

Why don't you turn into a house plant and die? Assuming you will survive by mind uploading may be simplistic.

Paul Almond

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

31 July 2011

Some people believe that if a copy of your brain were made – perhaps implemented in computer software in a process known as mind uploading – you would have survived even if the original brain were destroyed. The mind uploading brain damage thought experiment is described, which questions this idea. The thought experiment involves the production of copies with different degrees of damage. The simple idea that a copy is a continuation leads to absurd results, one of which is that your own death should be expected in the next instant. To avoid these results, we need to dismiss some copies as valid continuations, or assign different copies different probabilities of being your future, but the simple idea that a copy assures your survival tells us nothing about this. We might say that we just do not know what the answer is, but the point is that there is no answer to be had while using this simple view, and it should be regarded as incoherent. Instead, a more general approach is needed. Any expectation of survival through mind uploading 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 survival is assured through mind uploading, though the general approach may approximate this in some way, but with the subjective probability of survival depending on the details of the situation.

1 INTRODUCTION

Some people believe that if a copy of your brain were made – perhaps implemented in computer software in a process known as “mind uploading” [1,2] – you would have survived even if the original brain were destroyed. The idea here is that the copy could be considered a future version of you, just as your brain is, and that it is just as valid to assume that you will be continued through the existence of your copy as through the existence of your brain.

This article will argue that there is a possible problem with the idea that you should expect to survive through mind uploading. The *mind uploading brain damage* thought experiment will be described. This will show that the idea that that you can expect to survive if a copy of you is made is too simplistic to indicate when you will survive – and that it is really an incoherent position to take. A more sophisticated way of approaching personal continuity is needed, and it is not guaranteed that this will assure your survival in a given scenario where it depends on mind uploading.

2 THE MIND UPLOADING BRAIN DAMAGE THOUGHT EXPERIMENT

The mind uploading brain damage thought experiment has some similarity with the quantum brain damage thought experiment, previously discussed [3], and is as follows:

A very accurate brain scanning machine is able, very quickly, to capture an extremely high resolution scan of your brain. The accuracy of the machine is as high as is theoretically possible: even individual atoms can be imaged. The information captured by the brain scanning machine is stored in a “scan file”¹ as a series of binary digits (bits). The scan file can be read back in another process and used to create a computer simulation of your brain, which can be allowed to interact with the real world or with a virtual reality environment. The idea is that this can continue your conscious experience from the moment of the scan: the technology is intended as a way of prolonging life beyond biological

¹ The term “scan file” is taken from Greg Bear’s science fiction novel, *Permutation City* [4], which makes reference to mind uploading.

death. You, however, will be participating in a somewhat morbid experiment with this technology.

In a few minutes, you will be given a general anaesthetic and then your brain will be scanned by the brain scanning machine. As soon as the scan is complete, your brain will be destroyed: any conscious continuation that you are going to have after being anaesthetised must come from mind uploading.

First, let us assume that the scan file will be used to make a single software copy, which is activated after your death. From your perspective, just before being made unconscious for the brain scan, what do you expect the future to be like? You may think that the copy is not a continuation of you – that it is not part of your future. If you think this, you should expect to die shortly, and the existence of the copy is irrelevant with regard to your future: it may as well not have been made. Many people think, however, that the copy *would* count as a continuation of you – that you should actually expect to find yourself existing in the future as the copy. If this is the case, mind uploading can be used as a loophole in death. This idea will be explored by extending the thought experiment.

Imagine now that the thought experiment is to be performed as just described, but *two* copies will now be made instead of just one. One copy will be activated in a virtual reality simulation of London. The other will be activated in a virtual reality simulation of Paris. At the moment of activation, the copies will be the same, but they will soon diverge due to having different experiences.

What should you expect the future to be like? If you do not think that a copy counts as continuation, it is as before: the copies are irrelevant. If, however, you think that a copy does count as a continuation, you have the issue of there being two copies now. There is no profound difference between the two copies, so you have no grounds for thinking that one is a “valid” continuation and the other is not: if you accept one as a valid continuation, you must accept the other. The only way that you can deal with this is probabilistically: you must think that each copy is equally likely to be your future. You should think that there is a 0.5 probability of finding yourself in simulated London and a 0.5

probability of finding yourself in simulated Paris. An analogy for what is going on here is provided by the many-worlds interpretation of quantum mechanics (MWI) [5,6]: you are “splitting”, and splitting in MWI is dealt with in terms of probability.

