelligent AI.

If future advanced AI uses MI to reach some of its goals via a form of QS lottery, it will disappear from our world, which may be good. If the AI is an unfriendly “paperclip maximizer” that wants to kill all humans for their atoms (or any other reason), we could suggest some form of deal (Turchin 2017) in which AI kills all humanity instantly with 0.99 probability, and completely preserve us, or even plays benevolent AI with 0.01 probability. In that case, the AI receives 0.99 of its expected utility of killing humans (like their atoms), but humanity enjoys collective quantum immortality (Almond 2008) and gains complete survival. As a result, the values of both entities will be satisfied.

# Conclusion

In this article, one of the weird probabilistic arguments connected with our place in the universe, i.e. so-called quantum (or multiverse) immortality, was explored. MI has some probability of validity; though its validity is not certain, its probability is large enough to be wary of its bad consequences, or at least to use it as a possible multiplier to boost the chances of success of some already extant processes like cryonics.

Given the significant uncertainty regarding the validity of MI, and the fact that even if it is valid, it is more likely to contribute to bad outcomes, MI can be regarded only as a “plan D” for achieving personal immortality. In this scenario, plan A is life extension, plan B is cryonics and plan C is digital immortality (Turchin 2015). From a practical point of view, MI is best viewed as an additive which increases the chances of success of other life extension technologies.

**Notes**

1. *Use of terms*: ‘multiverse immortality’ is the technically correct term, which is an umbrella term for two separate subtypes: ‘big world immortality’ and ‘quantum immortality’. But as quantum immortality is the most accepted term, it is used in this paper where the difference between QI and MI is not substantial.

2. A good introduction can be found in Wilkins’ analysis (Wilkins 2012).

3. E.g. as seen at <https://www.reddit.com/r/Glitch_in_the_Matrix/comments/6j7pi5/quantum_immortality_is_the_glitch_the_fact_that/>

4. Note that here “endless” does not mean “infinite”, as circular timelines are possible, and actual infinity is impossible.

5. This topic is obviously controversial; for more discussion see the SEP review by (Friederich 2017).

6. As it seems has already happened with Everett’s daughter according to her suicide note where she wishes – obviously impossible from the point of QI thing – “end up in the correct parallel universe to meet up w[ith] Daddy” (Byrne 2010).

7. As in Harlan Ellison’s classic short story “I Have No Mouth, and I Must Scream” (Ellison 1967).

8. This assumes he will have the traumatic experience of something akin to regaining consciousness in a morgue and feel pain again without the help of painkillers, plus whatever damage may result from the method used in the euthanasia attempt.

9. More details about this resurrection approach can be found in (Turchin 2018c).

**References**

Almond, P. 2006. “Many-Worlds Assisted Mind Uploading: A Thought Experiment.” https://web.archive.org/web/20110513092111/http://www.paul-almond.com/ManyWorldsAssistedMindUploading.htm.

———. 2008. “Civilization-Level Quantum Suicide.” https://web.archive.org/web/20120310105846/http://www.paul-almond.com/CivilizationLevelQuantumSuicide.htm.

———. 2010. “Repeated Quantum Suicide Would Not Suggest That MWI Is Correct – Even to the Person Doing It.” https://web.archive.org/web/20120310205434/http://www.paul-almond.com/QuantumSuicide.pdf.

———. 2011a. “Delayed Quantum Suicide: Assuming You Will Survive Quantum Suicide May Be Simplistic.” https://web.archive.org/web/20130503122611/http://www.paul-almond.com/.

———. 2011b. “Quantum Brain Damage: Assuming You Will Survive Quantum Suicide May Be Simplistic.” https://web.archive.org/web/20130503122611/http://www.paul-almond.com/.

———. 2011c. “Quantum Suicide and Inconsistency: Assuming You Will Survive Quantum Suicide May Be Simplistic.” https://web.archive.org/web/20120309233248/http://www.paul-almond.com/QuantumSuicideInconsistency.pdf.

———. 2011d. “Repeated Quantum Suicide Would Not Suggest That MWI Is Correct: A Stronger Version of the Argument.” https://web.archive.org/web/20120310023143/http://www.paul-almond.com/QuantumSuicide2.pdf.

Aranyosi, István. 2012. “Should We Fear Quantum Torment?” *Ratio* 25 (3): 249–259.

Armstrong, Stuart. 2011. “Anthropic Decision Theory.” *ArXiv Preprint ArXiv:1110.6437*.

Bell, Gordon, and Jim Gray. 2000. “Digital Immortality.”

Bergström, Lars. 2012. “Death and Eternal Recurrence.” *The Oxford Handbook of Philosophy of Death*.

Bevers, Brett. 2011. “Can Many-Worlds Survive a Quantum Doomsday.”

Bitbol, Michel. 1994. “IS NOW A MOMENT IN TIME? A Discussion of McTaggart’s Argument Against the Reality of Time From a Transcendental Idealist Standpoint.”

Blanqui, A. 1872. “Eternity Through the Stars by Auguste Blanqui.” 1872. https://www.marxists.org/reference/archive/blanqui/1872/astres.htm.

Bocchieri, P., and A. Loinger. 1957. “Quantum Recurrence Theorem.” *Physical Review* 107 (2): 337.

Borges, Jorge Luis. 2000. *The Library of Babel*. David R. Godine Boston, MA.

Bostrom, N. 1999. “The Doomsday Argument Is Alive and Kicking.” *Mind* 108 (431): 539–551. https://academic.oup.com/mind/article-abstract/108/431/539/985823.

———. 2002. “Self-Locating Belief in Big Worlds: Cosmology’s Missing Link to Observation.” *The Journal of Philosophy* 99 (12): 607–623.

———. 2003. “Are You Living In a Computer Simulation?” *Published in Philosophical Quarterly (2003) Vol. 53, No. 211, Pp. 243-255.*

Bradley, Darren. 2009. “Multiple Universes and Observation Selection Effects.” *American Philosophical Quarterly* 46 (1): 61–72.

Braunstein, Samuel L., and Arun K. Pati. 2007. “Quantum Information Cannot Be Completely Hidden in Correlations: Implications for the Black-Hole Information Paradox.” *Physical Review Letters* 98 (8): 080502.

Brin, Glen David. 1983. “The Great Silence-The Controversy Concerning Extraterrestrial Intelligent Life.” *Quarterly Journal of the Royal Astronomical Society* 24: 283–309.

Brown, Geoffrey. 1984. *The Cosmological Theory of Empedocles*. De Gruyter.

