

Civilization-Level Quantum Suicide

By Paul Almond, 2 June 2008

Website: www.paul-almond.com
Email: info@paul-almond.com

© Copyright Paul Almond, 2008. All Rights Reserved.

Civilization-Level Quantum Suicide

By Paul Almond, 2 June 2008.

Abstract

If you think quantum suicide is valid then you should expect an advanced civilization to think that *quantum suicide at the level of an entire civilization* is valid, and probably to be more certain about it, as you should expect an advanced civilization to know anything that you know. If you also think the technological singularity idea is correct then you should expect a civilization to start performing civilization-level quantum suicide around the time it undergoes a technological singularity. Motives for quantum suicide could be *quantum suicide reality editing* – using quantum suicide to enter desirable situations – and *quantum suicide computing* – using quantum suicide to gain huge computing capability. This provides possible answers for the Fermi paradox, the Doomsday argument and the simulation hypothesis. A civilization might not engage in quantum suicide forever: it may perform it for a short time until its motivation becomes even more abstract and even stranger behaviour takes over.

This article is not arguing for or against the validity of quantum suicide, but merely considering implications of quantum suicide being valid.

Introduction

Quantum suicide is a controversial idea proposed by Max Tegmark and based on the *many-worlds interpretation of quantum mechanics* (MWI) [1]. MWI deals with the apparent conflict between quantum mechanics and classical physics which manifests itself in experiments like the double-slit experiment by suggesting that when a quantum event occurs all possible outcomes occur as a result of the world splitting into different worlds for different outcomes. In MWI time is not like a line, but like a tree. Quantum suicide is based on the idea that if MWI is correct and you set up an experiment in which you are immediately killed if a quantum event occurs with a particular outcome, but not killed if it occurs with another outcome, then there will always be branches in which you survive. The really controversial part of the quantum suicide idea is that you should only view those branches in which you can make observations as being possibilities for your future and so, from your point of view, should be certain of survival.

This article will not be arguing for or against the validity of MWI or quantum suicide.

What I Mean By “Civilization”

When I refer to *civilization-level* quantum suicide I am not necessarily talking about something that most people would recognize as a civilization. I am talking about whatever the sort of thing that we regard as a civilization turns into in the future. It could be a civilization of organic beings like ours, or it could be a civilization of organic beings

augmented by technology so that the boundaries between individuals are reduced in some way. It could be a civilization in which mind uploading [2,3,4] has been used to make the transition from organic brains to artificial intelligence, or one in which artificial intelligence has replaced organic brains in some other way. It could be a civilization in which all minds have been amalgamated to form a single mind, or one in which minds are split and amalgamated as needed.

Relevant Philosophical Ideas

Some existing philosophical ideas are relevant to the argument that I will make. These are the many-worlds interpretation of quantum mechanics, quantum suicide and the technological singularity.

The Many-Worlds Interpretation of Quantum Mechanics

Also known as the relative state formulation of quantum mechanics, the many-worlds interpretation (MWI) [1] is Hugh Everett's proposal that the wavefunctions in quantum mechanics should be regarded as a true representation of reality, rather than as abstractions that describe the probability of finding particles in particular locations, that when we experience one of many possible outcomes happening all possible outcomes actually occur and that "decoherence" of quantum wavefunctions prevents the wavefunctions associated with different outcomes from interacting with each other, effectively producing different "worlds" in which an each observer will only experience one outcome.

In some experiments, where quantum physics is an issue, weird experimental results are obtained. For example, in the double slit experiment light is passed through two slits and interference effects are observed consistent with light from one slit interfering with light from the other. The problem with this is that light is understood in terms of both wave and particle (photon) models, an idea known as wave-particle duality, and when the intensity of the light is reduced so that only a single photon should be passing through the apparatus at once the interference effects still persist, even though the single photon should pass through one slit or the other and not both. Some interpretations of quantum mechanics explain this as "wavefunction collapse" which is supposed to occur somehow as the effects of a quantum event become magnified and macroscopic. MWI explains it differently as wavefunction interference between different possibilities occurring before wavefunction decoherence has occurred.

If MWI is correct then, every time a quantum event occurs, the world splits into different worlds in which all the different outcomes occur and history is more like a branching tree than a single timeline.

Quantum Suicide

Quantum suicide is an idea related to MWI and proposed by Max Tegmark. The idea is that you arrange a mechanism to destroy you when a quantum event occurs with certain

results. According to MWI all outcomes occur in different branches and you will never experience those branches in which the quantum event occurs so as to activate your mechanism and destroy you, so you should not regard those as your future. You will only observe those branches in which the mechanism is not activated, so as far as you are concerned your survival is guaranteed because there will always be branches in which you survive and you can ignore the branches in which you do not.

This is a controversial idea.

There is an extension of the quantum suicide concept known as *quantum immortality*. The idea is that in any situation where you are facing death there would always be some branches where you somehow survive and you will only ever observe these branches. Quantum immortality therefore asserts that conscious beings are immortal.

This is also a controversial idea.

Not everyone who accepts MWI also accepts quantum suicide or quantum immortality and some people who accept quantum suicide in certain circumstances reject quantum immortality. Some people who accept MWI assert that both of these are “fringe” ideas, not taken seriously by most people who accept MWI.

The Technological Singularity

The technological singularity [5] is a hypothetical point in a civilization at which technology leads to the creation of self-improving intelligence and/or at which hyper-rapid technological progress starts to occur. The term singularity was proposed by Vernor Vinge [6] to imply that, if this happens, we will not be able to predict what will happen next. A singularity is often regarded as likely to result in ultra-intelligent machines [7].

Is quantum suicide valid?

