

# **Minds, Substrate, Measure and Value**

## **Part 1: Substrate Dependence**

By Paul Almond, 12 September 2007

**Website:** [www.paul-almond.com](http://www.paul-almond.com)  
**Email:** [info@paul-almond.com](mailto:info@paul-almond.com)

© Copyright Paul Almond, 2007. All Rights Reserved.

# Minds, Substrate, Measure and Value

## Part 1: Substrate Dependence

By Paul Almond, 12 September 2007

### Abstract

This is the first in a series of articles exploring the relationship between minds and physical systems (substrates) on which they are based. Strong AI (strong artificial intelligence) advocates typically maintain that the substrate is irrelevant, provided that the required computation can be performed on it, and only the computation matters. John Searle, an opponent of strong AI, argues that the substrate does matter and that a mind is not just computation on a substrate but is caused by specific physical processes. Searle states that there is no reason to assume that all substrates that allow general computation can support minds [1,2]. This article will show that the substrate matters, but not in the way that Searle thinks. It influences the probabilities that you are in various situations in some thought experiments in which there is uncertainty about the substrate on which you currently exist. The substrate is *statistically* important and influences the measure of minds associated with computing done on it. How this may relate to expectations of future status and influence ethics, in terms of the value that we assign to a thinking entity, are explored. There is an apparent paradox between the idea of associating measure with minds that emerges from the thought experiment in this article and our experience of seeing single instances of intelligent entities in the real world. This does not mean that we should avoid the issues raised by the thought experiment, but we need to consider how minds are associated with substrates to resolve it.

The second article will show how a many-worlds type view, arrived at by considering what the existence of objects means, resolves this apparent paradox and how the concept of measure can be meaningful when applied to a mind associated with a substrate.

The third article will discuss some implications of the arguments made previously.

The fourth article will consider the implications for strong AI and Searle's argument against it in particular. Searle's case relies on the nature of a substrate being important regarding whether or not it can support a mind and on the irrelevance of the substrate in strong AI. It will be shown that admitting *statistical* relevance of the substrate, or its effect on the measure of minds, does not discard so much of the idea of irrelevance of substrate that strong AI fails, but partially giving up irrelevance of substrate like this does not help Searle's argument, while it does, in fact, allow strong AI to be more clearly expressed in a way that allows Searle's arguments to be more easily dealt with. Strong AI's claim that different substrates can support minds is generally correct, although what is meant by "computation" needs clarification and the statistical importance of substrate should be acknowledged.

## Introduction

A widespread view is that human minds exist due to computation in brains. The brain is a physical system providing a substrate on which the mind is based. Computation can occur on different substrates; for example, electronic or mechanical computers.

Some proponents of a philosophical position known as *strong AI* think that the substrate is irrelevant and, provided that the correct computation is occurring, a mind exists. This article will consider the importance of the substrate supporting a mind. A thought experiment will show that the substrate is *not completely irrelevant*, but has some *statistical* importance. The nature of a substrate influences the chance that you are based on it if you are uncertain about your status and, by implication, the *measure* of minds based on it.

## Strong AI and Substrate Independence

The term *strong AI* (strong artificial intelligence) was given by John Searle to the proposition that computers can be conscious and he described it as follows:

“...according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations.” [2]

The strong AI case would seem to imply that:

- consciousness can be viewed as existing if symbols are being manipulated in the right way by a physical device.
- consciousness can be associated with the behaviour of a system. If a system is acting in the right way, as determined by observations of its external behaviour, then it can be viewed as being conscious, or as one scientist casually put it, “If it walks like a duck and quacks like a duck, it *is* a duck.”
- if a system were conscious then a properly constructed model of that system would also be conscious, irrespective of how it was physically realized. Computers can be built from semiconductors or wooden rods and string. The strong AI case would say that this choice of substrate is philosophically irrelevant.

Searle gives the term *weak AI* (weak artificial intelligence) to the position that computers are capable of modelling the behaviour of conscious entities. Weak AI differs from strong AI in that it does not regard such computers as necessarily conscious merely by virtue of them appearing to behave in the same way, to an external observer, as systems that *are* conscious.