Byrne, Peter. 2010. *The Many Worlds of Hugh Everett III: Multiple Universes, Mutual Assured Destruction, and the Meltdown of a Nuclear Family*. OUP Oxford.

Carrigan Jr, Richard A. 2006. “Do Potential SETI Signals Need to Be Decontaminated?” *Acta Astronautica* 58 (2): 112–17.

Chalmers, D. 1996. *The Conscious Mind*. Oxford University Press, New York.

Cirkovic, Milan M. 2004. “Is Quantum Suicide Painless? On an Apparent Violation of The Principal Principle.” *ArXiv:Quant-Ph/0412147*, December. http://arxiv.org/abs/quant-ph/0412147.

Copi, Irving, and M. Cohen. 1990. *Introduction to Logic*. New York: Macmillan. Google Scholar.

Daniel, M. 2017. “S-Risks: Why They Are the Worst Existential Risks, and How to Prevent Them (EAG Boston 2017).” 2017. https://foundational-research.org/s-risks-talk-eag-boston-2017/.

De Grey, Aubrey, and Michael Rae. 2007. *Ending Aging: The Rejuvenation Breakthroughs That Could Reverse Human Aging in Our Lifetime*. St. Martin’s Press.

Deasy, Daniel. 2015. “The Moving Spotlight Theory.” *Philosophical Studies* 172 (8): 2073–2089.

Deng, Natalja. 2015. “On Whether B-Theoretic Atheists Should Fear Death.” *Philosophia* 43 (4): 1011–21. https://doi.org/10.1007/s11406-015-9638-y.

Egan, Greg. 2009. “Dust Theory FAQ.” 2009. http://www.gregegan.net/PERMUTATION/FAQ/FAQ.html.

Eliade, Mircea. 1949. “The Myth of the Eternal Recurrence.” *Translation Feb Srkaraty,*.

Ellison, Harlan. 1967. *I Have No Mouth, and I Must Scream*. Galaxy Publishing Corp.

Feynman, Richard P., Albert R. Hibbs, and Daniel F. Styer. 2010. *Quantum Mechanics and Path Integrals*. Courier Corporation.

Friederich, Simon. 2017. “Fine-Tuning,” August. https://plato.stanford.edu/archives/fall2017/entries/fine-tuning/#InveGambFallChar.

Frischhut, Akiko M., and Alexander Skiles. 2013. “Time, Modality, and the Unbearable Lightness of Being.” *Thought: A Journal of Philosophy* 2 (3): 264–273.

Higgo, J. 1998. “Does the ‘many-Worlds’ Interpretation of Quantum Mechanics Imply Immortality?” https://web.archive.org/web/20050203192203/http://higgo.com:80/quantum/qti.htm.

Hossenfelder, S. 2018. “Scientific Theory And The Multiverse Madness.” *NPR.Org*, 2018. https://www.npr.org/sections/13.7/2018/01/22/579666359/scientific-theory-and-the-multiverse-madness.

Istvan, Zoltan. 2014. “Should Cryonics, Cryothanasia, and Transhumanism Be Part of the Euthanasia Debate?” *Huffington Post Tech: The Blog* 23.

Jones, Jonathan. 2017. *Technological Resurrection: A Thought Experiment*.

Kant, Immanuel. 1781. *Critique of Pure Reason*. Cambridge university press- 1999.

Kaufman, J. 2013. “Is Our Continued Existence Evidence That Mutually Assured Destruction Worked?” 2013. https://www.lesswrong.com/posts/9miMdHERJFjQDp5uq/is-our-continued-existence-evidence-that-mutually-assured.

Kaufmann, Walter A. 2013. *Nietzsche: Philosopher, Psychologist, Antichrist*. Princeton University Press.

LessWrong Wiki. 2018. “Updateless Decision Theory - Lesswrongwiki.” 2018. https://wiki.lesswrong.com/wiki/Updateless\_decision\_theory.

Lewis, David. 2004. “How Many Lives Has Schrödinger’s Cat?” *Australasian Journal of Philosophy* 82 (1): 3–22.

Loew, Christian. 2017. “Boltzmannian Immortality.” *Erkenntnis* 82 (4): 761–776.

Mallah, Jacques. 2009. “Many-Worlds Interpretations Can Not Imply ‘Quantum Immortality.’” *ArXiv:0902.0187 [Quant-Ph]*, February. http://arxiv.org/abs/0902.0187.

Marchal, Bruno. 1991. “Mechanism and Personal Identity.” In *Proceedings of WOCFAI*, 91:335–345.

Menzel, Christopher. 2017. “Actualism.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2014th ed. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/actualism/.

Minerva, Francesca, and Anders Sandberg. 2017. “Euthanasia and Cryothanasia.” *Bioethics* 31 (7): 526–533.

Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.

Mueller, Markus P. 2017. “Law without Law: From Observer States to Physics via Algorithmic Information Theory.” *ArXiv:1712.01826 [Physics, Physics:Quant-Ph]*, December. http://arxiv.org/abs/1712.01826.

Nietzsche, Friedrich. 1883. *Thus Spoke Zarathustra*. Jester House Publishing.

Page, Don N. 1999. “Observational Consequences of Many-Worlds Quantum Theory.” *ArXiv:Quant-Ph/9904004*, April. http://arxiv.org/abs/quant-ph/9904004.

———. 2007. “Susskind’s Challenge to the Hartle-Hawking No-Boundary Proposal and Possible Resolutions.” *Journal of Cosmology and Astroparticle Physics* 2007 (01): 004–004. https://doi.org/10.1088/1475-7516/2007/01/004.

Parfit, Derek. 1984. *Reasons and Persons*. OUP Oxford.

Penrose, Roger, and Martin Gardner. 2002. *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*. 1 edition. Oxford: Oxford University Press.

Pereira, Toby. 2017. “An Anthropic Argument against the Future Existence of Superintelligent Artificial Intelligence.” *ArXiv:1705.03078 [Cs]*, May. http://arxiv.org/abs/1705.03078.

Randall, Allan F. 2004. “Quantum Miracles and Immortality.” *Quantum* 5: 8.

Sandberg, A. 2008. “Bayes, Moravec and the LHC: Quantum Suicide, Subjective Probability and Conspiracies.” *Andart* (blog). 2008. http://www.aleph.se/andart/archives/2008/09/bayes\_moravec\_and\_the\_lhc\_quantum\_suicide\_subjective\_probability\_and\_conspiracies.html.

Shikhovtsev, E. 2003. “Biographical Sketch Of Hugh Everett, III.” http://space.mit.edu/home/tegmark/everett/everett.html#contents.