There is the issue of whether quantum suicide is a valid idea. It would be an *absolutely* valid idea if continuity of self were guaranteed, but whether or not continuity of self really objectively exists is debatable. It may be that continuity of self is an illusion and what we think of as continuation of us from moment to the next is an illusion. Rather than get into this issue of whether quantum suicide is valid in some absolute sense, it is better to consider asking about validity as really asking:

Would an advanced civilization take quantum suicide seriously enough for it to affect its behaviour?

or

Would an advanced civilization regard quantum suicide as being free of any existential cost?

If you think that quantum suicide is valid then you should expect an advanced civilization to agree with you: you can hardly expect it not to know something that you know. Alternatively, if you think that MWI and/or quantum suicide are nonsense then you should expect an advanced civilization to know this too.

Motives for Quantum Suicide

Quantum Suicide Reality Editing

If you accept the idea of quantum suicide then you should be open to the idea of using it for editing reality. You could construct some system that monitored events for you and would immediately cause you to cease to exist if events did not happen as you wanted them. The idea would be that you would continue to exist only in those future worlds in which events happened as desired, so that from your point of view, events would always happen as you wanted. You would be using quantum suicide to control your reality.

Quantum Suicide Computing

Basic Quantum Suicide Computing

A special case of quantum suicide reality editing would be what I will call *quantum suicide computing* – not to be confused with quantum computing.

Suppose you had some computing problem which would take a long time to solve, but you have some way of checking possible answers. You could set up some system which uses quantum events to generate a random answer to the computation and then automatically causes you to cease to exist if the answer is not the correct answer, or if it is not better, in some sense, than the previous answer that you obtained. The idea would be that future worlds would exist in which all possible answers were generated and you would only exist in those worlds where the answer was correct or better than previously generated hours, thereby giving you the perception of having enormous computing power.

Why would you would want to do this in preference to building a "conventional" quantum computer that does not require you to kill yourself as much to compute things? One reason could be that there are practical problems with building computers that will impose limitations on them, and it may be that, if you regard quantum suicide as reasonable you should expect this kind of method to be less limited. If you really think quantum suicide is viable you should think that this computing method is *free* in terms of existential cost.

Quantum Suicide Thinking

A special case of quantum suicide computing is what I will call *quantum suicide thinking*. This is when the actual computing being facilitated by quantum suicide is your own thought processes, so that quantum suicide is used to increase your cognitive capabilities.

Implications

Suppose that MWI or some similar idea is correct and also that quantum suicide is an idea that an advanced civilization would accept as not involving existential cost.

It follows that an advanced civilization would perform quantum suicide reality editing and quantum suicide computation, and possibly quantum suicide thinking, on a civilization-wide scale. The *entire civilization* would set up mechanisms to cause it to cease to exist when unwanted events happen, or when guessed results in computations are not the answers that were wanted, etc.

If you think quantum suicide is an absurd idea then you should hardly expect an advanced civilization to know less about philosophy than you do, so you should expect them to reject it as well. If, however, you think that quantum suicide is a reasonable idea then you should not expect an advanced civilization not to know something that you have worked out. Further, you should not expect them to have many doubts or intuitively based fears over it. They would have more cognitive resources than you to direct at the issue. They should either *know* that quantum suicide is nonsense or they should *know* that it makes sense.

If they do not think that there is any existential cost in performing quantum suicide once then they may not think there is any cost in doing it many times. They may therefore destroy themselves as often as possible as a way of performing quantum suicide reality editing, quantum suicide computing and possibly quantum suicide thinking. Deliberately engineered, civilization-wide extinction events could be ultra-frequent. If we ignore the complications of how large the civilization is and how long it takes information to be transmitted from one side of it to the other, we might even expect the civilization to destroy itself many times a second. Quantum suicide might even become the dominant means by which the civilization affects reality.

Now, suppose that a technological singularity occurs. If you think there is any validity in the quantum suicide idea you should expect a civilization that is undergoing the increase in thinking capabilities associated with a singularity to realize this immediately, and to realize it without significant uncertainty.

This means that, if you believe quantum suicide is reasonable and you believe in the technological singularity hypothesis then you should expect that *the first of many quantum suicide events may occur around the time of a technological singularity*.

We might imagine a variety of methods by which a civilization might perform an act of quantum suicide. Many advocates of quantum suicide think that it would need to be done very quickly, so that you could never actually observe that you were on the wrong branch. One obvious way of doing this might be to use nuclear weapons. A quieter method might be imagined for a civilization based on artificial intelligence, for example one in which mind uploading [2,3,4] has caused machine intelligence to supersede organic brains. If the civilization's thought occurs in a computer network instead of in

organic brains then the easiest, and quickest, way of performing an act of quantum suicide might be to turn the computers off, or erase the software for them, or use some software method to render them non-functional. This software kind of quantum suicide might address the concern that quantum suicide might only be partially successful and may lead to many branches in which you survive with significant damage. Computer systems running artificial intelligences could be designed so that, once some quantum suicide process is started, it will almost certainly end all functioning of the computer or fail completely, so that there is little chance of any middle ground of impaired functionality.

Deferred Quantum Suicide

Some criticisms of quantum suicide reality editing would be based on the idea that, by the time it is performed, reality has already got into a state such that all of the branches for any conventional future are ones in which (for example) some event which we do not want to happen is going to happen anyway.

Some people have suggested that quantum suicide could fail in practical situations, because there may not always be branches in which you survive. If you set up a situation involving a radioactive source so that you will be killed if a single particle is detected within some time and not killed if it is not then it is easy to see that, if MWI is true, there will be some branches in which you survive. It may be harder to see how there are supposed always to be branches in which you survive in situations where reality is in a state that is prejudiced against your survival in ways that would need something more than the occurrence or otherwise of a microscopic particle detection event to put right.