Strong AI asserts that the nature of the substrate is irrelevant to the issue of whether a mind exists on it or not, provided that the necessary symbol manipulation can be done on it. From this, an obvious implication, although not one necessarily stated as part of strong AI, is that the nature of the substrate has no philosophical relevance whatsoever. The thought experiment in this article will show that the substrate has *statistical* relevance. Although this may require the strong AI position to be revised slightly, this sort of relevance is too weak to help Searle's case that strong AI is wrong.

## Ideas Behind the Thought Experiment

The thought experiment will use two main ideas: mind uploading and uncertainty about your current situation.

Mind uploading [3,4] is the idea of using a hypothetical, very accurate brain scanning procedure to construct a digital representation of a human brain in a computer. This is then used to run a simulated model of the brain to make a digital "copy" of the person who was "uploaded". Mind uploading will form the basis of the thought experiment, but this does not mean that it must be considered practically plausible: it is merely being used as a philosophical device. Nor need we worry about whether a mind uploaded copy of you would be a continuation of your identity: the thought experiment assumes that you already exist as a mind uploaded entity. The issue of continuation of identity is tentatively explored later in the article. This article is not really about mind uploading, which is only being used to create a scenario in which you are unsure about your current situation. The type of mind uploading imagined to be occurring in this article requires you to be asleep while it happens, to make things convenient for me.

The uncertainty about your situation which will be used in the thought experiment is uncertainty about the substrate on which you currently exist. That is to say, you are unsure about the nature of the physical system which is causing your mental experiences. Most people probably are quite sure about the nature of their own substrate: it is their brain, if we ignore Nick Bostrom's simulation argument [5,6], to which this series of articles has some relevance, incidentally. It is possible, however, to imagine situations in which you are less sure. Greg Egan's novel *Permutation City* [7] has situations in which characters are unsure. At the book's start a character is unsure whether he is an organic person or a "copy" made by mind uploading running on a computer in a virtual reality. The thought experiment will use the same sort of idea, with a scenario in which you could be in a number of different situations, existing on any one of a number of different substrates.

## The Thought Experiment

Imagine finding yourself in the following situation:

You wake and see a window floating in front of you containing the text "You are one of the copies." Your last memory is of being in a reclining chair, about to go to sleep to undergo a brain scanning procedure which was to be used for "mind uploading" – making

a computer simulation of your brain. Three identical simulations of your brain were to be run on three different computers: A, B and C. A, B and C run identical software and start in the same state. Each of the uploaded copies was going to find itself waking in a virtual reality with a window apparently floating in front of it containing the text “You are one of the copies.” You know that you must be one of the copies: you cannot be the original, organic person. When this is happening to you, none of the simulations in A, B and C, after starting off in the same state, has yet interacted with the outside world. Although such interaction cannot be ruled out in the future, it does not matter now.

You are not told on which computer A, B or C you are actually running. You know that, whichever it is, two more versions of you are running on two other computers, each equally uncertain about his/her own status. As far as you know, there is still the original, organic version of you in the real world, unless something has happened to him/her but this person is irrelevant here.

How likely is it that you are in Computer A, B or C? You have no way of differentiating between them, so the probability is  $\frac{1}{3}$  in each case.

We now cosmetically change the computers by removing the exterior casing from computers A and B, enclosing them both in a single, larger case, so that from outside it appears to be a single computer. Internally, nothing has changed: this new box simply contains two computers, A and B. Although we appear to have two computers in the box, they are both doing the same processing and we really have a single computer with some redundancy, so we will call this new box “Computer AB”.

Computer C is unaltered: it remains in its original case and will continue to be called “Computer C”. The probability that you were in Computer C was originally  $\frac{1}{3}$ . We have merely altered some exterior casing and it is unreasonable to expect this to affect probability.

So, the probability you are in Computer C is still  $\frac{1}{3}$ .

You will be in Computer AB if you are in what was originally called Computer A or what was originally called Computer B, which are now inside the box called “Computer AB”. The probability is  $\frac{1}{3}$  for each of these possibilities.