We will now extend the thought experiment still further. Instead of two copies, a very large number of copies is made from the scan file. Before the process starts, a second file, known as the “random file” is generated. This has the same length (the same number of bits) as the scan file, but it consists entirely of randomly (or pseudo-randomly²) generated bits.

The first copy is made using the scan file with complete accuracy as would be expected.

In making the next copy, not all of the scan file is read: one bit is ignored. Instead of the value of this bit, the value of the corresponding bit (the one in the same position) in the random file is used.

In making the next copy after this, the bit that was not read in making the last copy is ignored, and the corresponding bit from the random file is used again, but now another bit in the scan file is ignored, and the value of the corresponding bit in the scan file used in its place.

For the copy after this, the two bits in the scan file ignored in making the last copy are also ignored, and the corresponding bits in the random file used in their places, but now a third bit in the scan file is also ignored, the corresponding bit in the random file being used instead.

This continues, with each subsequent copy involving one bit less of the scan file actually being read and a bit from the random file being used in its place. The penultimate copy to be made is made by reading only one bit from the scan file and the remainder from the random file. For the last copy, no bits from the scan file are used at all: every bit comes from the random file: the last copy has nothing to do with your brain.

The thought experiment raises the following question:

² The distinction does not matter here.

If mind uploading does result in actual continuation of you – if software copies of your mind do constitute your future – what is your expectation of what is going to happen to you?

3 IMPLICATIONS OF THE THOUGHT EXPERIMENT'S QUESTION

3.1 If you do *not* expect to survive the mind uploading process.

An obvious view to take, for many people, is that *none* of the copies are continuations. According to this view, when you are anaesthetised for the brain scanning process you expect never to wake up: you are going to die while asleep. The existence of the copies is irrelevant to your future expectations. This view might be held for a number of reasons.

One reason could be that the copies are computer programs, and that only biological brains, or systems of some specific kind, are considered to be conscious, as argued by John Searle [7]. In this case, the thought experiment could be modified so that it becomes one in which each copy is a biological brain, made in some kind of cloning machine.³

Another reason could be that the copies are *copies*: continuity of self might be considered to occur only in the original. A problem with this view could be gradual replacement of matter in the human body. One answer to this is that such replacement of matter is at least gradual. However, we might imagine scenarios in which the rate of replacement of matter increases, with there being no firm line between sudden replacement and gradual replacement.

3.2 If you *do* expect to survive the mind uploading process.

3.2.1 The Problem of Many Maximally Damaged Copies

If you expect to survive the mind uploading process, according to the simple idea that software copies of you count as continuations of

you, the thought experiment causes you more of a problem.

The only way that this situation can be approached is probabilistically. You should expect your future situation to correspond with one of the copies. The obvious approach is to view things in terms of probability, and to view every copy as being an equally likely future for you – just like a different branch in MWI.

If you take this position, what would be your expectation of the future? Expecting to find that you are a completely undamaged copy made from all of the scan file is clearly unrealistic. There are a huge number of copies, and only one of them is made with every bit from the scan file: all of the rest have at least one random bit in them. In fact, it seems that you should expect to be seriously damaged. The amount of damage in the copies ranges from zero – no loss of information – to total – not a single bit from the scan file being used. Only half of the copies would have been produced in a way that even involved use of 50% of the bits from the scan file, so you should expect only a 0.5 probability that you will find yourself as a copy made using at least half of the scan file – the missing data being made up with random bits. In half of the outcomes, from your perspective before this happens, you end up as a copy that is extremely damaged – being effectively made from a corrupted scan file in which half of the data is replaced by random data. This may seem to suggest that, while you should expect to survive, you should expect to be severely damaged, finding yourself with your mental faculties impaired.

Things are much worse than this, however. Suppose that instead of just one version of this last copy, $10^{10,000}$ versions of it were made, each made using a different random file and with all the bits being taken from the random file rather than the scan file. The number of such copies would far outweigh the numbers of all the other, less damaged copies, so it would be almost certain that you should expect to find yourself as one of these copies in the future. However, these copies would all be based on nothing more than random information. They would not have anything that we would imagine as consciousness. You would effectively be dead. The problem is that, whether it is done explicitly with the $10^{10,000}$ versions or not, a situation like

³ In this case we might assume that some artificial mechanism is available to keep the brain alive if it is so compromised that the autonomous systems that ensure survival are damaged.

this actually arises anyway: many maximally damaged “copies” of you will be out there in the world.