As an example, imagine that a train you is about to hit you in a few seconds. Can you really expect a branch in which this does not happen? You could argue that there could be a branch in which some large obstacle gets in front of the train and stops it, but that would probably already need to be starting. You could argue that someone could push you out of the way, but if there is nobody around how are microscopic quantum events supposed to get someone in a few seconds? It would seem that getting saved in any conventional way, when you are very close to death, would require you already to be living in a world in which things have already started happening to save you. Likewise, if you intend to use quantum suicide for reality editing, how do you know that there are any branches in which your desired reality emerges, and what if you have left it so late that a reality that you do not want is the only possibility?

One answer is to say that there will always be *very* extreme things that could happen to bring about the situation that you want. This same reply can be made to people who use a similar argument against quantum immortality: that there will be some situations in which there are no branches in which you survive. Even when a train is about to hit you, the way that the matter is behaving in you, the train, and the surrounding environment are all a product of quantum events. While it may not happen in many branches, there will always be some unlikely sequence of events that changes the world in a radical way. This is similar to the sort of speculation that suggests that you could asphyxiate if all the air

molecules in your room just happened to go to the same side of the room randomly, and it would be suggesting that some amazing event is supposed to happen to improve your situation

There is a problem with this, however. Even if we accept that extreme events can always happen, such events may be of such low probability that it could be very hard to predict what your world is like afterwards, potentially making this less attractive as a form of reality editing.

A possible solution is what I will call *deferred quantum suicide*. This would allow you to start the process of quantum suicide long before it is obvious whether or not you are on a desirable branch, but to defer making it permanent until it is known how desirable a particular branch is. In this way, providing that you start soon enough, you can rely on the outcomes of quantum events, or even a single quantum event, generating the branch you want. Deferred quantum suicide would require you to be able to suspend your thinking ability. Here is how it would work:

Suppose you want the world to meet some criteria in two weeks, and let us suppose that this is long enough in the future that you expect a future like this to be accessible from your world, down some sequence of branches involving individual quantum events, in conventional kinds of futures.

Assuming that your mind is running on a computer, you shut the program down, but you first set an automatic mechanism to restart you if, and only if, reality meets the required criteria in two weeks. If it does not meet that criteria then you will not be restarted and will “sleep” forever: you could actually arrange for automatic destruction of the computer system or its software at this point to ensure that you do not accidentally wake up.

After setting the system up you will be shut down and will know nothing more, until or unless you are restarted later.

If MWI is right, and provided that you have chosen some future that is accessible by a sequence of quantum decisions from your own world, then in some branches the desired future will occur, and the computer system running your mind will be restarted in these worlds. In the other worlds, where the desired future does not occur, you will never be restarted and will never know.

Deferred quantum suicide may initially appear no better than “conventional” quantum suicide, but it has the advantage of removing the need for reliance on strange events happening at the time that you do it. Instead, you can shut yourself down and rely on the butterfly effect (the sensitivity of chaotic systems to initial conditions) to generate many different worlds, most of which will seem quite conventional, only being restarted and becoming conscious again in those worlds that the automatic monitoring system finds acceptable.

With deferred quantum suicide you might just start the automatic monitoring system

running and shut yourself down, relying on quantum events naturally generating different futures, or you might set some device up which responds to quantum events to make different, (quantum) random changes to the world to generate considerably different futures. Another option would be to set up automatic systems to make different, controlled changes to the world, in different branches, with your automatic monitoring system determining whether or not you should be woken in each case.

We can see how this might work by considering the example of a lottery. Suppose that the lottery does not involve winning numbers being drawn by a machine after you buy a ticket. Instead, the winning numbers have been drawn *before* you buy your ticket and are written on a piece of paper in a locked safe. Announcing the winning numbers will involve the safe being opened and the numbers being read from the paper. If those numbers match the ones on the ticket you have bought then you win.

Suppose also that you exist as computer software: your mind is running on a computer.

This kind of lottery will not be influenced by any likely sequence of quantum events after you buy your ticket. The numbers have already been drawn and are merely being kept secret. In almost all worlds the numbers that win the lottery will be the same – those that are in the safe. There will be some worlds in which ultra-unlikely things happen to change this, such as the piece of paper in the safe randomly changing, but hoping to exist in these Dr Seuss worlds is hardly a good idea.

We can improve this, however, by using the following approach:

You shut down the computer that is running you. Before that, however, you set up an automatic mechanism to wait a short time and then choose a set of lottery numbers based on a series of quantum events and buy the lottery ticket on your behalf. The automatic mechanism will also restart the computer running you in the future, but only if you win the lottery. When the lottery is drawn, if your ticket wins, the computer that is running you will be restarted and you can collect your money. If you do not win the lottery then the computer will not be restarted again, and maybe some mechanism will even destroy it, or wipe you from it, to ensure that it never gets restarted due to accident or malicious action.

The idea is that, before shutting yourself down, you know that branches will be available for every possible lottery ticket purchase and you will only experience those futures in which you win. This is a *deferred* act of quantum suicide. We are assuming that all the branches will be undesirable and starting quantum suicide immediately, but it is only after the lottery result is known that the decision is made about whether or not to make that quantum suicide *permanent*. If you win the lottery then the act of quantum suicide is cancelled, or reversed, by restarting the computer.

The objection that desirable futures may not have a chance to happen in any predictable way is no longer valid, because we are not expecting some ultra-unlikely sequence of events to allow us win the lottery – merely the future in which the mechanism happens to

choose the correct numbers, which is hardly *ultra*-unlikely: it happens to people all the time. Although the mechanism of this will probably be clear, what really makes this work might be harder to see. The most important feature is not that your ticket purchase is based on quantum events: any ticket purchase is based on quantum events. It is the fact that the time delay between the initial act of quantum suicide and the decision about whether or not to make it permanent allows time for small changes to be magnified enough to generate radically different possible futures – some of which will be desirable to you. Leaving aside insane sequences of quantum events that twist the shape of the world, once you have bought a ticket and the numbers are on the paper it is too late: things have gone too far and reality has already committed itself to a final outcome. On the other hand, before you buy the lottery ticket is the time when reasonably likely, desirable branches still exist. The world is not yet committed to whether or not you win the lottery and the outcome can be changed, in a reasonable way, by a small number of quantum events - those that determine what numbers you choose.