“SHOULD WE FEAR QUANTUM TORMENT? - Aranyosi - 2012 - Ratio - Wiley Online Library.” n.d. Accessed January 2, 2018. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9329.2012.00540.x/abstract.

Şimşekoğlu, Özlem, Trond Nordfjærn, Mohsen Fallah Zavareh, Amin Mohamadi Hezaveh, Amir Reza Mamdoohi, and Torbjørn Rundmo. 2013. “Risk Perceptions, Fatalism and Driver Behaviors in Turkey and Iran.” *Safety Science* 59 (November): 187–92. https://doi.org/10.1016/j.ssci.2013.05.014.

Smolin, Lee. 1992. “Did the Universe Evolve?” *Classical and Quantum Gravity* 9 (1): 173.

Soares, Nate, and Benjamin A. Levinstein. 2017. “Cheating Death in Damascus.” Technical report, Machine Intelligence Research Institute, 2017. https://intelligence. org/files/DeathInDamascus. pdf. https://intelligence.org/files/DeathInDamascus.pdf.

Sotala, Kaj, and Lukas Gloor. 2017. “Superintelligence as a Cause or Cure for Risks of Astronomical Suffering.”

Squires, Euan J. 1994. *The Mystery of the Quantum World*. 2 edition. Bristol ; Philadelphia: CRC Press.

Steinhardt, Paul J., and Neil Turok. 2002. “Cosmic Evolution in a Cyclic Universe.” *Physical Review D* 65 (12): 126003.

Tegmark, M. 1998. “The Interpretation of Quantum Mechanics: Many Worlds or Many Words?” *Fortschritte Der Physik* 46 (6–8): 855–62. https://doi.org/10.1002/(SICI)1521-3978(199811)46:6/8<855::AID-PROP855>3.0.CO;2-Q.

———. 2017. “The Universes of Max Tegmark.” http://space.mit.edu/home/tegmark/crazy.html.

Torres, P. 2014. “Why Running Simulations May Mean the End Is Near.” 2014. https://ieet.org/index.php/IEET2/more/torres20141103.

Torres, Phil. 2018. “Phil Torres - Question for Friends More Scientifically Literate...” *Facebook Post* (blog). 2018. https://www.facebook.com/phillip.torres.752/posts/2010728329041160.

Turchin, A. 2015. “Immortality Roadmap - Less Wrong.” *LessWrong* (blog). 2015. http://lesswrong.com/lw/mjl/immortality\_roadmap/.

———. 2017. “Messaging Future AI.” *Manuscript*. https://goo.gl/YArqki.

———. 2018a. “Approaches to The Prevention of Global Catastrophic Risks.” *The Human Prospect*, no. Winter.

———. 2018b. “Digital Immortality: Theory and Protocol for Indirect Mind Uploading.” https://philpapers.org/rec/TURDIT.

———. 2018c. “Resurrection of the Dead via Multiverse-Wide Acausual Cooperation - LessWrong 2.0.” *LessWrong* (blog). 2018. https://www.lesswrong.com/posts/BJfb2hqtdnaRijAfz/resurrection-of-the-dead-via-multiverse-wide-acausual.

Turchin, A., and M. Chernyakov. 2018. “Classification of Approaches to Technological Resurrection.” *Under Review in Post-Humans Studies*. https://philpapers.org/rec/TURCOA-3.

Turchin, A., and D. Denkenberger. 2018. “Surviving Global Risks through the Preservation of Humanity’s Data on the Moon.” *Acta Astronautica*, no. Volume 146, May 2018,: Pages 161-170.

Turchin, A., and B.P Green. 2017. “Aquatic Refuges for Surviving a Global Catastrophe.” *Futures* 89: 26–37. https://doi.org/10.1016/j.futures.2017.03.010.

Watkins, James. 2014. *“Black Mirror” White Christmas (TV Episode 2014)*. https://www.imdb.com/title/tt3973198/reviews.

Wiley, Keith B., and Randal A. Koene. 2015. “The Fallacy of Favoring Gradual Replacement Mind Uploading Over Scan-and-Copy.” *ArXiv Preprint ArXiv:1504.06320*.

Wilkins, Alasdair. 2012. “Quantum Suicide: How to Prove the Multiverse Exists, in the Most Violent Way Possible.” *Io9*, 2012. https://io9.gizmodo.com/5891740/quantum-suicide-how-to-prove-the-multiverse-exists-in-the-most-violent-way-possible.

Wilson, Robert Charles. 1998. “Divided by Infinity.” https://www.tor.com/2010/08/05/divided-by-infinity/.

Young, George M. 2012. *The Russian Cosmists: The Esoteric Futurism of Nikolai Fedorov and His Followers*. Oxford University Press.

Yudkowsky, E. 2008. “How Many LHC Failures Is Too Many?” *LessWrong* (blog). 2008. https://www.lesswrong.com/posts/jE3npTEBtHnZBuAcg/how-many-lhc-failures-is-too-many.

———. 2009. “The Anthropic Trilemma.” 2009. http://lesswrong.com/lw/19d/the\_anthropic\_trilemma/.

# Appendix 1. Multiverse immortality and probabilities

## MI immortality formalism

Imagine that any observer consists of a discrete timer T and identity block I, which is just a constant string of numbers, not changing in time. Any observer-moment O—that is the sum of the experiences of the observer in the moment T—could be described as O(I, T).

In that case, the multiverse immorality claim is: for any O(I,T) for any T and any I exists O(I,T+1):

$∀O\left(I,T\right)\rightarrow ∃O\left(I, T+1\right), for ∀I,∀T$

## MI and probabilities

The questions about probability is notoriously difficult as they always come to the question “what is probability”, especially in the situation with multiverse and many observer’s copies. See great short overview of problem by Wei Dai: <http://lesswrong.com/lw/1iy/what_are_probabilities_anyway/>

My view on the “probabilities” is that it is incorrect to ask “what are they”, as they are not real objects, but instruments which we constructed to predict the world and act in it. So, we can’t pretend that they actually exist.

As “probabilities” are prediction instruments sometimes we could use different prediction instruments, for example, expected utility, which is combination of the probability and our preferences.

It is not easy to discuss QI without applying probabilities to it, and here we explore some probability related questions.

Many authors prove or attack QI exactly based on some probabilistic calculations, like (Mallah 2009), and (Almond 2011d).

D.Page wrote that there will be observational difference between single and many past histories interpretations of QM, because of change of the number of observers (Page 1999).