Probability you are in Computer AB = probability you are in Computer A + probability you are in Computer B

$$= \frac{1}{3} + \frac{1}{3}$$

$$= \frac{2}{3}$$

We have two computers with different internal workings and:

Probability you are in Computer AB =  $\frac{2}{3}$

Probability you are in Computer C =  $\frac{1}{3}$

*The probabilities are different for these two computers.*

If you object to Computer AB being regarded as a single computer, saying that it is *two* computers labelled as a single computer in a contrived way, then imagine that the computers are mechanical and made of metal components. Imagine sliding the two computers together, closer and closer, so that they overlap, until equivalent parts are adjacent and can be welded together. Alternatively, imagine two electrical computers of some kind which are moved together to make a single computer with thicker wiring, or two electrical computers with lots of gaps between atoms allowing us to merge the machines together by overlapping them so that they fill each other's gaps. Even though it may be practically implausible we should be able to imagine, philosophically, various merging processes like this, each producing what can only sensibly be considered a single machine. Would it be reasonable to say that the probabilities for computers A, B and C, or the combined probabilities for any group of such computers, must change appreciably during such a merging process? Is this supposed to happen magically in the instant when equivalent components touch each other?

If we get different probabilities like this for computers working in essentially the same way, just by altering the amount of redundancy or merging components together, it follows that more extreme changes to the substrate, such as completely altering the design of the machine or its working principles, can also cause a change in the probabilities.

Therefore:

*If you are in one of a number of different situations which involve you existing on different substrates, then the nature of each substrate affects the probability that you are running on it. The substrate is not irrelevant, but has statistical significance.*

## **Substrate and Redundancy as a Placeholder**

In the thought experiment, and similar ones that we may consider, the feature of a system that makes it particularly likely that you are inhabiting it is *redundancy*. Computer AB is a substrate that provides more redundancy than Computer C because it can be used to run two versions, each equivalent to Computer C. It is not just “how many computers you have in the box”, though. When the individual computers inside the case of Computer AB are moved closer together, being ultimately combined, it may not be immediately obvious that redundancy still exists, but it does in one obvious sense: Computer AB is made from twice as much matter as Computer C, making it equivalent to two versions of Computer C which have been merged. The redundancy, therefore, relates to inefficient use of

matter. Computer C will also have its own redundancy, but when we are considering the probabilities it is the *relative* redundancies of Computer AB and Computer C that matter.

Many aspects of a computer's design relate to its redundancy. In any computer, duplicates of components would mean more redundancy. In an electrical computer thicker wires would seem to involve more redundancy, as would larger gears and levers in a mechanical computer. If two computers based on the same principle, such as two mechanical computers or two electrical computers, can have different amounts of redundancy, then when the operating principles are different there is even less reason to expect the redundancy to be the same. We should not expect the redundancy provided by a given electronic computer to be the same as that provided by a given mechanical computer and we should not expect the redundancy provided by the human brain to be the same as that provided by various other types of system.

*In scenarios like the one in the thought experiment, substrates which work on different principles could have very different probabilities.*

The argument is not claiming that redundancy is the only characteristic of a substrate that determines probabilities like this. Redundancy merely tells us how we should assign probabilities *all else being equal*. Situations in which you might exist could have other features which can be shown by different philosophical arguments to affect the probabilities. This “all else being equal” qualification applies throughout this article.

Some readers may have issues with this “redundancy” concept because it is hard to formally define, and they would have a good point, but this is not the end of the matter. “Redundancy” is a cruder version of a more sophisticated explanation of what is going on that will be given in the next article. This more sophisticated idea is based on numbers of different ways of algorithmically extracting patterns or meanings. For now, however, “redundancy” is a useful “placeholder” idea.

