

Quantum Suicide and Inconsistency: Assuming you will survive quantum suicide may be simplistic.

Paul Almond

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

23 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. Such an idea involves inconsistency. If many-worlds is not true, you must accept that there is a probability that you will die – whether you regard this as a “future” or not or whether you think this means “having no future” – and regardless of the fact that you will not be there to observe such an outcome. Whatever the semantics, or however it is philosophically viewed if many-worlds is not true, this seems, subjectively, to be a possible outcome in some sense. When many-worlds is true, however, we are supposed to forget all this and regard such a possibility as incoherent, because the future in which you survive always occurs as well, and is always available. How can a possibility that is accepted as incoherent suddenly become incoherent when a “better” alternative is on offer? People who think that you are assured of your subjective survival in the quantum suicide experiment seem unable to decide whether dying is incoherent or not. Because of the inconsistency in the argument used to support it, we should be sceptical of the idea that, given the many-worlds interpretation being correct, survival is guaranteed in the quantum suicide thought experiment. Nevertheless, this should be regarded as a mitigated attack on the idea of expecting to survive quantum suicide: a more general approach to personal continuity – one based on “measure”, that takes account of details of what is happening in the substrate – might allow the idea to survive in some form, though if this is the case, the details of the situation would probably be important with regard to probabilities, and it is unlikely that survival would be completely assured.

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 a process 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 only a 0.5 probability of surviving in each iteration.

It will be argued here that there is an issue of inconsistency with the view that you should expect to survive quantum suicide.

2 THE JUSTIFICATION FOR EXPECTING TO SURVIVE

Let us consider an individual iteration of the quantum suicide experiment. The quantum event is about to happen with the result of “1”, meaning that you will live, or “0” meaning that you will die. Some people think that, if MWI is correct, you should consider your survival of this iteration guaranteed, while if MWI is not correct you should think you have a 0.5 probability of living and a 0.5 probability of dying. We will consider what we might expect if MWI is *not* true and if MWI *is* true.

2.1 If MWI is *Not* True

The justification for thinking that you have a 0.5 probability of dying if MWI is *not* true is that you are only going to have one future. There is a 0.5 probability that it will be a future in which you are dead and a 0.5 probability that it will be a future in which you are alive. Unlike the case where MWI is true, you cannot assume that you will be in the future in which you are alive because there may not be such a future for you to inhabit. You may end up in the future in which you are dead because, if that is the future that actually happens and there is nowhere else for you to be: the fact that you will not be able to observe this is irrelevant. Presumably, if such a future occurs, you will be aware until the second that the machine kills you and then... nothing: your future will be one of having no future. You might think in terms of there being a probability that you have no future.

2.2 If MWI is True

The justification for expecting to survive quantum suicide if MWI is true is that you should eliminate futures that you will not observe from your consideration: subjectively, futures which you will not inhabit as an observer are no longer regarded as being futures. When you consider your possible future situations, you should only consider those situations in which you are an observer. There will be a future in which you survive, so you should expect to find yourself in it – not in the future in which you are dead: you can clearly not find yourself in a future in which you are dead.

3 THE INCONSISTENCY

There is an inconsistency between the way that futures in which you die are being dealt with in the situation in which MWI is not true and the situation in which MWI is true.

If MWI is *not* true, you have probabilities for two different futures, and one of these is a future in which you die. It is regarded as coherent to say that you might end up this future. This, of course, does not imply a claim that you will “observe” such a future. Clearly, questionable metaphysical ideas aside, you will not. However, the fact that you will not observe such a future does not mean that it is regarded as incoherent: it is regarded as something that might happen and it is assigned a probability of 0.5. You might say that this is not really “a future” for you because you will not be able to observe it, but that would really be semantics: in that case, you would be regarded it as “a future in which you have no future”, or you might say that there is a 0.5 probability that you have no future. However you describe it, the fact remains that you regard this “thing” or this “absence of anything” as having a probability of 0.5. You might try to dismiss this outcome completely, but it leaves you with a problem: your chance of survival is only 0.5, so what does the other 0.5 needed for the probabilities to add up to 1 actually mean? It seems unavoidable that the future in which you die must be regarded, in some sense, as being coherent, and that it needs to be accepted as one of the possible outcomes facing you.

If MWI *is* true, you have two possible futures, but then, after some consideration, you remove the one in which you die as a possibility. The idea is that it is incoherent to talk of this being your future, because you are not there to observe it. Instead, you expect to find yourself in the future in which you do not die.