## Difference in the probabilities of outcomes in the multiverse and quantum immortality

One interesting thing is that in the multiverse immortality the number of “very rare survival” observers could be higher than in the quantum immortality. For example, you are falling on Sun. The probability that some quantum effects temporary change gravity and move you away is extremely small, like the number with millions zeros. However, in the other part of the big universe there are another copies of you, which have an illusion that they are falling on Sun, but actually not. In that case big immortality provides more probable way of survival.

## Marginalization: Strange outcomes will dominate after several QS experiments

Almond called “Low Measure Marginalization”, that is, survival in an unexpected ways (Almond 2011d).

There are several possible types of marginalization:

* Saved by aliens,
* In a dream,
* In a short-term hallucination
* Data stamping mistake
* In a simulation with afterlife
* As a Boltzmann brain
* An error in estimation of risk or in firing of QS weapon
* A jump to past observer-moment
* Respawning

Marginalization means appearing in less probable worlds which are also less likely to be ordered. It has surprising consequence that QI is more likely to work in them! Because the size of set O(I, T+1) is not much smaller than O(I, T) set. That is, “measure” is declining slowly after marginalization. The reason for it is that in the strange non-ordered world all possible observer-moments are more or less equally frequent, so frequency is not declining.

For example, in a bizarre dream you are much more likely to be saved by “aliens”. However, long term stable outcome could be reached only via AI as I will show in the next subsection.

## Many futures and different probability distribution at the moment of death

In the MW interpretation of quantum mechanics, any agent faces many futures which will actually happen with him, so if we don’t have problems with “many futures” idea before the QS experiment, we will not have problems with the idea after the experiment.

However, in the case of the big world acausal immortality, a person has only one future in his normal life, but after the moment of his death, there are infinitely many exact and non-exact copies of him in very different situations. Thus, a reasonable question arises: where exactly his timeline would jump, or, if we rightly decline the idea of “jumping”, the question is what to do with infinitely many futures, which will actually happen simultaneously.

The question, however, is not factual, but normative, the same way as all questions about personal identity.

Multiverse-immortality is not answering to the question *what to expect?* It answers to the question *what you should not expect,* and it tells that *non-existence is impossible*: there always will be next observer-moment. However, correctly predicting observations about the next observer-moment may be easy or difficult depending on many circumstances.

“Many futures” explains that there will be many infinite future timelines with different topology. Some will be circular, some will include infinite growth inside superintelligent AI.

Carl Shulman: “I think this sidesteps the underlying intuitions too quickly. We have cognitive mechanisms to predict "our next experience," memories of this algorithm working well, and preferences in terms of "our next experience." If we become convinced by the data that this model of a unique thread of experience is false, we then have problems in translating preferences defined in terms of that false model. We don't start with total utilitarian-like preferences over the fates of our future copies (i.e. most aren't eager to lower their standard of living by a lot so as to be copied many times (with the copies also having low standards of living)), and one needs to explain why to translate our naive intuitions into the additive framework (rather than something more like averaging). <http://lesswrong.com/lw/6op/preference_for_many_future_worlds/4iuy>

## “Many futures” imply that several types of immortality will happen

“Many futures” imply that several types of immortality will happen, but we don’t know with which probabilities, as to estimate it we need the perfect model of the whole multiverse, humanity possible futures, alien civilizations types, Boltzmann brain physics, nature of time, nature of consciousness etc. There are several plausible candidates:

Long-term outcomes:

**Small time loops** – if there is a series of states of mind where first and last states are equal, they could be regarded as a time loop. But no time loop is eternal, as the same states of mind may be in another larger loop.

**Infinite time loop** – an “empty consciousness” “runs” across all possible beings and returns back to any given being. Equal to eternal return.

**Continuous “reincarnation”** – reseting of the part of the memory and clock

**Becoming an AI with infinitely long memory** – This is actual full immortality as it assumes extremely long existence in linear time with constantly growing memory and time-clock.

**Locked in a simulation** – like in “White Christmas” a character was locked for 30 mln years.

Short-term immediate outcomes after death:

**Low-measure marginalization.**

**Respawning** – appearing in another world but with memories of this world. Requires some other physical laws

**Very long aging in damaged state**. But it eventually become so improbable that the person will be cryopreserved and resurrected by AI – even if such worlds are very rare.

**Appearing on the higher level of the nested simulation**.

**Resurrection after cryopreservation** – even for those who didn’t signed up for cryonics, as there will be world with such resurrection

**Resurrection by a future AI** which will use digital immortality and quantum randomness.

Most short-term outcomes will eventually end up in being either AI or inside the AI-driven simulation, so they converge, except low-level marginalization and respawning. From long term outcomes we are interested only in those where a person actually knows that he has surprisingly long age, so we could ignore loops and reincarnation – and end up with worlds controlled by advance AI (not necessary a computer, but very powerful and stable mind, capable to exist indefinitely long. But the only rational type of such mind – which we could imagine now – is an advance AI).

Such an AI may be either benevolent to humans, so I could merge with it or live in pleasant simulations, or an Evil AI, which is designed to torture human beings: either for blackmail or because an error in its utility function, as described here: <https://reducing-suffering.org/near-miss/>

## “Believer against one’s will” thought experiment

After Mary survived several rounds of QS, she could say: “I still don’t believe that MWI is true, and I think that my survival is pure randomness. I also think that I will most likely completely disappear after the next rounds. I want to drink water now, but I will not drink water, as the next several rounds of QS will happen before I will able to get to the bottle, and I will be dead in 999 of 1000 of futures.” After several rounds she becomes even more thirsty, and decides to go for the water bottle anyway.

Here her subjective surprise of survival will decrease with each round, and she may even start making plans which include survival after several QS, and which put zero utility to the dead ends.

That is, after several survivals, subjective experience and planning will be in contrast to theoretical understanding of Mary. This is similar to the *survivor fallacy* or *gambler fallacy* by its psychological effects. However, the gambler will always be punished by his overconfidence in his luck— and Mary may become a true believer, but will ultimately be punished by what Almond called “low measure marginalization”, that is, she will survive in unexpected ways (Almond, 2011d).

## Multiverse-immortality for computer algorithms

Yudkowsky wrote “The Anthropic Trilemma” <http://lesswrong.com/lw/19d/the_anthropic_trilemma/>, about manipulation of observed probabilities by creation of large numbers of virtual copies, and manipulation of this number. But any computer program could exist in many copies, and the Internet is a big world for it.