## **Considering Organic Brains**

The thought experiment and argument just given relates to organic brains as much as anything else. In the scenario just given you knew that you were the uploaded copy, but what if you did not know? Imagine this thought experiment:

You wake in a chair in a brain scanning room. Your last memory is of being about to go to sleep for a scan. You had arranged that, after the scan was made, a number of uploaded copies of you would be run on various computers. Each of these copies would be presented with a virtual reality simulation of waking in the brain scanning chair with no immediate clues from the environment that it is no longer the original, organic version. There is no way of knowing, just by looking around, whether you are the original, organic person waking in the chair after the scan or one of the uploaded copies in a virtual reality simulation of waking in the chair. Maybe, of course, the simulated environment has limits. If you are simulated and try to leave the room, the building or whatever city you are supposed to be in, then maybe you will find it impossible, for

example (as happens to a character in *Permutation City* [8], but this need not concern us: right now, you do not know. How likely is it that you are the original, organic person?

If each version of you is as statistically important as any other then the answer is simple: there is one organic version of you and two computer simulations, so the probability that you are the organic version is therefore  $\frac{1}{3}$  and the probability that you are an uploaded copy is  $\frac{2}{3}$ . The argument previously given about substrate importance, however, throws doubt onto this. If the contents of the computer cases were previously important in the thought experiment, then they are important now. We cannot just count versions. Instead, we need to look at the substrate in each case. We need to consider the substrate for the organic version, which means considering the physical nature of the brain itself, and we need to consider the substrate in the case of each of the computers running uploaded versions. The more redundancy in a particular implementation of a version you, then the more likely it is that that implementation is causing your experiences.

## Expectation of Future Status

An argument like this about your current status could possibly be extended into a tentative argument about expectation of *future* status.

As before, imagine that a scan will be made of your brain and used to make a number of copies, each of which will be presented with a virtual reality simulation of waking in the brain scanning chair with no immediate clues from the environment that it is no longer the original, organic version.

On waking in the chair you will be unsure whether you are the original, organic version or one of the copies. You will only be able to assign a probability to each possible situation and, from the previous argument about substrate and probability, the types of substrate will affect this probability.

If you can determine the probabilities when apparently waking in the chair, just from considering different substrates, you can as easily work out the probabilities that you will later assign to different situations on waking *before the brain scanning occurs*.

As an example, let us imagine that, before the brain scan, you decide that when you wake in the chair, with no sensory information available to resolve your possible situations apart, you will conclude that there is a 60% chance that you are the original, organic version. This means that right now, as an organic person, you think that in the future you will think that there is a 40% chance that you are the copy. Maybe, if you trust your reasoning to be valid after the mind uploading procedure, you should assume the same reasoning to be valid now when you use it to determine your expected future status? Maybe you should assume, now, that after the mind uploading procedure there is a 60% chance that you will find you are the original, organic version and a 40% chance that you will find that you are the copy?

We do have the issue, here, that we are assuming that you do not have any experiential way of resolving the situations apart. What if such experience were available? Maybe, for example, you would expect to be in a different room if you were the uploaded copy? Either this is a reasonable argument about expectation of future status or it is not. If it is reasonable then it is implausible that your expectation of future status should vary depending on how many clues you are given from the environment. Saying that you allocate a certain probability to being the copy or the original accepts the possibility that you can find out later which you are if some experiential information becomes available. Even if experiential information starts to become available immediately (that is to say, there is no attempt to fool the copy), it will take you time to acquire and process sufficient information. On the other hand, if you have thought about issues like substrate difference in advance you may already know what the probabilities will be of finding yourself in different situations. No matter what experiential evidence is available, or whether or not it is available immediately, there will be at least an instant when you have no observational evidence on which to base any assessment of your situation and in which you will have to base your consideration on the issue of substrate discussed here and, possibly, other issues.

When multiple copies of you are going to exist, one of which could be your original, biological brain, a case can be made that you should view this as involving multiple futures and you should assign yourself a probability of finding yourself in any one of these futures. That probability would be determined by ignoring any sensory evidence that could be presented to you in this in these situations, assuming that you would have no experiential way of telling the situations apart, and considering how your knowledge about the substrate in each case would affect the probability you assigned yourself of being in each possible situation after the copying process.