The very last copy made is one in which not a single bit of the scan file was actually read: every bit of data used to make the final copy came from the random file. This final copy has not been affected by the scan file in any way, meaning that it has not been affected by your brain. It has not had any causal interaction with your brain. It is not even as though it was made from the scan file and then corrupted: the scan file was not read at all in making it. Effectively, the final copy is just a random computer model that has been put together with no relation to you. If this is admitted as a “copy” *any* program produced in a similar way, or in fact *any* way, will also be a badly damaged “copy” of you. For example, the same day, someone in another country who does not know you might run a computer program which makes a computer model from a random file, and this would also be a badly damaged “copy” of you. The fact that such a model would have nothing in common with you, and that it would have no causal relationship with you, would be irrelevant: both of these are ignored for the final copy. Furthermore, each copy is effectively just a computer program, and any computer program made by anyone or anything anywhere could also be considered a “copy” of you. Someone in a software company could write an accountancy program and it would be a valid “copy” of you. Things are not even restricted to programs made by computers. We could consider the neurological processes in all kinds of animals as “computer programs” and include them. We could go even further. The idea of a “computer program” is only a particular interpretation of a physical system. Reality is full of natural processes that could be described in algorithmic terms: they could be viewed as computer programs. If we are not to require any causal connection, or any degree of accuracy, a huge number of copies of you could be found all over the universe. This is somewhat similar to John Searle’s idea of finding the *WordStar* computer program in a wall [8] – except that this should be much less controversial. No claim is being made here that sophisticated computer programs are found all over nature due to universal realizability: instead, it is merely being claimed that trivial computer programs doing random things are all over nature.

Reality, then, seems to contain a huge number of copies of you – each with no causal connection to your brain, and each extremely badly damaged in the sense that it cannot be meaningfully said to contain any information from your brain. As the number of such copies will far outweigh the number actually created by the process in the thought experiment, it is almost certain that you should expect to end up as one of these copies. Yet these copies have no meaningful mental activity. This gives an answer to the thought experiment’s question:

If mind uploading does result in actual continuation of you – if software copies of your mind do constitute your future – what is your expectation of what is going to happen to you?

...and the answer is that you should expect to end up dead.

Things can still get worse, however. If you regard the copies as continuations of you, then they are being regarded as just as valid as the continuation that would be provided by your biological brain if it were left alive, but what if your brain is left alive? The copies can hardly stop being continuations just because the original brain is still alive. If there is a possibility that your present can connect to a future in some copy, then this should remain the case whether your biological brain is alive or not. With your biological brain alive, it would not make sense to say that you can no longer end up as one of the copies. Rather, many different futures would branch off from your present – and just one of these would involve your biological brain.⁴ The other futures would involve any copies that exist. However, trivial “copies” corresponding to the maximally damaged version of you – the one produced in the thought experiment without any reference to the scan file – will exist all over the universe – whether the mind uploading process in the thought experiment is occurring or not – so at any time, you should regard your present as branching into many futures, one of which involves your biological brain and a normal continuation of existence and almost all of which involve “finding” yourself as an extremely damaged copy with no resemblance to your

⁴ We have previously argued that a measure approach should be taken with regard to the relative measure of versions of you that exist in some substrate [9], but that is being ignored here.

existing mind and no causal connection to your current situation. Right now, you should expect to find yourself, in the next instant, as a “maximally damaged” copy, somewhere outside your brain, that exists due to some trivial process in nature and that is too damaged to allow conscious thought. You should expect, in the next instant, to die.

3.2.2 Similarity with Quantum Brain Damage

There is a similarity here with the *quantum brain damage* thought experiment, discussed previously [3]. In that thought experiment, assuming that MWI was true, many branches were made to split off from your present, with versions of you with varying amounts of damage. The thought experiment was used to show that it may be simplistic to be sure of your survival in the quantum suicide scenario. The situation here is similar, except there is no reliance on MWI being true. Instead, the different futures are made to split off in the same world through copying.