# Appendix 2. Multiverse immortality and other similar probabilistic arguments

## Multiverse-immortality and simulation argument

Simulation argument contribute powerfully to the big world idea and, if we buy it, most my copies (and me now) are in some forms of simulations which are created by advanced civilizations for whatever reason. The number of my copies in simulations is much larger than the expected number of my copies in any other big world model (because simulations are optimized to create conscious observers).

If my simulation is turned off or if I die in it, where will be always another simulation where my copy is still running, so soon I will find myself in the simulations which are designed to preserve the life of characters.

If I am not in a simulation, the simulation probability (that is I am in a simulation) increases after number of “deaths”, because in a simulation even weirdest ways of survival are possible. If I fail in 1 km abyss I will likely to survive only in the simulation where miracle ability to fly is turned off – or I will wake up from the dream (as dreams are in fact most widespread type of simulations).

However, dying is also increasing the probability that I will “wake up” on a higher level of the simulation, as if I was a player on computer game. Surely, it is possible that our world is a computer game for a higher-level character, the same way as we play in RPG (or have dreams). So, after waking up I will “remember” who I actually am, but there are many different variants of it (see subsection below): the same game-character could be played by different players. These players could be played by even higher level players in the nested simulation, so the idea of multiverse immortality stacked well with it (P. Torres 2014).

There is probably also a share of simulations which simulate “afterlife” according to human believes about it. This share will also increase after a number of my “deaths” as I will be able to find myself only in the simulations where there is afterlife, miracle survival or awakening, and will not find myself into simulations which simply shutdown or terminate my file.

The phenomenon of respawning often reported by people (where they find themselves before a deadly accident and live this moment the second time without dying) may be also explained by the fact we are in the simulation.

## Multiverse-immortality and Doomsday argument

 As my “measure” decreases, it is less likely that I will find myself in the later stages of “immortal timeline” if I think that current moment is random.

Almond on DA and QI: “The Doomsday argument, if correct, suggests that our civilization will end soon, based on the idea that our own position in time should not be too unusual, but in an MWI context, the Doomsday argument could be interpreted as meaning that our civilization will have a huge and ongoing decrease of measure in the future: that our civilization will continue in some futures, but that its measure will decrease all the time as something like the quantum suicide experiment – whether contrived by our descendants or imposed on them – occurs on a continual basis. In such a situation, observers should expect to find themselves in relatively high measure parts of the civilization’s history – maybe where we are now” (Almond 2011d).

Note, that merging timelines may counteract DA.

The Almond’s argument is valid only for the “new born” observers, which are more likely to be located in the regions with high measure. However, if some has a counter (t+1), only the observers with the same timestap should be counted: if I know that it is Tuesday, I can’t be randomly selected from all who are located in Monday and Tuesday, even if there are much more observers on Monday.

## Modal realism and back causation

If everything possible exists, I should exist. However, if I exist, there should exists most probable way of my appearing, and Darwinian evolution seems to be the one. This necessity to create “me” create something resembling back-causation pressure on the past events, the same way as my father and mother has to meet no matter how difficult it was.

## Multiverse-immortality reset and infinite timelines in the space of possible minds

Another possible type or resetting could happen in the case of “slow death”. In that case, my state of mind will be simpler and I will forget part of my personal identity as well as my timestamp. This partial simpler observer-moment could be equal to many more observer-moments of larger groups of people, which also experience these narrow states on consciousness.

This could produce something similar to “reincarnation”. For example, I forget my surname and my age. Now I am just Alexey and my state of consciousness could be equal to a state of any child with the same name.

However, it creates frightening possibility that every night when my state of mind narrows, I could become everyone else and randomly jump inside any person with similar narrow state of mind. One guy on LessWrong had panic attacks because of it.

## The same “me” in different worlds and superposition of interpretations

*Universal reference class* by Almond: “In previous articles, the author gave the term ultimate reference class to the set of all possible descriptions of your situation – all possible situations in which you could be right now” (Almond 2011d). *Ultimate reference class* includes all possible situations which creates given observer-moment, and includes all the same observer-moments in these situations.

It means that questions about actual nature of reality is useless, as the same observer-moment could be Boltzmann-brain, in different simulations and in different real worlds, and even belong to different peoples.

The next step in this logic is to suggest that these different interpretations of the same observed reality do not just linearly sum up, but could interfere with each other in the similar way as different paths of a quantum particle interfere. In that case we will find ourselves in the reality which resists attempts to any final interpretation and this will fundamentally limit our ability to predict future outcomes.

## Multiverse-immortality and qualia

Observer-moments consist of *atoms of experience* which are qualia. That is, qualia is minimal living parts of consciousness. In the moment of death consciousness dissipate, but individual qualia can’t cease to exist, because *green* can’t stop to be *green*. Simple qualia could exist in may be in nature, that is like dust universe, or in other people’s minds. So, qualia are themselves immortal, but also don’t have anything personal.

In other words, an observer could fail to smallest pieces which are separate qualia. But it is no more immortality than immortality of my physical atoms. But eventually memory identity and qualia identity could merge again.

# Appendix 3. Other objections to QI and wrong interpretations

## Aging and slow death

Tegmark suggested that the main counterargument for multiverse immortality is that it is not applicable to the situation of slow death, like death of aging. In these situations, I slowly lose most of my identity which is based on memories. At the end rather simple mind remains, but it is not me. This mind maybe still able to suffer, but it lost most of my valuable traits.

One outcome of it is that I will become equal to so many other simple minds that I could become almost anyone else, and it is similar to “reincarnation”.

Here I explore another outcome, based on the situation where I will regain most of my memories.

Imagine a situation that I am gradually losing memories (and memories is all that is needed for personal identity). For example, my memories consisted of 10 numbers, like 1234543210, but now I remember only first 3, that is 123. However, there is still some possible future timelines where I will regain my memories, where my brain function improves, and I will again remember 1234543210. In that case, I could percept all the story as jumping from the last moment where I had full memories to the first moment when I again have all full memories, similar to the situation where will lost our memories during night sleep but without problems regain them in the morning and often feel it as if we jumped from the evening to the morning, and forgetting almost all period of narrow consciousness during night dream.

It means that if my identity-memory will fall below some threshold, I will jump over a period of narrow consciousness to the next observer-state where I will regain my memory, and thanks to the multiverse immortality where will be always such next me, who regain the memories.

Almond suggested “Quantum Brain Damage”:

“A more general version of the thought experiment, the quantum brain damage thought experiment, shows problems with such a simplistic view. In this thought experiment, the quantum outcome determines a degree of damage to your brain, ranging from no damage at all to complete destruction, with many intermediate degrees of damage”.