This may seem a strange view. It would mean, for example, that if you are going to have a copy of yourself made from a brain scan then, from your point of view, before the scan is made, there is a chance that you will become the copy after it is made and activated. Readers may be sceptical about this, but it is hard to find a firm reason against it. Both the copy and the future state of your brain are effectively patterns propagated into the future by some causal sequence of events linking your current and future states. What I would be sceptical about is the idea that substrate is the only issue when considering your chances of finding yourself in any future given situation. No assumption should be made, for example, that the original, organic version and any simulated copies automatically have the same status until we start considering substrate differences. It is possible that what has happened in the past of a particular version, whether it has just relied on the biological brain continuing to exist or on various processes associated with mind uploading, *does* matter. The sort of consideration of substrate that we have been discussing would tell us how to assign probabilities *all else being equal* and other factors could play a part.

Objections can be made against the idea that all that matters is future uncertainty. One objection involves constructing a thought experiment in which we know *now* that one of

the situations that we will consider possible in the future is actually impossible. This is the scenario:

You have volunteered to go and live in the Mars colony for the rest of your life and in a moment you will go to sleep in a hibernation chamber on a spaceship. When you wake up you will have arrived at Mars. You know that they have trouble getting enough volunteers to go to Mars and they get round this by manufacturing colonists in unused hibernation chambers on the spaceships en route to Mars, using advanced technology that can actually build a human being, atom by atom, from raw materials. The practical plausibility of this should not be an issue, but if you need an idea of how the manufacturing process might work, imagine molecular nanotechnology [9,10,11,12,13], as proposed by K. Eric Drexler, being used. Fake memories are implanted into the manufactured colonists as part of the manufacturing process: the brains are built with neuron wiring patterns containing memories. These memories are of a made-up life on Earth prior to volunteering to go to Mars, right up to the point of going to sleep for the journey. On arrival at Mars the colonists who actually lived on Earth and volunteered (whom we will call “real” colonists) and the colonists who were manufactured en route and only think they lived on Earth (whom we will call “fake” colonists) are all woken to colonize the planet. For every “real” colonist, one “fake” colonist is manufactured en route. There is no attempt to hide this process from the inhabitants of Mars. They are all told that half of them never existed prior to the journey and proof is made available that this is actually the case. We will not worry about what form this proof takes, but it is available on Mars. The colonists are not told, however, which of them are “real” colonists and which are “fake” colonists and no communication with Earth is permitted, so colonists cannot find out, for example, if people remember them on Earth. You know all this while waiting to go to sleep for the journey.

On waking at Mars you will be a member of a population of people, half of which are “real” colonists and half of which are “fake” colonists. You will not know which, however. In the absence of anything else on which to base your judgement, when you are on Mars you should think that there is a  $\frac{1}{2}$  chance that you are a “real” colonist with real memories of life on Earth and a  $\frac{1}{2}$  chance that you are a “fake” colonist with fake memories of life prior to waking at Mars. What makes this situation strange is that right now, before going to sleep, you know that you are one of the “real” colonists because, right now, you are actually experiencing life prior to waking up at Mars – experience that manufactured colonists never really have. You also know that this knowledge will not help you when you arrive at Mars. Even though you know yourself to be a “real” colonist now, you also know that when you wake at Mars you will not know if your current experience of being sure is a fake memory and that you will still need to assign yourself a  $\frac{1}{2}$  chance of being a “real” colonist.

If knowledge of future probability estimates always indicates future expectation then the knowledge that in the future you will assign yourself a  $\frac{1}{2}$  probability of being a “real” colonist and a  $\frac{1}{2}$  probability of being a “manufactured” colonist indicates that in the future there is a  $\frac{1}{2}$  chance that you will be a “real” colonist and a  $\frac{1}{2}$  chance that you will be a “manufactured” colonist, even though you now know that you a “real” colonist. This

does not seem to make sense: you can hardly change from a “real” colonist into a “fake” colonist. Where does this leave us? Substrate in different possible future situations was related to the idea that future uncertainty is significant in determining the probability of being in future situations, but this objection seems to throw doubt on this. My own view is that the objection may not mean that uncertainty is not worth considering, but rather than we need to be cautious about it and if we do use expected future probability values to indicate future expectation we should not expect this to apply in all situations.