This sort of approach seems to allow quantum suicide reality editing without having to enter strange realities where ultra-low probability events happen.

The lottery example here was a bit contrived. We are discussing the quantum suicide of an entire civilization and, even if quantum suicide is valid, it is unlikely that civilization-level quantum suicide would happen for something as trivial as a lottery.

Civilization-level Quantum Suicide and Measure

While not essential to an understanding of the main ideas in this article, my previous article *Minds, Substrate, Measure and Value, Part 3: The Problem of Arbitrariness of Interpretation* [8] is relevant. It proposed that minds should be associated with any valid, formally expressed interpretations of physical reality, rather than associated with some ad hoc, intuitive, explicitly computational interpretation. It was suggested that, if MWI is correct, the continual “thin-slicing” of quantum wavefunctions as worlds split means that, all else being equal, the measure of a mind in a single branch will tend to decrease as time passes.

In the absence of any events likely to end your existence in any branches this should not make much difference. For example, if your measure is M before some event and after that event 10 branches split off then it would seem reasonable to say that your measure is $0.1M$ in each branch, but your total measure remains the same.

With quantum suicide occurring, however, things are different: it would eliminate you in some branches and reduce measure. Suppose, for example, that your initial measure is M and after some event there are 10 branches, only one of which involves your survival, then it would seem reasonable to say that your measure in that branch is $0.1M$ and this is now your total measure: 90% of your measure was “lost” in the event.

This would mean that if repeated common suicide occurs then observer moments will become progressively rarer. Even if you view your future is extending infinitely into the

future, most of the observer moments would actually be contained in some finite part of your history.

We can get an idea of this by considering infinite sequences of numbers. For example:

$$1 + 0.5 + 0.25 + 0.125 + 0.0625 + 0.03125 + \dots \text{ etc}$$

This sequence goes on forever and adds up to 2. Although it goes on forever, half of the total comes from the first number, 1, because the other numbers are decreasing exponentially. Similarly, even if observer moments go on forever in time, if they keep decreasing in number then earlier observer moments, before the decrease starts, could disproportionately contribute to the total. A significant fraction of observer moments could be those occurring before you started performing quantum suicide.

The explanation just given was in the context of the many-interpretations view described in my earlier article [8], yet some people would say that a conventional understanding of quantum mechanics gives the same result and that splitting of worlds should be regarded as decreasing numbers of observer moments. I will not go into detail on this here, instead assuming that many-interpretations is the correct approach, as I think that is the case.

Possible Problems

Are there always survival-branches?

In objecting to quantum immortality, which is not the main subject of this article and not something that I will try to defend, the objection is often made that you could find yourself in situations where none of the branches in your future involve your survival. This sort of situation is still relevant when considering quantum suicide reality editing because now the intention is to make your survival conditional on particular events happening or not happening, so the question of whether or not survival branches exist becomes the question of whether or not branches exist in which reality is as you want it. The problem, mentioned previously, is that by the time some automated system makes the decision about whether to commit you to quantum suicide or not, reality may have already got into such a state that no conventional sequence of events is enough to cause a desirable event to happen or prevent an undesirable event.

One answer is that, no matter how unlikely it is, some sequence of quantum events could always occur to bring about the situation that you want, but the problem is that, even if this is valid, in some situations this will be so ultra-unlikely that things would be unpredictable. Relying on such a justification for quantum suicide would be like pressing the “Hyperspace” button in old videogames like “Asteroids” or “Defender”.

Another way in which you might get the reality you want is by “unconventional continuation”. A “conventional” view of quantum suicide would be to look for some branch in the future of your own world in which you survive, but some ideas about continuity and mind are more flexible. As an example, some people regard the idea of

mind uploading [2,3,4] as suggesting that a digital copy of your mind could be said to be you, even though it comes into existence somewhere other than your own brain. If you think that a copy of you would be adequate for continuation then you do not need to rely on very low probability events in the future of your world: you can rely on the probable existence of someone in another world who has the correct mental state to be regarded as a copy of you, with a different past that has allowed survival.

Let us consider the example of a train about to hit you, and let us presume that you could be saved if someone jumps up from hiding in bushes nearby and pushes you out of the way. In most worlds in which you could exist, you should expect that there is nobody in the bushes to jump out, so no plausible sequence of quantum events is likely to make it happen – although you might still expect branches where freak sequences of quantum events could still ensure your survival in crazy ways: let us ignore those very low probability sequences for now and assume (possibly simplistically) that no branches from your own world result in your own survival. If MWI is correct, however, there is sure to be some other world in which someone almost exactly like you was in the same kind of situation, except someone was hiding in the bushes and does come out to push you away in the last second, and if you would accept the sort of continuation involve in mind-uploading then you should possibly expect this as a continuation of you, even though it has no direct causal connection with your own world. You might also imagine strange ways for such unconventional continuity to arise, for example, if you die in one world then, if MWI is true, some of the vast number of other worlds with different pasts would be likely to be worlds in which you had even be tricked into thinking that a train was even approaching you. This prospect of unconventional continuation seems to be overlooked in many discussions of quantum suicide and quantum immortality, which is surprising as many enthusiasts of this sort of subject are also interested in concepts like trans-humanism and mind uploading.