In that case, arises the problem of personal identity and death definition, as it is not clear how to draw the line between “me” and not-me (Almond 2011b). But existence of timelines where I will remember, who I am, makes the problem of the threshold irrelevant: I will always emerge above any threshold.

Another thought experiment by Almond: “A more general version of the thought experiment, the delayed quantum suicide thought experiment, shows problems with such a simplistic view. In this thought experiment, there is a variable time delay between the quantum event and your death occurring, ranging from a very long time delay to a negligibly short time delay, with many intermediate time delays being possible” (Almond 2011a).

## Not-exactly-me copies and jumping over bad event via forgetting

Not-exactly-me my copies dominate total number of copies. For example, if I go out and have a fatal accident, there is a lot my copies in other worlds which didn’t go out at all in that day. The next day, the fact of did they go out or not will completely vanish from their memory and will not become a part of their personal identity. So, timelines (went out on Monday and not hit by a car) and (didn’t go out on Monday) will merge on Tuesday, and the first one could dominate over time as more and more timelines will forget going out. As a result, the most of timelines where I survive the car accident will be not the ones where I miraculously dodge the car or survive the crash, but those where **I didn’t remember going out at all**. See the similar lines of reasoning in (Aranyosi 2012).

It is especially true for older people whose personal identity is very rigid and is not evolving every moment and whose memory is not very strong. They could be regarded and fell themselves as the same person even after years between two moments in life. So, they could even “jump over” the whole deadly illness. For example, John (t=60) gets deadly cancer. For John (t=61) the most likely outcome to be alive is that he didn’t become ill at all.

## Resetting of timeclock, respawning and reference class problem

As we said before, multiverse immortality could be formulated in the form that for any observer-moment O (I, T) exists observer-moment O (I, T+1).

However, Almond showed (Almond 2011d) that the “measure” (roughly speaking, the number of copies in quantum world) is decreasing in the transition from 1 to 2, and if a subject can’t measure the time T properly, he will be in the situation similar to Sleeping Beauty experiment, and (with some assumption how we solve Sleeping beauty) he will more likely find himself before the QS experiment.

For example, if he doesn’t know if the quantum suicide experiment already happened or not in the moment T0, he should think that is doesn’t happened yet, as the number of copies before experiment is higher.

However, there is another possible solution of the QS experiment. That the experimenter dreamed about being in the moments T and T+1, but in fact he is still in the moment T-10. For example, he had a false awakening in the morning before the experiment and dreamed that he performed the experiment, but after experiment he awake again in his bed in the moment T-10, but having memories about realistic dream where he has the experiment.

In the real settings, probably 1 of 100 experimenters will have dreams about the planed QS experiment the night before it. If the probability to survive the QS suicide is set 1 in 1000, it means that from observational point of view the most probable outcome of experiment us to find one self before the experiment with false memories about the experiment!

However, the person could awake from the dream about QS experiment not in his timeline. I mean that after the experimenter actually performed QS experiment, he will find in the most probable intstance of (I, T+1) in the universe, and many such moments will be the false awakening dreams of another experimenters in another world, who are not yet started the experiment. That is, actual timeline could continue in someone’s dream. (Remember, that it still not enough to explain reports about respawning, as classical big world immortality doesn’t allow any information transfer or premonition visions, which often described in the respawning cases. Only if some form of evolutionary evolved ways of orchestrating quantum collapse (Penrose and Gardner 2002) exists for probability manipulation (Yudkowsky 2009), it could become possible. However, rigorous thinking on this topic is not easy as it is field of quantum woo and out of scope of this article.)

The less likely is the survival in QS experiment, the more likely is the outcome of it is which the next observer state will be in the world where he (or his copy) only hallucinated about the experiment.

Almond about time resetting: “if you had somehow lost track of how old you were – say you had severe amnesia – you should find it unlikely that you were in one of the observer moments in a category with a low combined measure, if these exist, corresponding to being, for example, millions of years old” (Almond 2011d).

## General anti-immortality principle based on observed age

The idea of this objection is that if immortality will be possible, including Multiverse-immortality, when a random observer should find his age infinitely higher that median age of people around himself (Mallah 2009). His idea is exactly the Doomsday argument, applied to the individual human life expectancy. It is surprising that after all attempts to disprove it as invalid, DA is now was revived only to kill another weird probabilistic idea.

However, all anti Doomsday argument arguments works here, like that my position is not random, as I still don’t know is multiverse immortality is applicable to me. Also, the fact that my measure decreasing may play the role here, as a random moment would be more likely to be in beginning of times, but the moment of now may not be regarded random for those who know about QI for a long time.

## QI is only artefact of thinking, not the property of the world

Nesov suggested that QI appears only in thinking in the world, but not a real thing in the world: “Note that this statement talks about a ritual of cognition, not about the world: it talks about what one can *remember*, but obviously it's possible to *infer* that in some circumstances you'll die, or that in counterfactuals following different past events you've died. So, this kind of "immortality" is an artefact of an artificial limitation on the ways of conceptualizing real world, one that can easily be lifted and thus shown to be not about an actual property of the world”. <http://lesswrong.com/lw/87o/a_pessimistic_view_of_quantum_immortality/548k>

However, actual statement of QI is about the state of the world, not about the cognition: It states that O(I, T+1) moment always actually exists.

## Adding up to normality

One of meta-level objections to this type of arguments is that “its all should come to normality”. However, the correct presentation of this principle is that “its all should look like normality in the near vicinity” of us, and MI doesn’t contradict it. This addition of *near vicinity* is based on the fact that as we live in a normally looking world, all theories which contradicts my current experiences should be false. However, not-near events may be more and more strange. For example, black holes are possible, but just not near my planet.

We are already in a very strange position. Most beings are animals, and even most humans are not interested in quantum immortality. So, is not it strange that we - the author and a few possible readers - are interested in QI? But it could be easily answered by conditioning on those who are capable to ask the question. Only QI-aware beings are counted, so it is nothing surprising to be one of them – it is trivial, even tautological.

# Appendix 4. Next step of QI theory: Dust theory and Boltzmann brains

## Multiverse-immortality and Boltzmann brains

There is a dangerous thought – what if I just a Boltzmann brain? There is popular objection to this idea, that most of BBs have very random experiences, but I am not, so I am not a BB. However, it doesn’t work, for two reasons.

First, is similar to the reasons why QS experiment is not an evidence of MWI: that is, non-random BBs also exists and the fact that I am nonrandom one is not the evidence for that I am not BB. It is also similar to the Sleeping Beauty thought experiment, in which the Beauty is told that she is non-random mind and has to estimate the probability that she is BB. For one-theirder solution of the Sleeping beauty, she can’t update its estimation to be a BB based on non-randomness of her environment.

