

Quantum Brain Damage: Assuming you will survive quantum suicide may be simplistic.

Paul Almond

<http://www.paul-almond.com> info@paul-almond.com

17 July 2011

One view of the quantum suicide thought experiment is that you should be certain of your subjective survival if the many-worlds interpretation of quantum mechanics is correct, because you never observe the branches in which you do not survive. 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. If you think that, given the many-worlds interpretation being correct, your subjective survival is guaranteed in conventional quantum suicide, to have any expectation of the state in which your brain will be after the quantum brain damage experiment, you need a way of deciding how much damage must occur to your brain to allow some branches to be discounted as possible futures. This is a problem, because there is no obvious way of deciding where to draw the line – where enough brain damage has occurred that the brain cannot “observe”. The simplistic view used to justify the expectation of survival in quantum suicide is inadequate for this purpose. Instead, if the idea of expected survival in quantum suicide is to be viewed as viable in any way at all, a more sophisticated approach is needed for this more general case. Any expectation of survival in conventional quantum suicide would merely be a special case in such a more general view, and it is far from certain that the special case will be one of guaranteed survival. In fact, given that any general approach will have to examine the details of the situation, we might reasonably expect the details of the situation to affect the subjective probabilities of survival. We should therefore be sceptical of the idea that, given the many-worlds interpretation being correct, survival is guaranteed in the quantum suicide thought experiment, though an argument might be made that the true situation approximates this in some way, but with probabilities being influenced by various details of the situation.

1 INTRODUCTION

Quantum suicide is a thought experiment originally proposed by Hans Moravec [1] and Bruno Marchal [2], and further developed by Max Tegmark [3] that relates to the many-worlds interpretation of quantum mechanics (MWI) [4,5]. In the thought experiment, you participate in an experiment in which a sequence of iterations occurs. In each iteration, a quantum event occurs with two outcomes and a 0.5 probability of each. If one outcome occurs, the mechanism kills you. If the other outcome occurs, the mechanism does nothing: you remain alive.

As most people would understand the matter, each iteration has a 0.5 probability of killing you, so as the number of iterations increases, the chance of you surviving to the end decreases.

An alternative view, however, is that things are different if MWI is correct. According to this

view, whenever the quantum event occurs, the world splits and both possible outcomes occur in different worlds: there is always a world in which you survive. That would seem to follow from MWI, but some people take a stronger position in which your subjective survival is guaranteed. In this view, if MWI is true, in every iteration, from your point of view, there is always a future in which you survive, which is experienced by you, and a future in which you die, which is not experienced by you. As you never experience the future in which you die, you will never make an observation of that outcome, and so it can be eliminated from your possible futures. You should therefore always expect to survive.

This is a controversial position, and many people think that MWI being true would not necessarily imply that you should expect to survive the quantum suicide experiment: they think that even if MWI is true, you should still think you have a 0.5 chance of surviving in each iteration.

This article will argue that there is a possible problem with the idea that you should expect to survive quantum suicide. The *quantum brain damage* thought experiment will be described, which will be similar to the quantum suicide experiment, but with a difference: the mechanism does not kill you, but instead causes varying degrees of degradation of mental functioning. The idea that you can eliminate a future if there is no conscious observer will be shown to be too simplistic in this scenario, which will in turn suggest that it is too simplistic for the “conventional” quantum suicide scenario. A more sophisticated way of approaching personal continuity is needed, and it is not guaranteed that this will assure your survival in the quantum suicide thought experiment.

2 THE QUANTUM BRAIN DAMAGE THOUGHT EXPERIMENT

The quantum brain damage thought experiment has some similarity with the conventional quantum suicide thought experiment.

As with conventional quantum suicide, a sequence of iterations occurs. In each iteration, a random outcome is generated by some quantum device. With conventional quantum suicide, there are only two possible outcomes, but here, there are many possible outcomes. The outcome is a quantum-randomly generated integer. The lowest possible value is zero. The highest value is the number of neurons in your brain. Every number is equally likely. For example, if your brain contains 25 billion neurons, a random number between 0 and 25 billion inclusive is quantum-randomly generated. The random number could be generated in a number of ways. One way could be by sending low-intensity light through an aperture, so that it will be diffracted, before it reaches an array of detectors which is divided into separate regions, one for each possible number, with an individual photon being detected corresponding to generation of a number with a value corresponding to the relevant region, and with the size of the regions adjusted so that all numbers are equally likely.