If you have an artificial intelligence based civilization, however, the best answer to this problem may be one that has been described already: deferred quantum suicide. By suspending your thought processes soon enough, while many branches still exist in your future where the desired outcome happens in conventional ways, and using an automatic mechanism to restart you if reality is desirable, almost all of the futures in which you experiencing anything would be desirable and quite conventional.

The Problem of Imperfect Quantum Suicide

In a lecture, *How Many Lives has Schrodinger's Cat?* [9], the philosopher David Lewis stated that we should hope that the many-worlds interpretation is false, because if quantum immortality occurs then there will be few branches in which you survive unscathed – and many more in which you survive maimed. Any civilization embarking on quantum suicide would need to consider this. Whatever means of suicide was used, the chances of complete destruction would need to be large enough, compared to the chance of some kind of permanent damage or loss of capability, to make it worthwhile.

I will not rule out the idea that, even if quantum suicide is valid, this sort of prospect

could actually cause a civilization to limit the extent to which it engages in quantum suicide. Against the idea I would make the following comments:

The idea, if valid, seems more relevant to considerations of quantum immortality rather than quantum suicide. With quantum immortality you are depending, for conventional continuation, on quantum events causing some ultra-unlikely sequence of events to save your life, or at least keep you conscious. Even then, however, there is the issue of unconventional continuation. If you think that the existence of someone with a mind in an appropriate state in a world which is not in one of the branches in your future would be a valid continuation then you should expect most such observers to be in comparatively normal situations (using the term in a decidedly relative sense), rather than to be maimed but kept implausibly conscious in some way. As an example, which of these is likely to be most common – a world which appears almost like ours, or maybe a more different world in which someone like you exists whose life is being maintained by fairly mundane means – or one in which you are being kept alive due to an ongoing sequence of freak quantum events, which should end at any instant as sensible statistics takes over and normal service resumes, but refuses to do so?

This article, however, is about civilization-level quantum suicide rather than quantum immortality and a civilization performing quantum suicide would be able to use deferred quantum suicide to ensure to select outcomes with reasonable probabilities of occurring. “Reasonable” is not the same as “likely” here. Some “reasonable” probabilities for “success” in deferred quantum suicide could be for outcomes that most people would regard as unlikely. With the lottery example, the probability of winning might be about 1 in 14 million (depending on the rules of the lottery), so you would continue to exist in about 1 in 14 million branches. Although there might be some branches in which the quantum suicide mechanism malfunctions and leaves you permanently damaged, or in which you win the lottery (or in which the quantum suicide mechanism otherwise fails to activate) as a result of some weird sequence of unconventional events that leave you existing in some unpredicted, miserable state, branches with events like this could occur in the many-worlds interpretation even if you are not performing quantum suicide and you would merely need to ensure that the probability of events like this happening is much less than 1 in 14 million – for example, if your civilization is based on AI systems, ensuring reliability in the systems and the quantum suicide mechanism, to make branches like this much less common than those for the “conventional” sort of future in which you simply win the lottery.

The Problem of Civilization Fragmentation

Quantum suicide is usually considered in the context of a single mind. Indeed, it is frequently pointed out that anyone performing it cannot expect to prove anything to doubters: most of the future branches for onlookers would be ones in which they watch someone committing suicide. If we are considering quantum suicide involving an entire civilization this issue becomes relevant. What if the entire civilization performs an act of quantum suicide and then individual minds in that civilization are continued in different branches, so that the civilization as a whole is not continued in any significant number of

branches?

This issue is addressed by the use of deferred quantum suicide. Rather than depending on strange, unpredictable events, the civilization can plan for relatively conventional events to occur in which the entire civilization continues to exist. With the lottery example given, the future with the greatest number of branches, in which any minds continue to exist at all, would simply be the one in which the civilization wins the lottery, provided that it has been made very unlikely (compared to winning the lottery) for individual minds to end up existing in isolation.

What if they get the calculations wrong?

Deferred quantum suicide might answer objections about the rarity of branches corresponding to desirable futures, but what if a civilization performing a deferred quantum suicide gets its calculations wrong, or acts on flawed information, so that there really are no branches in which the events needed to cause the automatic mechanism to restart the civilization can occur conventionally?

Using the lottery example again:

Suppose a civilization uses deferred quantum suicide to try to win a lottery. The civilization is running as AI software on a computer system and stops itself running. An automatic mechanism chooses a set of lottery numbers, based on quantum events, and buys a lottery ticket. The civilization is then restarted if, and only if, the civilization wins the lottery. Unknown to the civilization, however, the lottery has been rigged in a way prejudicial to the civilization, so that, whatever sequence of lottery numbers the civilization chooses, whoever is running the lottery will make sure that those numbers are not announced as the winning numbers. This would cause problems, because the automatic mechanism will only restart the civilization if it wins the lottery, but the civilization will not win the lottery in any conventional futures, and it will not be restarted in any conventional branches. The quantum suicide attempt has now backfired.

This does not necessarily mean that the civilization has committed *actual* suicide, if quantum suicide is valid at all as an idea. It could be argued that there will always be branches in which ultra-unlikely, unconventional events cause the civilization to be restarted, however this would be undesirable and would also raise issues of imperfect quantum suicide and civilization fragmentation.

A solution to this would be to perform deferred quantum suicide with a mechanism that has a low probability of restarting the civilization even if the desired situation is not achieved. In the lottery example, the mechanism could be set to have a low probability of restarting the civilization even without a lottery win. This probability could be set low enough so that almost of the branches in which the civilization gets restarted will be ones in which it has won the lottery, but high enough so that even if the civilization does not win the lottery (for example due to it being rigged) branches in which the civilization is restarted, having lost the lottery, greatly outnumber branches in which the civilization is

restarted due to some unconventional, unpredictable sequence of quantum events.