The situation presented here is actually worse than the situation presented by quantum brain damage. The quantum brain damage scenario merely suggests that it may be simplistic to expect your survival to be guaranteed, but the situation presented here seems to be suggesting that you should actually expect to die – and not just if you perform something like the quantum suicide experiment, but *right now*. The difference between your fates in the two situations is caused by the huge number of maximally damaged copies that happen to exist in reality.

The similarity between the two situations suggests that consideration of one might inform us about the other.

3.2.3 But we do not *really* expect to die right now.

Assuming that any copy is as valid a continuation as any other seems to indicate that you should expect to die right now. In the earlier discussion about quantum brain damage [3], we mentioned this issue, giving the example of finding yourself, in the next instant, in a house plant, existing due to some trivial computation that is inadequate to produce consciousness, so that you die. This example would seem absurd to most people. The general idea that your

continued existence is so unsafe due to the likelihood of finding yourself existing as a maximally damaged copy, will also seem absurd. In everyday life, we expect continuity to work in a “sensible” way, with successive moments of our experience relating to states of our brains at different moments in time. A plausible view of reality cannot involve the idea, apparently suggested by the thought experiment, that all copies, including maximally damaged copies with no causal connection to your present situation, can be equally valid continuations. This leads us into absurd situations like the one that gives this article its title: a possible future for you could be “finding” yourself, in the next instant, in a house plant – your mental processes reduced to something lacking in anything that we associate with human consciousness so that you are effectively dead. This position is telling us nothing about how to avoid situations like this, and it should be viewed as incoherent.

A way of dealing with the thought experiment, and its apparent implication, that prevents this is needed.

4 TRYING TO RESOLVE THE PROBLEM

4.1 We cannot treat all copies equally.

The problem of severely damaged copies is an inevitable result of assuming that any copy must be a valid continuation of you, and that all copies should be treated equally. To resolve the problem, it will be necessary *not* to treat all copies equally, and consideration will now be given to this. A problem that will be shown to arise here is that any approach that can work will be beyond the scope of the simple idea that a copy is a continuation.

4.2 Can we dispose of the most damaged copy as a special case?

It may seem tempting to think that the problem can be resolved just by assuming that the most “damaged” copy in the thought experiment – the final one that was made, with all the information coming from the random file and no information from the scan file – is not a continuation of you. It may seem to be a special case on account of containing nothing from the scan file and having no causal relationship at all with it – meaning that it contains nothing from your brain and has

no causal relationship with you – and it may seem that it is this copy that is causing all the problems by implying that all the other maximally damaged “copies” out there in the world with no information from you and no causal connection with you are also continuations.

This, however, would be naïve. Suppose that we decided that this copy could not be a possible continuation. The *next* most-damaged copy is the one which was made by reading just one bit from the scan file and all the other bits from the random file. This copy is almost as badly damaged as the one that was just eliminated – it only has one bit from the scan file – and it would be trivial enough that we should still expect to see many “copies” with this degree of damage out there in the universe. We could say this for the *next* most-damaged copy – the one made with two bits from the scan file and the rest from the random file, and the next after that: near this end of the damage scale, the amount of information from the scan file is insignificant. Unless these copies can be eliminated too, it seems that you should still expect your imminent death.

Against this, it could be argued that these copies at least have *some* causal connection with you: when they are made, at least a small amount of interaction with your brain occurs, whereas most similarly damaged copies that are “out there” will have no such causal connection – so it may seem that, while you could end up as one of the extremely damaged copies in the thought experiment, the similarly damaged copies that are “out there” in the world are not important and you should not expect to die right now.

A problem with this is that the causal connection between you and the most severely damaged copies is extremely weak and it would be easy, even with current technology, to make a copy like this. All that is needed is the capture of a small number of bits from your brain, followed by the addition of a lot of random data to make the “copy”. (In fact, not even the random data is *really* needed: a version of the thought experiment could be described in which varying amounts of random data are used.) It is possible to create a huge number of copies which are similarly damaged and have been made by a process which involves a similar degree of causal interaction with you – which is to say, hardly any

causal interaction. If all that is protecting you is absence of causal interaction, you should expect to die if this were done – yet it should seem absurd to us that your death could be arranged that easily, by just making a huge number of copies which are only causally derived from you in a trivial way.