After the random number has been generated, a mechanism automatically destroys that number of neurons in your brain. You can imagine the neurons to be destroyed as randomly selected. This could be done by some classical

mechanism, or it could be done using quantum mechanics. Alternatively, some consistent method could be used for selecting the neurons; for example, the mechanism might start with neurons nearer the outside of your brain first.

That is what happens in a single iteration. After each iteration, a similar process occurs in the next one. If neurons in your brain have been destroyed in previous iterations, then the range of random numbers generated is adjusted accordingly. For example, if in some iteration, the number of neurons in your brain has already been reduced from 25 billion to 13 billion, then a random number between 0 and 13 billion inclusive is quantum-randomly generated.

There is the issue that part of your brain is needed to control autonomous functions that keep you alive. Loss of too many neurons from such parts of the brain would cause clinical death and further loss of all brain functioning. This is a complication that is undesirable. To deal with it we will assume that such neurons are excluded from the thought experiment, or that your brain is connected to some life support system that can take their place – for example, by pumping oxygenated blood to your brain.

With the conventional quantum suicide thought experiment, the question posed by the thought experiment is as follows:

If MWI is correct, do you subjectively expect to survive?

Here, the question is different and is as follows:

If MWI is correct, after some number, n , of iterations, what is the average (mean) number of neurons that you subjectively expect to have in your brain?

n can be any number that interests us. For example, if $n=1$, the question is about how many neurons (on average) you expect to have after going through just one iteration. If $n=100$, it is about how many neurons you expect to have after going through 100 iterations.

3 TRYING TO ANSWER THE THOUGHT EXPERIMENT'S QUESTION

3.1 If you do *not* expect to survive conventional quantum suicide

An obvious view to take, for many people, is that you should not expect to survive quantum suicide, even if MWI is true – that you should view yourself as having a 0.5 probability of dying. If you take this view, the quantum brain damage thought experiment should cause you little difficulty. You can simply say that one outcome is as likely as any other, irrespective of whether you are conscious after it has happened, or how much damage is done to your brain.

According to such a view, if at the start of some iteration you have 20 billion neurons, the minimum number of neurons that will be destroyed is zero and the maximum is 20 billion. Any value in-between is equally likely, so the mean number of neurons you expect to have left after this iteration is 10 billion.

3.2 If you *do* expect to survive conventional quantum suicide

If you expect to survive quantum suicide if MWI is correct, according to the simple idea that you only experience those branches in which you continue to exist as a conscious observer, the thought experiment causes you more of a problem. It would seem that you need to eliminate those branches in which you do not continue to exist as a conscious observer, but *how do you do this with partial destruction of your brain?*

You cannot reasonably say that you should experience no brain damage at all. You should be able to imagine existing as a conscious observer if you lose one, two, ten or even a hundred neurons: your chances of coming out of this unscathed are extremely remote. But how far can the process of destruction go before the rules of the game say that “too much of you is gone” and you are no longer a conscious observer? Could you lose a thousand neurons? A million? Could you be left with just sixty percent of your brain? What about fifty percent? If we start with your brain completely intact – and conscious – however many neurons you think are needed, the case can always be made for saying that the

number of neurons can be reduced just a bit further without substantially changing the situation. We could therefore end up with a situation in which every outcome, from none of your neurons being removed to all of them, counts as a possible future.

Similarly, you can start at the other end of the scale with no neurons left at all. If you have no neurons left, you are clearly not conscious, so you should not regard the outcome where your brain is completely destroyed as subjectively possible: you should remove it from your subjective list of possible outcomes. But what about the outcome where you are left with one neuron? It is hard to see how one neuron counts as being a conscious observer, so that outcome should be removed too. But what about two neurons? Ten? A hundred? What if you are left with ten percent of your brain? Fifty percent? However low you think the number of neurons must be for you to your brain to be “dead” – in the sense of you not being a conscious observer – and for the path not to “count” – it should be hard to see how we cannot take a brain with that number of neurons and add one neuron without substantially changing things. However, if we extend such reasoning to its limit we end up with the result that *none* of the branches can be counted.

There does not seem to be any point at which an obvious line is crossed and you cease to be conscious. Suppose we say that there is some minimum number of neurons that you need. Why? What is so special about this number? The simple idea that only the branches where you continue as a conscious observer should be counted is inadequate for dealing with this situation.