If MWI is not true, the idea of a future in which you die is being treated as coherent, but if MWI is true, the idea of a future in which you die is being treated as incoherent. This is the inconsistency. If you think that survival is assured in quantum suicide, we should ask you:

Is the idea of a future in which you die a coherent idea or not?

A person who insists that his survival is assured in quantum suicide, if MWI is true, seems unable to make his mind up!

Now, if you think that your survival is assured in quantum suicide, with MWI being true, you could reply, “Yes, but you do not understand. The many-worlds situation is *different*, because you have an alternative to death: there is always going to be a future in which you survive.” The problem with this is that the future in which you die has supposedly become incoherent because an alternative is now on offer – not due to any real change in the nature of this future. How can this be the case? How can something which is considered a *coherent* description of the fate that awaits you when there is no alternative suddenly become *incoherent* when there is something else on offer? This should seem suspicious.

4 THE INCONSISTENT MATHEMATICIAN ANALOGY

An analogy for what seems to be occurring here is as follows:

Suppose a mathematician solves a problem in applied mathematics and obtains an answer which is supposed to correspond to some feature of the real world. He initially has some concerns about whether this answer really does relate to the real world, but he decides that it must do so: there is no other answer and the calculation itself must provide at least one answer that relates to the real world. He accepts the answer as valid and corresponding to some feature of the real world.

The mathematician then performs a slightly different version of the calculation and obtains two answers. One answer is the same as the one he just obtained, and causes him the same concerns again. This time, however, he rejects the answer and decides that it cannot relate to the real world. Why? He does not need this answer now: he also has the second of the two answers, and this one causes him no concerns. He decides only to use the second answer.

The mathematician’s approach should seem questionable. In the second version of the calculation he decided that an answer could not correspond to reality, but it was good enough for him first time round! He seems happy to accept that the answer corresponds to reality when he

really has to in order to solve a problem, but not when he has any alternative. How can that be justified? It is the same answer in each case! How can an answer that corresponds to something in reality stop corresponding to something in reality when a different answer is available? The answer itself has not changed.

5 AN EXTREME RESPONSE

The argument given here is about consistency in how futures in which you die are viewed in the situation in which MWI is not true and the situation in which it is true. There is one response to the charge of inconsistency that will occur to some advocates of expecting to survive quantum suicide: this is to dispose of any inconsistency by saying that MWI is true.

This would be a bold position to take. It would amount to saying that it is incoherent to expect your death to result from the quantum suicide experiment and, as this is exactly what seems to result sometimes in a reality in which MWI is not true, a reality in which MWI is not true is incoherent, meaning that accepting that MWI is true is the only coherent position. Incidentally, such an approach might imply that the more extreme idea of *quantum immortality* should also be adopted.

We are *not* arguing for this position here, but merely observing that it is a possible reply.

6 DOES THIS MEAN THAT EXPECTING TO SURVIVE QUANTUM SUICIDE IS MISTAKEN?

6.1 A measure approach might be imagined.

The argument given here may seem to be suggesting that any expectation of surviving quantum suicide is misguided. The argument is certainly intended to show that it is incorrect to think you will survive quantum suicide solely on the basis of the simple idea that you cannot observe a future in which are dead. We are, however, not intending this as a complete attack on the idea, but rather as a mitigated attack on it. It is conceivable that the idea of expecting to survive quantum suicide might survive the argument presented here in some form – with the justification made somewhat different and the certainty lost.

Some idea *with some resemblance* to the argument that you can only have futures which you can observe may relate to intuition about how our minds seem to relate to our brains. Your mind seems to be associated with your brain: where your brain goes your mind follows. Now, this may seem to be a very obvious thing to say. Where else would your mind go? It can hardly slide out of your head and into a nearby houseplant: such things lack the sophistication to hold your personality in any form that you would recognize. But why do we require that anything that is a destination for your mind has enough sophistication to produce something like your mind as it is now? When you die, your mind seems to end up being reduced to nothing over a period of time, as the substrate on which your mind is based becomes increasingly lacking in the sophistication needed to produce what you normally experience as a mind: your mind can end up following a path into the future in which the substrate becomes inadequate to support anything like thinking and in which your consciousness fades to nothing. That being the case, what is there to stop the process happening all at once? How am I protected from some hypothetical situation as follows?

My mind is here, now, in my brain, and then it abruptly ends up in a houseplant nearby – and, because a houseplant lacks any of the properties needed for my mind to exist in it with anything like the form my mind has now, I actually end up with no mind to speak of: I end up dead.