The Problem of Marginalization

If the many-interpretations view which I discussed in a previous article [8] is correct then, if quantum suicide is being performed, later observer moments will be less common than earlier ones. If quantum suicide is performed to a great extent then later observer moments would be much more unlikely than early ones. If the sort of approach used in the Doomsday argument is correct – that you should find it surprising to be experiencing uncommon types of observer moments – then you should be surprised to find yourself in the situation of an observer who has committed quantum suicide many times.

This would be the case even if quantum suicide is valid. Even if you think that only those observer moments in which you survive constitute your future, it does not change the issue that those observer moments, once you are experiencing them, are ones that you should view as very unlikely.

This creates a potential problem with quantum suicide that I will call *marginalization*. Marginalization is when you experience observer moments that are so uncommon, due to an unusual history, that observer moments with similar kinds of experiences and more mundane histories are more common, creating the issue that maybe you should doubt that you are in the uncommon situation in which you seem to be.

As an example of marginalization, suppose that repeated quantum suicide puts you in the situation of experiencing an observer moment in which you remember performing quantum suicide many times. You now know that you have survived many such acts. However, you also know that you are experiencing an uncommon observer moment. Many more similar observer moments will be experienced by observers who are deluded and just think they have performed quantum suicide many times, so how do you know you are not one of those people? Does the rarity of the observer moment that you are supposed to be having not suggest that that is the case? We do not need delusion as an example. Suppose you think that at some point in human history simulations of all kinds of minds might be made. What if some of these simulations are of entities who *think* they have performed quantum suicide many times? Would such observer moments not greatly outnumber observer moments experienced by entities that have really survived quantum suicide many times? (See the later discussion of the simulation hypothesis, however.)

This suggests that, even if quantum suicide is valid, as you perform it many times you will become increasingly marginalized and, as you experience progressively less common observer moments, maybe you should doubt your own status.

There are a number of answers to this. A civilization performing repeated quantum suicide might somehow try to arrange things so that it can be more sure of its status. This might be possible by minds in the civilization increasing their capabilities after a singularity and also using quantum suicide computation to acquire massive thinking abilities. On the other hand, what about any deluded humans who might think they are

super-intelligent members of such a civilization? Should a mind in the civilization worry that it might really be in that sort of situation? This creates a problem of reference class and what you should regard as possible candidates for your situation, complicated by the fact that non-human, or post-human, minds with much more capability than ours could be involved, so I will not try to go into it further here. A civilization might deliberately compromise its own thinking ability, so that every time any mind in the civilization is about to doubt its own status the thought is stopped – an unusual situation for a civilization to arrange for itself – or it might compromise its memories (though some people would say that would make the value of any longevity questionable). A civilization might simply accept marginalization, and all the doubts that might go with it, as part of its future.

Issues Answered By Civilization-Level Quantum Suicide

The Fermi Paradox

The Fermi paradox [10] is the issue of why we do not have any observationist evidence of alien civilizations, given that the size and age of the galaxy is, in the belief of many people, more than adequate to have allowed many such civilizations to have come into existence by now.

Civilization-level quantum suicide, if it occurred, would resolve the Fermi paradox. The reason that we did not encounter other, more advanced civilizations would be that they have long since edited themselves out of almost all worlds, including ours. From their point of view they would not have edited themselves out of our world: they would have edited *our* world out of *their* futures.

If this were the case, how to interpret it is semantics. Some people would say that is a special case of the answer to the Fermi paradox which states that almost all civilizations eventually destroy themselves. If a civilization were about to perform an act of quantum suicide, an external observer, able to see the multiple paths branching off into the future, would see the civilization being destroyed in most of these branches, and as far as any practical statistics were concerned it would make sense to interpret this as a high probability of destruction. A civilization which has adopted the practice of civilization-level quantum suicide would accept this while simultaneously thinking that MWI and the concepts behind quantum suicide provide an escape from the existential cost.

One issue that might be raised here is that of anything left behind by the civilization. For example, what if a civilization releases self-replicating probes into the galaxy before embarking on civilization-level quantum suicide? A possible answer is that a technological singularity will almost always occur before things like this happen and that quantum suicide almost always starts to happen first.

The Doomsday Argument

The Doomsday argument [11,12], otherwise known as the Carter catastrophe, is an

argument originally proposed by Brandon Carter observing that if humanity is going to survive a very long time then our position in all of human history – from the start of humanity to its end – would be very unusual: we would be among the first humans ever to exist. Our status becomes more special when we consider exponential growth and the larger populations that will presumably exist in the future. If humans colonize space then the total population could become much larger than it is now and the total number of humans born after the 21st century, compared to all those who exist up to the 21st century, would be vast, making us a special, tiny minority of humans – the earliest ones born near the dawn of human history.

The Doomsday argument says that this is implausible and that it is much more likely that we will go extinct soon.

This is controversial.

There are many objections to the Doomsday argument; however it is also important to note that it can be stated in different ways: I will not be going into these here as the Doomsday argument is not the focus of this article. Some refutations of the Doomsday argument might refute a simplistic version of it, but fail to deal with more sophisticated versions.

If civilization-level quantum suicide were a normal practice for an advanced civilization then it could provide an answer to the Doomsday argument. A significant fraction of all observer moments could occur prior to the adoption of civilization-level quantum suicide and, after, civilization-level quantum suicide is adopted, frequent suicide events would reduce measure and numbers of observer moments, according to the many-interpretations view [8] and, possibly, the conventional view of quantum mechanics.

If the statistical method in the Doomsday argument is valid, if a technological singularity is expected soon, and if civilization-level quantum suicide is going to start happening then our position in history is where we should expect to be. We would be in a high population period of history, just before a singularity is going to usher in a future of repeated quantum suicides which will collapse the measure of future worlds and the numbers of future observer moments.