Another reason that randomness argument doesn’t work is that a BB is not able to make coherent conclusions about its own randomness. So, BB could think that he has non-random experiences, but still have random experiences, as most BBs are not logical.

So where is no way for BB to prove that he is not BB, and it seems to be bad as he will disappear after a microsecond. However, Multiverse immortality gives a solution. For any BB-observer-moment exists another next observer-moment in the real life, and if we agree with Multiverse-immortality, my mind will jump from BB to the real me. Loew presented similar theory (Loew 2017).

It is not hard to see that “real me” is not necessary state here, as the next observer-moment could be also BB (let’s call it BB2), if it has memories about last moment equal to BB1 and the continuous line of experiences. This idea is known as flux-universe and will be discussed in the next subsection. However Mueller showed that it is possible to create a stable solution with observable physical laws just assuming existence of observer-chains (Mueller 2017).

"What we can do, however, is recognize that it’s no way to go through life. The data that an observer just like us has access to includes not only our physical environment, but all of the (purported) memories and knowledge in our brains. In a randomly-fluctuating scenario, there’s no reason for this “knowledge” to have any correlation whatsoever with the world outside our immediate sensory reach. In particular, it’s overwhelmingly likely that everything we think we know about the laws of physics, and the cosmological model we have constructed that predicts we are likely to be random fluctuations, has randomly fluctuated into our heads. There is certainly no reason to trust that our knowledge is accurate, or that we have correctly deduced the predictions of this cosmological model.” <https://arxiv.org/pdf/1702.00850.pdf>

##  Multiverse-immortality and Dust-theory universe

The idea of dust universe comes from the novel of Greg Egan *Permutation city*. The idea is that if the physical world (including consciousness) is just a succession of states, then why would it matter that these states occur sequentially in time and at the same place. If any state of mind is just some information, that is sequences of numbers, the same sequences of numbers could appear as random sequences of atoms inside physical bodies.

The one difference of state of minds in dust theory with BBs is that they are not vacuum fluctuations but a random information sequences inside actually existing physical bodies. They could even include the same causal connections, as in thoughts in human mind, as atoms hit each other because of thermal movement, and these hits create different causal structures, which could be similar to causal stricture of neuronal exchange in mind or to calculation inside a computer.

In other words, complex random calculation in nature could be exactly the same as calculations which create human experiences, and some observer-moments could exist inside a stone in your garden.

This difference – quantum fluctuations create random minds or random thermal processes create random observer-moments – is not important, as the result is the same. What is really interesting in the dust theory, is the idea that close to each other observer-moments could produce subjectively continuous chain of experiences without having any causal connection with each other, which is exactly the same way as multiverse immortality works. Multiverse-immortality connects each state of consciousness with the next one in the dust theory.

If the dust theory is enough to explain the existence of my line of consciousness, it means that existence of any actual universe is not needed, or at least principally unknowable from the inside such line. Thus, we don’t know what kind of the “dusts speckles” creates needed combinations. However, pure “mathematical dust” is possible, if each state of consciousness is just a random number or is hidden inside Pi number.

This may have some observable consequences, if “primordial randomness” will able to break inside the stream of the subsequent states of minds. But this speculation is outside our topic.

## Dust theory and Flux-universe

A poster on LessWrong (<https://www.lesswrong.com/posts/is7ieoWyiyYRc7eXL/the-consequences-of-dust-theory>) expressed strong emotional concern about the consequence of the dust theory which he called *flux-universe*. Every night, he goes to sleep and his consciousness “simplifies”, that is he has much simpler observer-moments, during which he forgets most information about his life. Many other people also could have the same simple observer-moments, and, if dust theory is true, only current observer-moment defines next observer-moment – thus if he forgets all his life and everything he loves – he could “jump” into another observer-moment of another person and never return to his original self! This thought terrified him, as it was like death of his original personality and he was afraid to sleep.

I suggested several counterarguments:

1. If he “jumps” into another person after infinitely many jumps he will eventually return back and will not remember anything, so he is not losing his life forever.
2. It is better to think about the same situation from the point of view that he has many “futures”
3. *It also not a bug, but a feature,* as it creates possibility to perform some form of “*magic*”. If a person is in bad but very rare situation, for example, he had just lost his win in negative lottery, he could put his mind in very simple state, may be meditation or full concentrating on very common activity – so simple, that he completely forgets everything about bad luck event. In that case, much larger number of observers, who didn’t have this bad lack event, will also meditate in the same way, and he will jump into them and and then randomly become one of them. And this is exactly the same situation he feared about, but now used as instrument to change observed reality. (Surely, this magic, even if it works, fails altruistic test of updateless decision theory, as someone else will be on his place.) Also, if it works, much more complex magic is possible, in which I consciously manipulate my uncertainties about some aspect of the observed world in order to get some control over it. In a long run, it seems to be zero sum game, as simple observer states will be exploitable limited resource by those who want to escape bad life-lines.
4. Another reason, why we should not be afraid that dust theory throws us in a kaleidoscopic *flux universe* is that if lines of similar observer-moments exist in the space of all possible minds, such lines will probably come to some “strange attractors”, where some form of stabilization mechanism exists. Stabilization mechanism means that observer-moments rather coherently follows one another. There are several candidates for such mechanism: a) nature of human dreams means that we never have really simple *blank* observer moments, we start to have some form of random thoughts and images (hypnogogia) as soon as we become sleepy, and they are very individual b) human observer-moments consist of qualia as elements and qualia could be very individual: I can’t jump into a person with different red color if I see now my red color. c) we constantly have perceptive memory of last moment. d) even more weird stabilization mechanisms are possible (like Man in Black), or *counterglitches*, which glitch in the way to make the world more coherent. These stabilization mechanisms could fight expected randomness of Boltzmann brains (discussed above) – but not completely, and some strange things like *glitches in the matrix* (but there are no matrix) could be observed.

One reason why all this is related to the Multiverse-immortality is that dust theory is about constant jumping between “worlds” and copies, and in the Multiverse-immortality it happens only in the moment of death.