Another problem with this is that, even if causal interaction is demanded, nobody even has to perform, explicitly, the kind of trivial copying process just described. Your brain is interacting with the outside world all the time. Matter in your brain is in contact with matter in the rest of your body, which is in contact with the outside world. There will constantly be a huge number of processes going on around you which, while they have practically nothing in common with your thought processes, could be considered to have a very weak causal connection with your brain by being affected by it in some trivial way. More than enough extremely damaged “copies” that *are* causally connected to you could be found to make your imminent death almost certain if all that is required is a causally connected copy.

A final problem with this is that requiring causal interaction with the original in production of a copy seems to be getting away from the simple idea that a copy is a valid continuation: extra demands are now being of a copy. Trying to use causality is actually admitting that there is more to this than the existence of copies – and this suggests that a deeper approach is needed. This is irrelevant anyway: trying to resolve the problem merely by disposing of the most damaged, completely random copy fails.

4.3 Can we dispose of some of the damaged copies?

Disposing of the single, completely random copy in the thought experiment achieves nothing, but might the problem be resolved by disposing of a large number of copies – the most extremely damaged ones? The idea here is that if we can justify the idea that the most severely damaged copies in the thought experiment – the ones derived almost completely from the random file – cannot be continuations, then the possibility of equivalent copies of you existing in nature – meaning that you should expect your death – is removed.

A problem with this is that it seems arbitrary. Where should the line be drawn between copies that are admitted as continuations and copies that are not? Wherever the line was drawn, there would have to be one copy that was admitted as a valid continuation, and another copy, with just one less bit from the scan file and one more bit from the random file, that was not accepted. It should seem implausible that there can be a philosophical justification for treating two copies so differently, when the difference between them is so minor.

If such a line is drawn, *where* should it be drawn? It might be said that the position of the line is simply a metaphysical fact. This should seem implausible. It would mean that we could never know where the line is, because no statement about where it is could ever be justified: there would be nothing to refer to in making the justification. Further, if such a line exists as a metaphysical fact, continuity of consciousness, and therefore consciousness itself, is being treated in a radically different way to anything else in reality: consciousness is having fundamental metaphysical truths asserted about it that are just supposed to be true: this might as well be a claim of Cartesian dualism.

Wherever the line is drawn really needs to be justified. Such a justification might involve consideration of the amount of consciousness (whatever that means) in a copy, the degree of resemblance of the copy to the original or the extent to which there has been causal interaction with your brain in making it – which in this case would mean the indirect interaction with your brain associated with interaction with the scan file that has been derived from your brain. Such an approach would seem to have to take into account the physical details of the situation. It would have to consider the physical substrates on which these minds are made, what is happening on them, and possibly how they causally relate to whatever they are supposed to have been copied from. The idea that a copy is a continuation would merely be a simplification of all this, but when thought of as a profound position it should be viewed as incoherent.

Given that we do not know exactly how such a general approach would work, we cannot be sure about what answers it will give for various situations. We may imagine some scenario in which a mind is copied, and the simplistic

assumption that a copy is a continuation may suggest that it is a valid continuation – but a proper analysis using the more general approach may indicate that, due to some features of this copy or how it is produced, it is not a valid continuation.

4.4 Can we regard some copies as *more likely* futures for you?

The idea that there is a line between copies that are continuations and copies that are not has just been questioned. A general approach that takes into account the physical details of the situation would be needed to draw such a line, but even so, it should be hard to see why there should be such an abrupt change in status from a copy that is a valid continuation to one that is not merely by taking one less bit from the scan file and one more bit from the random file. It may make more sense to say that copies are dealt with in a probabilistic way – that all copies are valid continuations, but that you should view some copies as more likely to be your future than others. The idea here is that the most severely damaged copies, or those with least causal connection to you, would still be valid continuations of you, but they would have extremely low probabilities of being your future assigned to them and you would be extremely unlikely to find yourself in one of the futures represented by these copies.

This deals with the problem of an implausible, abrupt change from continuation to non-continuation when a single bit in a copy is set up in a different way, but – unless we implausibly say that the probabilities are a certain way because it is a metaphysical fact – some way of justifying the probabilities is still needed. As is the case with actual elimination of copies, this will need some general approach that takes into account the physical details of the situation. As before, such an approach would have to consider the physical substrates, and possibly what is happening when copies are made and how they causally relate to whatever they are supposed to have been copied from. Again, the idea that a copy is a continuation would merely be a simplification.