As with the Fermi paradox, this could be viewed externally as just a special case of our civilization having a high chance of destroying itself soon, and therefore as supporting the Doomsday argument. A civilization doing this would accept the statistical reality, but have a different subjective view of the existential cost.

The Simulation hypothesis

Nick Bostrom's controversial simulation hypothesis [13] has received media attention and some public interest. The argument considers the possibility that we are living in a computer simulation, not the sort of simulation in the film *The Matrix*, in which people's brains are plugged into a computer simulation, but one in which we do not even have

organic brains in the world outside the simulation, our minds being simulated by the computer system along with the world that we experience. If we lived inside such a simulation we would be artificial intelligences that just think they are organic. The simulation hypothesis is based on the idea that computer power will be greater in the future and, if we survive long enough, we can one day expect to have the computing power to run such simulated realities in which the simulated inhabitants have mental states and are unaware that they are simulated.

According to the simulation hypothesis one of these possibilities must apply (to quote Bostrom):

1. The fraction of human-level civilizations that reach a posthuman stage is very close to zero.
2. The fraction of posthuman civilizations that are interested in running ancestor-simulations is very close to zero.
3. The fraction of all people with our kind of experiences that are living in a simulation is very close to one.

Civilization-level quantum suicide could provide a resolution to this issue. If almost all civilizations that end up launching simulations start engaging in civilization-level quantum suicide before launching large numbers of simulations the measure of the branches containing these simulations would be greatly reduced, and according to the many-interpretations view [8] (and, according to some people's understanding of the conventional view of MWI, even a conventional view of MWI) so would the measure of any observers in them and the distribution of observer moments. The number of observer moments inside, for example, simulations of the 21st Century could be much less than the number of observer moments inside the real 21st Century, meaning that, even if the simulation hypothesis is correct, if you find yourself living in the 21st century you should still regard yourself as almost certainly living in the real 21st century. This situation would not be changed if the civilization doing the simulating went on making vast numbers of simulations over an arbitrarily long period of time, provided that they continued performing quantum suicide so that the measure continued to decrease. Issues could be created if the simulating civilization exponentially increased the number of operating simulations while they were performing civilization-level quantum suicide, so that any loss of measure in individual simulations would be compensated for by an increase in the number of simulations.

As with the Fermi paradox and the Doomsday argument, whether this is a possible alternative to the simulation hypothesis depends on your perspective on it. From an external, statistical perspective it is merely a special case of Bostrom's first option:

1. The fraction of human-level civilizations that reach a posthuman stage is very close to zero.

The civilization performing the quantum suicide might, however, see things differently from the point of view of existential cost.

Would the quantum suicide era be short?

If an advanced civilization viewed quantum suicide as a rational thing to do, and I am not claiming it would, then this would be an abstraction of the survival motivation. Motivation has become more abstracted in our past. Starting from systems that simply responded to what was happening in the present, that evaluated the desirability of situations based on the present, and going to systems that assessed situations based on what was going to happen in the future, we have evolved with abstracted views of “survival” that allow us even to accept replacement of body parts provided that we are still “us”. Our survival motivation has probably already reached at least the level of abstraction where we want to preserve our brains and some people seem to be taking abstraction further, thinking of survival as merely “continuing thinking” in some abstract way.

Quantum suicide, if regarded as valid by an advanced civilization, would be a further abstraction of survival motivation, and one which you should expect an advanced civilization to accept if you accept it. For how long would they do it?

Some ideas of a technological singularity suggest that it should be impossible to make predictions of what will happen after one, but there are different ideas of exactly what a singularity would mean. Even if you expect things to be completely unpredictable after a singularity, if you find quantum suicide valid you should still expect them to start at some point during or just before a singularity, provided that you think the civilization will have enough thinking ability to make a conclusive decision before the actual point of unpredictability is reached. A sequence of quantum suicides would then occur before and/or during a singularity until the singularity arrives and all predictions are irrelevant.

Quantum suicide might not go on forever then. We could imagine it might stop when the actual singularity is reached for a “brick wall of prediction” type of singularity. If we think we can meaningfully discuss what happens after singularities then quantum suicide might occur after a singularity, but it might also stop if the civilization changes its views.

Regardless of the semantics of the term “technological singularity”, then, quantum suicide should be expected at some point if you think quantum suicide is valid, but it might stop at some stage. It is unlikely to stop because the civilization decided that it was invalid after all: the sorts of thinking resources that civilizations like this could spend on the problem would easily be able to deal with a problem about which we can make an educated guess. What could happen is that, just as the step to quantum suicide is an abstracted motivation, a further abstraction of motivation could occur, possibly to something that we cannot imagine. We cannot even guarantee that the concept of “survival” would be recognized after such a step in abstraction. Quantum suicide would seem to many people to be a strange, abstract step for a civilization to take. Anything that replaced it would be even more alien to us.

Conclusion

This article has not argued that quantum suicide is a valid concept. It has been proposed, instead, that if you think that quantum suicide is a valid idea then you should expect an advanced civilization to know that too and to use it at the level of an entire civilization to control the situations in which it finds itself. An advanced civilization could use civilization-level quantum suicide as a form of reality editing and to gain extra computing power, which may be used to facilitate its own thought processes – though in an advanced civilization the distinction between thought processes and computing processes performed in tools may no longer be relevant. If you also accept the technological singularity idea then you should expect a civilization to start performing quantum suicide at the level of the entire civilization around the time that it undergoes a technological singularity.

Some problems of civilization-level quantum suicide could be dealt with by deferred quantum suicide, a type of quantum suicide in which the civilization stops its thought processes, restarting them later only if the situation is desirable. The advantage of deferring quantum suicide in this way, if the idea of quantum suicide is valid at all, is that it would allow the act of quantum suicide to be started soon enough, when there is still time for quantum events to produce the desired situation in conventional futures. Deferred quantum suicide could only be performed by a civilization which had enough control over its own thought processes to temporarily interrupt them, such as one based on artificial intelligence.