4 THE PROBLEM

4.1 There may not be an answer to be had.

The thought experiment causes a problem if you think that you should subjectively expect to survive quantum suicide. The conventional quantum suicide thought experiment presents you with two distinct possible outcomes – one in which there is a conscious observer and one in which there is not – with the distinction being the relatively clear-cut one between your normal self and a corpse in whatever state it is left by the mechanism that kills you. The *quantum brain*

damage thought experiment described here, though, removes this clear distinction between outcomes. Instead, it presents you with a *scale* of possible outcomes, and no obvious way of deciding which outcomes to count as containing conscious observers. It is clear that the completely intact version of you at one end of the scale *can* be counted and that the completely destroyed version of you at the other end of the scale *cannot* be counted. It is less obvious what should be done about the versions in-between: how many neurons need to be intact in a branch for it to be counted? If you cannot resolve this issue then you cannot answer the question: *If MWI is correct, after some number, n, of iterations, what is the average (mean) number of neurons that you subjectively expect to have in your brain?*

It may seem that, if you think that you should expect to survive conventional quantum suicide, the problem with the quantum brain damage scenario is that you do not know the answer. The problem, however, runs deeper than that: *the scenario brings into question the very idea that there is an answer to be had with this way of thinking.*

4.2 The Problem with an Abrupt Transition

The idea that there is a transition point between being a “conscious observer” and being “unconscious”, and that a brain can move across that point by the removal or addition of a single neuron, should appear implausible. The brains on either side of the transition point – wherever it is supposed to be – are essentially the same, so a profound change in something’s nature is happening with an insignificant change to the underlying physical structure. Someone might object that the change is not really very small, but we could always respond by making the steps still smaller. Instead of quantum-randomness being used to generate a number of neurons that are going to be removed in each iteration, we could use it to generate a number of molecules: there would then presumably be the situation where removal or addition of one molecule is supposed to move across the transition point. The idea of a profound change in something’s nature being caused by a trivially small change in the underlying physical system is not something we normally associate with physical systems and their properties, and it is so unusual that, as well as being implausible, it

suggests that minds are being treated in a radically different way to everything else, and this should be suggestive of dualism.

The problem also has similarities with a problem that some people regard as existing with the “consciousness causes collapse” interpretation of quantum mechanics: how do you draw the line between conscious observers and everything else? Many people who think that MWI is likely to be true might have problems with this issue when they see it in that context, so the same issue should suggest problems to them when they meet it in the context of quantum brain damage.

Although it presents problems of plausibility and is suggestive of dualism, let us suppose that there *is* some point along the scale from completely destroyed to completely intact at which a “conscious observer” starts to exist: any branches with at least this number of neurons are going to be subjectively counted as possibilities for your future. What is it that is supposed to determine where that point is? We should demand that whatever rules govern this are justifiable and relate to the underlying physical system – the underlying substrate – but it should be hard to imagine how such rules are going to deliver the abruptness of transition that is needed when nothing else about the system seems to have this abruptness. If the rules do *not* relate to the underlying physical system then we seem to be in the realm of some “metaphysical rules” about what is conscious or not that work in some mysterious, unknowable way, and which have some kind of existence apart from things like brains. This, again, should be suggestive of dualism and should be implausible.

So, there must be some rules that determine when a thing is conscious and when it is not – whether these relate to the underlying physical system or whether they are “metaphysical” in some sense – and they give the abrupt transition from consciousness to absence of it that is needed to declare some branches irrelevant. You do not know what these rules are. Nobody does. That being the case, how can you be sure that the rules are even going to give the abrupt transition from consciousness to absence of it as a single small change is made? A serious problem has crept in here. The idea that you should expect subjectively to survive quantum suicide is based on an idea that there are branches in which you

are not conscious and branches in which you are, but that idea has now been shown to rely on something else – some other way of approaching this and dealing with consciousness. The appeal to the simple idea that you are conscious in some branches and not others no longer works. Instead, you are reduced to *hoping* that the underlying, unknown approach to consciousness – the correct one, whatever it is – agrees with this idea and gives this abrupt transition as a neuron is removed or added. We should be far from convinced that this is the case: with metaphysical rules that have nothing to do with your brain, we do not have any idea what they would do, and with an approach based on the physical system in some way, such an approach is going to have to get involved in the actual substrate that causes human minds – the underlying mess of matter – and things may be more involved.