Again, given that we do not know exactly how such a general approach would work, we cannot be sure that it will agree with the simplistic view. In some scenario in which a mind is copied, the

simplistic assumption that a copy is a continuation may suggest that it is a valid continuation, but a proper analysis using the more general approach may indicate that, due to some features of this copy or how it is produced, this copy should be assigned a very low probability of being your future.

Further, if your brain is scanned and then your brain dies, it does not necessarily follow that you will escape death by copies being produced. The physical details of the situation matter, and it may be that the general approach tells us that “occupying” your dead brain should be assigned a relatively high probability of being your future. Without actually knowing how the general approach would work, you should not be sure that it gives you a low chance of ending up in your dead brain and a high chance of ending up as a copy: the physical details of the situation might cause the ending to be less happy.

5 CONCLUSION

Many people think that “mind uploading” – copying a human mind into a computer – could be used as a way of avoiding death – that if your mind were copied into a computer, and your brain then died, your personal survival would have been assured by this. The “copy” is supposed to provide continuity of self.

A thought experiment has been used to show that there is a problem with this idea. The thought experiment, the *mind uploading brain damage* thought experiment, involves creating a large number of copies of you, with a proportion of the information in each copy coming from a scan of your brain and the remainder coming from a random file. The proportion of information from the brain scan is variable, and many different copies are produced, with the proportion of information coming from the brain scan ranging from 100% to 0%. As the proportion of information from the brain scan in a copy decreases, the copy effectively becomes more “damaged”. The problem here is that there is no obvious way of saying which of these copies are possible continuations and which are not. The copy with the most damage – the one with no information at all from your brain scan in it – seems to be a possible future. The situation is worse than this, though: the most damaged copy is equivalent to a huge number of equivalent processes out there that might now be considered

continuations of you, yet none of these processes really involve anything like conscious thought: they would involve continuing you in an extremely damaged way in which your mental existence has effectively ended. If the simple idea that a copy is a continuation is applied, you should worry that you might end up “in” something like a house plant – with the physical events going on in a house plant being the continuation of you, but lacking any of the qualities you associate with consciousness and therefore meaning that you are dead. If this is true, you should expect to die right now, as there are more ways for your mind to be continued in some arbitrary, destructive way than there are for it to be continued in a way that preserves your consciousness. We do not expect this to happen, and to make sense of this we must clearly assume that some copies cannot be valid continuations, or are very unlikely to represent your future. The idea that copies are continuations – that you would survive if a copy is made – is too simplistic to deal with this. There is no answer to be had using this approach, and the idea should be viewed as incoherent. We need to look beyond this simplistic idea for answers.

If we decide to eliminate some copies as valid futures for you, we are assuming an abrupt line between those that can be continuations and those that cannot be, even though the difference between copies on either side of the line is trivial. This should seem implausible. It may make more sense to treat things probabilistically – to assume that all copies are valid continuations, but that different copies have different probabilities of actually being your future, but in either case, we must use some more general approach that tells us which copies to include, or what probabilities to assign.

Such a general approach will need to take into account the physical details of the situation, such as the physical substrate and the physical events that have occurred in a copy’s production and how it causally relates to the original.

There is no guarantee that such a general approach is going to provide the same assurance of survival that you might think is provided if you think that a copy provides continuity. The form that a copy takes and the way in which a copy is produced is likely to be important with regard to the likelihood that it is your future.

Furthermore, if you are about to die, the possibility that your future simply consists of ending up in a dead brain cannot be eliminated: it may be more likely than the probability of ending up in a copy! Without knowing how a general approach would work, we cannot be sure what the probabilities will be – and it is likely that any attempt to use mind uploading for continuity, even if the process is viable in principle, would need to meet certain physical requirements to provide a copy that has a reasonable probability of being your future.

The idea that you should consider your survival through mind uploading as certain should be therefore be treated with some degree of scepticism. While this may seem to be an attack on the concept of mind uploading, it is a mitigated attack. We are not claiming that mind uploading would not work, but rather that the details of the situation – the form the copy takes, the degree of similarity to the original and how it is produced – would be more relevant than people might think to your chance of experiencing continuation through mind uploading.