Most people would regard such an idea as absurd, but that is the point, really: we do not worry about such things. We can conceive of our deaths when we imagine our minds, in a non-MWI reality, inhabiting decreasingly functional brains, so that the mind, with nowhere to go, must ultimately inhabit something unable to support it in any meaningful form, and must fade away. At the same time, however, when presented with scenarios in which our minds follow paths into absurd substrates, such as houseplants, where this happens, more quickly, we reject it as absurd: we expect our minds to stay in our brains – but one view on this might be that we should expect our minds to stay in our brains when they are functional enough – that, somehow, we should expect our minds to have a tendency to follow paths that prefer highly functional substrates to less functional ones. The possible connection of this to survival in

quantum suicide should be clear: an argument to justify an expectation of survival in quantum suicide might say that you should expect your mind to tend to follow paths where a functional substrate is available, where they are available, and expecting to “enter” the branch where you are dead might be viewed as somewhat similar to expecting to “enter” the houseplant in the hypothetical situation just described.

A proper justification would be needed for such a position, and it is not our intention to provide it here, or even claim that such an argument must be valid. If such a justification were to be obtained, it would probably imply an approach to personal continuity that takes into account many details of the situation, such as the physical details of the substrates on which the mind is based at various times, and the changes happening to them. There would be no sense in which an expectation of survival in quantum suicide would be justified based on a particular situation being one that you cannot observe: in fact, to avoid having the same problem that is the subject of this article, such an approach would have to allow futures in which you die as possibilities. Rather, the approach would have to give some preference to various futures based on the physical details of the underlying situation with a result that has some similarity with what is claimed to happen by those who expect to survive quantum suicide. However, given that the approach would be based on physical details, and death would not be rejected in principle, it is more likely that the approach would be one that deals with things in terms of *measure* and there would still be some probability of dying. How low this probability can become – how closely the situation resembles the one of guaranteed survival justified by the simplistic reasoning about quantum suicide – might depend on the particular characteristics of the situation – but it might be considerably different from it in many real-world experiments that we might imagine performing.

6.2 This is a *mitigated* attack.

The assumption that you would survive quantum suicide and the justification typically given for this idea should be viewed as simplistic. Nevertheless, it is conceivable that the idea might survive in some form as a consequence of some kind of “measure” approach to continuity which favours the branches in which you survive

to some degree. Even if this is the case, however, it is not clear that it would make much difference in most of the real-world scenarios that we might imagine. We are not claiming here that such a measure approach *would* be viable, or that this “weak” form of survival in quantum suicide would apply, but we have stated this to show that, while we have attacked the idea of assuming survival in quantum suicide, we have intended this to be a *mitigated* attack on the idea, and that we accept that, while the justification for the idea has serious problems, the idea might survive in *some* form.

7 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 should be questioned on the grounds of consistency.

If MWI is not true, you are required to accept the possibility that a possible outcome, from your point of view, is death. Whether you are there to observe this or not is immaterial: we all know what it would mean to perform an experiment, in a reality in which MWI is not true, and to think that you may die as a result. It might be said that the future in which you die cannot really be a future for you, as you have no future if you are dead, but this is really just semantics: in that case, we would have to accept that, if MWI is true, *there is a probability that you have no future*. There is no way of avoiding this probability. If we ignore it we are left with the remaining probabilities not adding up to 1. You may be happy with this: you may decide that

death is somehow a “hole” in your set of future situations, but it remains that if MWI is not true, the idea of dying, in some form, has to be accepted: your statistical calculations have to acknowledge it, for it may happen.

Given that the future in which you are dead must be acknowledged, in some sense, as a possibility when MWI is *not* true, it should seem inconsistent to dismiss it as a subjective possibility when MWI *is* true, purely on the grounds that a “better” alternative” – and by this we mean a less philosophically bothersome one – will always be available – yet that is precisely what people who expect to survive quantum suicide are doing when they dismiss the possibility, subjectively, of dying in the quantum suicide experiment on the grounds that you are not there to observe the outcome. The idea of an outcome which you are not there to experience is regarded as incoherent – yet this outcome was fine when there was the possibility that there would be no alternative!

We may use the analogy of an inconsistent mathematician who rejects an answer as meaningless, when he has another answer available, yet accepts that answer when it is all that he has and an answer is demanded.

While this may seem to be an extreme attack on the idea of expecting to survive quantum suicide, we actually intend it as a *mitigated* attack. It should dispose of the idea that it is *incoherent* to think that you may die in the quantum suicide experiment. The idea might, however, be justified using some more general approach to personal continuity. Such a general approach, however would need to take account of various details of the situation, and it is unlikely that it would give a result that agrees exactly with the idea that survival is subjectively assured. Nevertheless, it is conceivable that the idea of expecting to survive quantum suicide could be maintained in some form.

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&ipq=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].)