4.3 The Problem with a Gradual Change

You may think that there is no abrupt transition between consciousness and absence of it as a single neuron is removed or added. What if there is, instead, a gradual decline of consciousness as individual neurons are removed? The problem here is that it seems to leave all branches as ones in which you are there to observe – even the ones with no neurons or a single neuron in which you would clearly view yourself as dead in quantum suicide experiment. This approach therefore seems to be unworkable.

If you think there is a gradual decline as individual neurons are removed, you may think that some kind of measure approach can be used – that the probability of finding yourself in a particular future is somehow related to how conscious you are in it – so that as the consciousness in some future decreases, it is increasingly unlikely that you will find yourself in that future. The problem with this is that it is, again, a significant departure from the simple idea that you are there to observe in some branches and not in others, and that just the former constitute possible futures for you. Instead, you must think there are some underlying rules that describe what the probabilities are. As before, these rules must be metaphysical rules that are just there – for no reason (which is implausible and suggestive of dualism) – or they must somehow relate to the underlying physical system. Either case has the

problem that, whatever they are, the rules may not agree with the simple assumption in the quantum suicide thought experiments: the assumption is again unsafe.

4.4 The Problem with Consciousness Only Being Relevant at Some Levels of Neurological Sophistication

You may think that it is a mistake even to talk about a transition, abrupt or gradual, from consciousness to lack of it as neurons are removed – that consciousness is only a meaningful idea when there is a certain amount of sophistication in the brain, and that when this is reduced, consciousness, rather than decreasing, becomes irrelevant. With this view, the question of where to draw the line between consciousness and lack of consciousness may not be a meaningful one: the concept would just become less “useful” as neurons were removed. The problem with this view is that it leaves no way of saying which branches should be considered as possible futures and which ones should not be.

4.5 The Problem with Saying that it is “Something Else” that Determines which Branches are Possible Futures

You may think that your consciousness declines gradually as neurons are removed, or becomes less relevant, but that this consciousness is not the same as what is needed to say whether you are there to observe a particular branch. Maybe something else is needed to say that you exist as an “observer”, and maybe this ceases to exist when your consciousness or number of neurons falls below a certain level? Such an idea will be of no help. This merely involves positing some kind of “secondary consciousness” – whatever it is that you are supposed to have that makes you an observer – and we have the existing problem of how it is supposed to change between existing and not existing with a small change to your brain. However this is dealt with, it must be done using an approach that uses general, underlying rules, with the problems that have previously been identified for these.

We return now to the question asked in the quantum brain damage thought experiment:

If MWI is correct, after some number, n , of iterations, what is the average (mean) number of

neurons that you subjectively expect to have in your brain?

If you think that your subjective survival is assured in the conventional quantum suicide experiment, the problem is that the simple idea used to justify that idea – that you can eliminate branches in which you are “not there to observe” from your set of possible futures – has been shown to be inadequate – whether you think that consciousness abruptly ceases to exist when a certain number of neurons are used, that there is a gradual reduction in consciousness as neurons are removed, that consciousness is only a meaningful idea at some level of brain sophistication or that it is not directly the degree of consciousness but “something else” about your brain that determines which branches should be viewed as possible futures. The idea that you should consider your survival in the quantum suicide scenario as certain should be therefore be treated with some degree of scepticism.

5 CONCLUSION

The quantum suicide thought experiment involves a situation with two possible futures, one of which involves your survival and one of which involves your death, with the future that happens being determined by some quantum event. If the many-worlds interpretation of quantum mechanics (MWI) is true, both futures happen on different branches. One view of this is that you can eliminate the branch in which you die as a possible future for you, subjectively, and that therefore you should view your subjective survival as guaranteed. This idea has been questioned by the quantum brain damage thought experiment.

In the quantum brain damage thought experiment, there is no clear distinction between killing you and not killing you. Instead, a degree of damage occurs to your brain, with some neurons being lost, ranging from no loss of neurons at all to complete destruction of your brain, with many possible outcomes in-between in which varying numbers of neurons are lost. The thought experiment suggests the following question:

If MWI is correct, after some number, n , of iterations, what is the average (mean) number of

neurons that you subjectively expect to have in your brain?