A further problem could be that the civilization might get things wrong and perform an act of quantum suicide so that there are no branches corresponding to conventional futures satisfying the criteria for the civilization to exist. This could be resolved by using a safer version of deferred quantum suicide in which, even if the desired situation does not arise, there is a chance of the civilization being restarted.

In an earlier article I have discussed the many-interpretations view and applying this to the many-worlds interpretation, on which quantum suicide is based, suggests that later many-worlds branches have fewer interpretations that produce observer moments than earlier ones, so that earlier branches will tend to be associated with more observer moments. Normally, this effect is compensated for by the proliferation of branches as time passes, but if quantum suicide is repeatedly occurring then progressively fewer observer moments will be at later times. This ongoing collapse of measure should possibly be a problem for people who think that quantum suicide is valid and in my earlier article about substrates [8] I pointed out that a similar decrease in measure may occur in mind uploading, making considerations of quantum suicide and mind uploading possibly closer than many people think. This article has taken no position on whether or not such a measure collapse should disturb anyone about to undergo it.

This provides possible answers to the Fermi paradox, the Doomsday argument and the simulation hypothesis.

The Fermi paradox [10] is answered by saying that alien civilizations start to perform quantum suicide early in their history, before we can detect them. This could be regarded as a special case of the answer to the Fermi paradox which states that almost all civilizations destroy themselves.

The Doomsday argument [11,12] is answered by applying the many-interpretations view which I previously mentioned and saying that quantum suicide would dramatically reduce the number of observer moments experienced in a civilization, and with repeated civilization-level quantum suicide this would be ongoing, so that most observer moments would actually occur before the civilization started doing this. This might be considered a special case of the simple answer to the Doomsday argument which is that the argument is correct and our civilization will end soon, the means of our civilization ending being civilization-level quantum suicide starting around the time of a technological singularity.

The simulation hypothesis [13] is answered by saying that one possibility is that a civilization might start civilization-level quantum suicide before it constructs any simulations, effectively reducing the measure for those branches in which the simulations are constructed and reducing the number of observer moments experienced in them. This could be considered a special case of the possibility suggested in the simulation hypothesis that almost no civilization will reach a technological level capable of producing simulated realities.

Civilization-level quantum suicide would imply more abstraction in the civilization's motivation than there is in ours. If civilization-level quantum suicide started to occur it could stop at some point when the civilization's motivation would become even more abstract, becoming something that we may not understand.

References

- [1] Everett, H. (1957). Relative State Formulation of Quantum Mechanics. *Reviews of Modern Physics* 29, pp454-462.
- [2] Web Reference: Strout, J. *Mind Uploading Home Page*. (2002). Retrieved 22 June 2003 from <http://www.ibiblio.org/jstrout/uploading/MUHomePage.html>.
- [3] Web Reference: *Mind Uploading Research Group*. (2002). Retrieved 22 June 2003 from <http://minduploading.org/>.
- [4] Egan, G. (1994). *Permutation City*. London: Millennium. (Fiction).
- [5] Web Reference: Bell, J. J. (2003). Exploring The "Singularity". *Mindfully.org*. Retrieved 25 May 2008 from <http://www.mindfully.org/Technology/2003/Singularity-Bell1may03.htm>.

[6] Vinge, V. (1993). The Coming Technological Singularity: How to Survive in the Post Human Era. *VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, March 30-31, 1993*. Retrieved 25 May 2008 from <http://rohan.sdsu.edu/faculty/vinge/misc/singularity.html>.

[7] Good, I. J. (1965). Speculations Concerning the First Ultraintelligent Machine. *Advances in Computers*, Vol 6, pp31-88. Retrieved 25 May 2008 from <http://web.archive.org/web/20010527181244/http://www.aeiveos.com/~bradbury/Author/Computing/Good-IJ/SCtFUM.html>.

[8] Web Reference: Almond, P. (2008). *Minds, Substrate, Measure and Value, Part 3: The Problem of Arbitrariness of Interpretation*. Retrieved 11 May 2008 from <http://www.paul-almond.com/Substrate3.pdf>. (Also at <http://www.paul-almond.com/Substrate3.htm>).

[9] Lewis, D. K. (2004). How many lives has Schrodinger's cat? *Australasian Journal of Philosophy*, Vol 82, No. 1, March 1 2004, pp 3-22. (A copy of the third Jack Smart Lecture delivered by David Lewis at the Australian National University on 27 June 2001). Retrieved 25 May 2008 from <http://www.arts.ualberta.ca/~pex/wordpress/wp-content/uploads/2007/04/lewis.pdf>.

[10] Jones, E. M. (1985). "Where Is Everybody?": An Account of Fermi's Question. *Los Alamos National Laboratory Report*, LA-10311-MS UC-34B, March 1985. Retrieved 25 May 2008 from <http://www.fas.org/sgp/othergov/doe/lanl/la-10311-ms.pdf>.

[11] Carter, B. (1983). The anthropic principle and its implications for biological evolution. *Philosophical Transactions of the Royal Society of London*, A310, pp347-363.

[12] Gott, J. G. III (1993). Implications of the Copernican principle for our future prospects. *Nature*, Vol 363, pp315-319.

[13] Bostrom, N. (2003). Are you living in a computer simulation? *Philosophical Quarterly*, 2003, Vol. 53, No. 211, pp 243-255. (Bostrom circulated a draft of this paper in 2001). Retrieved 8 September 2007 from <http://www.simulation-argument.com/simulation.html>. (Further information about this subject by Bostrom and others is at <http://www.simulation-argument.com>.)