## Miracles and glitches in the observer-moments chains

(This section does not belong to this article per se, but I leave it here for a while. I do not believe in the following, but it has high enough a priory probability to be worse be discussed)

If we suggest that human life is just as a line of similar observer-moments, like 4567, 4568, 4569, it has the following consequences:

1. For any random observer-moment there is another observer-moment, which follows from it, and makes it less random. If you see static, the next observer-moment will be just you looking on TV screen.
2. If from A observer moment follows B moment, and from B follows C moment, it doesn’t mean that from A follows C, as only near observer-moments must be similar, but A may be not coherent to C. So, observer-moments have not transitive relation. The consequences of it is fluidity of the world, when we compare A and C – we could find some disagreements.
3. For observer-moments chains there is no time, and past observer-moments could be somehow similar to the future moments, which could take form of “precognition”
4. Here should be some connection with quantum theory, that is any observer-moment may result from infinitely many different paths of past global wave function, or in other words, any observer-moment is surrounded by uncollapsed world. But is seems to be different level of description of the same phenomena.

# Blog post and comments about the decline of measure and its compensation by the merging timelines

<https://www.lesswrong.com/posts/5TYnquzAQPENyZXDa/quantum-immortality-is-decline-of-measure-compensated-by>

I wrote an article about the quantum immortality which, I know, is a controversial topic, and I would like to get comments on it. The interesting twist, suggested in the article, is the idea of measure increase which could compensate declining measure in quantum immortality. (There are other topics in the article, like the history of QM, its relation to the multiverse immortality, the utility of cryonics, impossibility of euthanasia and the relation of QI to different decision theories.)

The standard argument against quantum immortality in MWI runs as following. One should calculate the expected utility by multiplying the expected gain on the measure of existence (roughly equal to the one's share of the world’s timelines). In that case, if someone expects to win 10.000 USD in the Quantum suicide lottery with 0.01 chance of survival, her actual expected utility is 100 USD (ignoring negutility of death). So, the rule of thumb is that the measure declines very quickly after series of quantum suicide experiments, and thus this improbable timeline should be ignored. The following equation could be used for U(total) = mU, where m is measure and U is expected win in the lottery.

However, if everything possible exists in the multiverse, there are many my pseudo-copies, which differ from me in a few bits, for example, they have a different phone number or different random child memory. The difference is small but just enough for not regard them as my copies.

Imagine that this different child memory is 1kb (if compressed) size. Now, one morning both me and all my pseudo-copies forget this memory, and all we become exactly the same copies. In some sense, our timelines merged. This could be interpreted as a jump in my measure, which will as high as 2power1024 = (roughly) 10E300. If I use the equation U(total) = mU I can get an extreme jump of my utility. For example, I have 100 USD and now my measure increased trillion of trillion of times, I supposedly get the same utility as if I become mega-multi-trillioner.

As a result of this absurd conclusion, I can spend the evening hitting my head with a stone and thus losing more and more memories, and getting higher and higher measure, which is obviously absurd behaviour for a human being - but could be a failure mode for an AI, which uses the equation to calculate the expected utility.

In case of the Quantum suicide experiment, I can add to the bomb, which kills me with 0.5 probability, also a laser, which kills just one neuron in my brain (if I survive), which - let's assume it - is equal to forgetting 1 bit of information. In that case, QS reduces my measure in half, but forgetting one bit increases it in half. Obviously, if I play the game for too long, I will damage my brain by the laser, but anyway, brain cells are dying so often in aging brain (millions a day), that it will be completely non-observable.

BTW, Pereira suggested the similar idea as an anthropic argument against existence of any superintelligence <https://arxiv.org/abs/1705.03078>

For the sake of argument, we assume that the information is erased sufficiently good for two pseudo-copies to be become exact copies - and this information is not immediately restored by interaction with the outside world. If we were digital minds, it will be simpler: I can generate a random string of data in advance and then delete bit by bit.

Even if the cut-off is not sharp, something – which was completely not me – may become partly me after my simplification. Adding diffuse personality border is a correct step to the proper calculations of "me" and "not-me", if we ever come to this level, but this doesn't change the idea of the post: The simplification of the definition of "me" results in wider "bell curve" and thus larger share of all possible observer which are me.

**My comments:**

## Comment 1. Three solutions of the conundrum: utility calculation with QI, measure manipulation magic or open individualism

You right - I didn't mean the merge of quantum worlds, but mean the merge of personhoods, which could happen even in classical but infinitely large universe. I wanted to show that assuming the possibility of such merger has absurd consequences. However, this absurdity is also applicable to other calculations where the "measure of an observer" changes, as it happens in the one of important objections to the quantum immortality.

In the other words, we can't kill the quantum immortality idea by saying "the measure will decline to infinitely small values", as in fact, the measure could even grow if we properly calculate it.

What could be done to resolve this conundrum? We could ignore changes of absolute measure, and look only on relative measure (thus assuming that QI is true), that is relation between shares of the different outcomes where observer is alive. For example, if QS thought experiment has 3 outcomes: a) non-existence 0.9, b) winning 1000 usd with 0.09 с) losing 100 000 with 0.01 probability (e.g. injury) , we should in that case ignore (a) and compare expected utility of (b) and (c), which are +90 and -1000, so the game in this case has negative utility of -910.

Another solution is to accept that we could change our measure in the world by forgetting things, and to build something like a magic based on it. (This idea was discussed on LW as "flux universe" on a series of posts where a person had panic attacks based on idea that if he forgets parts of his personality before sleep, he would be never able to return back to his initial self.) This "magic" may look like: (1) a person learns that he has rare deadly disease. (2) he meditates and remove from his mind all clues about this fact (3) his observer-moment becomes so simple that it is equal to zillions of observer-moments of other people (4) if human mind is only numbers, they "merge", (5) now he returns to awake state, but his probability to be in the world-line where he has the rare disease is equal to the level of incidents of the disease in the population, that is very low. (6) Profit. But this thing seems more absurd than quantum immortality.

Another possible solution is to get rid of the idea of identity in favor of some form of open individualism. The problem here is that most human preferences are formulated in the way that they assume existence of some form of identity: "I want a cake".

## Comment 2: any definition of identity creates possibility of "pseudo-copies”.

Most human preferences have an embedded idea of identity as a receiver of the profit. However, the idea of "beings similar to me" assumes that there are "beings which are not enough similar to me for to be regarded as me" - but still have some of my traits. In other words, any definition of identity creates possibility of "pseudo-copies": if we define the identity wider, the circle of the pseudo-copies around it will become also wider, but will not disappear until we include all possible beings and end up with open individualism.

If we assume total "open individualism", it results in perfect effective altruism and the utility function will be akin "I prefer that total wellbeing of all sentient beings in the universe will increase on 100 pounds". However, this is not how most human preferences work, and there is also a risk of starvation.

So, playing with the definition of identity will not help to escape the problem of existence of pseudo-copies, which could become "real me", if some information is erased from both of us.