If you think your subjective survival is assured in conventional quantum suicide, the problem here is one of deciding which outcomes can be discounted as possible futures with quantum brain damage. You can clearly discount the branch in which your brain is completely destroyed – you are obviously dead – and you should clearly count the outcome in which no damage at all occurs: this is just like the conventional quantum suicide outcome in which you survive. It is the intermediate outcomes that are the problem. You need to decide which of them involve a conscious observer – which of them will be counted as possible futures. The idea that you count branches where you are there to observe, and do not count branches where you are not there to observe is too simplistic to deal with this. Instead, there is a need for a more general approach. Such an approach would seem to need some general, underlying rules that indicate when branches should be counted and when they should not.

We might imagine that the general, underlying rules are “metaphysical”, having nothing to do with the physical nature of your brain, but instead dealing with the issue in a different way. This should seem to be dualism and should seem implausible. If such metaphysical rules exist, your survival in conventional quantum suicide would only be assured if that occurred as a special case of these rules, but there should be little reason to think that that would be the case: we really have no idea of how such rules would work.

We might imagine, and this is more plausible, that the general, underlying rules take account of the physical situation, and that what happens is implied by the details of the physical situation in some way and can be justified just by examining the physical situation. Such rules would have to come out of some more general approach – some way of analyzing the situation that indicates which branches should be counted and which should not. If you think that you would survive conventional quantum suicide, this would need to follow as a special case of such general rules. However, we do not know what such rules are, so we cannot be sure about what any special cases of those rules would be like. Such general rules might need to involve many things not

typically considered in the conventional quantum suicide experiment, such as the method of death, how quickly it occurs, other aspects of the experimental setup and the way the human brain works. In fact, we might have every reason to think that a general approach *would* have to take account of such things. The very thing that the approach would have to do – deal with a continuum of brains, each slightly different from its neighbours – would seem to require that whatever approach is used takes into account the fine details of the situation, and that should seem to make it likely that details of the situation could affect the outcome. This should give us reason to doubt that the simple reasoning behind the idea that your survival in conventional quantum suicide is guaranteed gives an accurate representation of the situation.

One way of answering this is simply to reject the entire idea that any branches can be eliminated in

quantum suicide scenarios, and many people will already think that this is the correct approach.

What has been said here, however, does not mean that the idea that you should expect to survive quantum suicide must be viewed as *completely* wrong. Conventional quantum suicide is an extreme of the kind of situation being considered here, and we might think that it tells us *something*. If, however, we accept that the idea has any validity at all, we should at least accept that the reasoning used to support this idea is a special case of a more general approach and that it may only approximate the actual situation. In this view, you should possibly think that your chances of surviving quantum suicide are based on various things, and your survival, while possibly considered more likely due to some general version of the quantum suicide argument, may be far from assured.

REFERENCES

- [1] Moravec, H., 1988. The Doomsday Device. *Mind Children: The Future of Robot and Human Intelligence*. Harvard: Harvard University Press. p.188. (Also available at: <http://books.google.com/books?id=56mb7XuSx3QC&hq=PP188&pg=PA188> [Accessed 10 December 2010].)
- [2] Marchal, B., 1988. Mechanism and Personal Identity. *Proceedings of WOCFAI 91* (Paris. Angkor), pp.335-345. (Also available at: http://iridia.ulb.ac.be/~marchal/publications/M&PI_15-MAI-91.pdf [Accessed 10 December 2010].)
- [3] Tegmark, M., 1997. *The Interpretation of Quantum Mechanics: Many Worlds or Many Words?*. [Online] arXiv:quant-ph/9709032. Available at: <http://arxiv.org/abs/quant-ph/9709032> [Accessed 8 December 2010]. (Also available at: http://xxx.lanl.gov/PS_cache/quant-ph/pdf/9709/9709032v1.pdf [Accessed 12 December 2010].)
- [4] Everett, H., 1957. Relative State Formulation of Quantum Mechanics. *Reviews of Modern Physics*, 29, pp.454-462.
- [5] Price, M. C., 1995. *The Many-Worlds FAQ*. [Online] The Anthropic Principle. Available at: <http://www.anthropic-principle.com/preprints/manyworlds.html> [Accessed 12 December 2010]. (Also available at: <http://www.hedweb.com/everett/everett.htm> [Accessed 12 December 2010] and <http://kuoi.com/~kamikaze/doc/many-worlds-faq.html> [Accessed 29 June 2011].)