That said, you may have cause to be pessimistic about your expectation of survival in many common “science fiction” scenarios involving mind uploading and other kinds of duplication. Continuation in a human brain is clearly more direct, in some sense, than continuation through scanning machines, computer files, different machines, etc. The chain of processes linking one state to the next state is shorter, and the physical substrates in consecutive states are particularly similar to each other. This does not mean that mind uploading would not work: for us to say that would be to anticipate what this unknown general approach is going to say, and would go against everything that we have been arguing. However, it should possibly seem that, if this general approach is going to view the connection between consecutive brain states as particularly strong – if it is going to give us the result that a path involving conventional continuity within your biological brain, even one that is dying, is one of the more likely futures for you – then it might have a lot to work with. While it may be possible in principle, statistically “wrenching yourself free” from continuity in your dying brain may be somewhat difficult. One approach that may be considered –

and here things possibly become more morbid – is to work on the problem from both ends – by trying to establish some alternative continuity in a way that will be statistically strong and also “doing something” to your brain to try to weaken continuity through it. Whether this could work, even in principle, is beyond the scope of this article.

The mind uploading brain damage thought experiment described here has some similarity with the quantum brain damage argument discussed previously. The thought experiment here is concerned with continuity in mind uploading, whereas the quantum brain damage thought experiment is about continuity in quantum suicide if the many-worlds interpretation of quantum mechanics is true. The same issue is raised by both thought experiments: that of deciding whether a given “version” in the future, with some level of damage, is really a continuation, and both thought experiments suggest that the position that survival is assured, based on a simple view of continuity, is an incoherent one.

The previous discussion of quantum brain damage was sceptical of the idea that your survival should be viewed as assured in quantum suicide. The discussion given here, may give some *limited* support to the idea that a more general approach *might* indicate that branches in which you survive in quantum suicide might be viewed as more subjectively likely than those in which you do not. If you view all “copies” as equally likely continuations, you end up with the result that your imminent death should be expected. We do not expect this: we expect, generally, to have a future involving more conventional continuity, and it could be the case that the extent to which a future version relates to the past version, and the extent to which it is conscious is relevant. This may open the door to some kind of expectation of survival in quantum suicide, in which your survival is not necessarily assured, but in which the branches in which you survive might be made more likely due to what happens in those branches. Such an effect, if it occurred, would depend on the specific details of the situation, though, and may be very weak. Nevertheless, the argument given here might be taken as giving some limited support for such a position.

REFERENCES

- [1] Strout, J., 2002. *Mind Uploading Home Page*. [Online] Mind Uploading Home Page. Available at: <http://www.ibiblio.org/jstrout/uploading/MUHomePage.html> [Accessed 31 July 2011].
- [2] The Society of Neural Prosthetics and Whole Brain Emulation Science, 2002. *The Society of Neural Prosthetics and Whole Brain Emulation Science*. [Online] minduploading.org. Available at: <http://minduploading.org/> [Accessed: 31 July 2011].
- [3] Almond, P., 2011. *Quantum Brain Damage: Assuming you will survive quantum suicide may be simplistic*. [Online] paul-almond.com. Available at: <http://www.paulalmond.com/QuantumBrainDamage.pdf> or <http://www.paul-almond.com/QuantumBrainDamage.doc> [Accessed 31 July 2011].
- [4] Egan, G., 1994. *Permutation City*. London: Millennium. (Fiction).
- [5] Everett, H., 1957. Relative State Formulation of Quantum Mechanics. *Reviews of Modern Physics*, 29, pp.454-462.
- [6] 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].)
- [7] Searle, J. R., 1980. Minds, brains and computers. *The Behavioral and Brain Sciences*, 3, pp.417-457.
- [8] Searle, J. R., 2002. *The Rediscovery of the Mind*. Cambridge, Massachusetts: The MIT Press. 9th Edition. pp.207-212. (Originally Published: 1992. Cambridge, Massachusetts: The MIT Press.)
- [9] Almond, P., 2011. *Minds, Substrate, Measure and Value – Part 6: Further Justification*. [Online] paul-almond.com. Available at: <http://www.paul-almond.com/Substrate6.pdf> or <http://www.paul-almond.com/Substrate6.doc> [Accessed 31 July 2011].