We could make another, stranger interpretation of the Mars colony thought experiment. Instead of considering the memories of the “fake” colonists as being really fake we might decide that these memories are of real experiences associated with whatever computational process was used to generate the fake memories. In this view, there would be no such thing as truly “fake” memories because all memories would have to be generated by some process and your “fake” memories would really be of your experience of that process. If true, this would mean that, in the thought experiment, while waiting to go to sleep in the hibernation chamber, you could actually not be sure that you are a “real” colonist: you could be a fake colonist whose current experience is based on the substrate of the computer used to generate your “fake” memories or on the process of constructing your neurons with all the wiring to contain them, or even on the processes that occur later in your brain during recall of the memories. A feeling of this sort of idea is provided in Alastair Reynolds’s novel *Revelation Space* [14] in which an implanted memory is described by narrative, as if the events were actually experienced by the character, though I am not sure if any philosophical suggestion is being made by this or if it is just an interesting, or convenient, literary device.

## Ethics and Value

It could be argued that the probability of finding yourself in various situations like this is linked to the value that should be placed on things. As an example, consider the following thought experiment:

You are a simulated copy running in a virtual reality on one of two computers: Computer A and Computer B. Both computers are running identical versions of you and you are given technical descriptions of both computers, but you do not know whether you are in Computer A or Computer B. You receive information from the “real world” that both computers are at risk of attack by the *HellSim V99.2* computer virus and you can assume that the other version of you has received similar information. This nasty piece of software “infects” computer systems running uploaded minds in virtual realities and twists the virtual reality simulation to cause great suffering to any inhabitants – and that means you. The danger to each computer can be reduced by spending money on it, and the more money that is spent on a given computer, the lower is its risk of viral infection. You have a limited amount of money to spend in the real world and someone in the real world will follow your instructions about how to spend it to buy safety for each computer. This person will not tell you which computer is running you. You need to decide how much money to allocate to reducing the risk of viral infection for each computer.

You could spend the same on each computer, but that could be simplistic. If the two computers differ in construction – if there is a difference in the possible substrates – then the previous argument suggests that you could be more likely to be in one computer than the other and it may be rational to spend more on reducing the risk to that computer. The “all else being equal” qualification applies. Substrate may not be the only issue – there could be others – but it is an issue as it affects probability. If you would spend different amounts of money on protecting different computers then you are assigning different *value* to the different computers. When considering computers that could be running you, value appears to be related to substrate.

If you would apply this reasoning to computers running copies of you, why should you not apply it to other entities? This suggests that if there is a computer running a thinking entity then the value placed on that computer should not be based just on the entity that it is running but on the physical nature of the computer itself. All else being equal, it seems that the greater the degree of redundancy in the substrate, the greater the value that we should assign to that substrate. Different computers running the same kind of entity, with the same behaviour, might be assigned different value, and different levels of rights, due to different physical construction. This is different to the view that many strong AI advocates would have – that the nature of the substrate has nothing to do with the rights that a machine should have.

I need to be clear on how this should be interpreted. I am not saying that intelligent computers should not have rights: as I think that human thought processes are computational it would not be sensible for me to suggest that. Nor am I saying that computers should automatically have less value, or fewer rights, than humans. The physical nature of a system is simply one characteristic that may play a part in assigning value: other features, such as the nature of the entity being simulated, may be far more important in determining the relative value of different systems in many situations. I am also not suggesting that we need to deal practically with ethics in this way in everyday situations. Even if philosophical arguments can be made about the nature of a substrate and value, we may decide not to make fine judgements in all instances, instead choosing to assign the same value to all systems of a certain general type.

## **The Apparent Paradox**

A paradox may seem to result from the conclusions reached so far. It has been argued that the type of substrate on which you could exist in different situations could influence your probabilities of being in these situations if you cannot resolve them apart and that this can influence the value that we assign to physical systems “running” different thinking entities. This would seem to need multiple versions of you and other thinking entities to justify it, but this is not what we see when we observe systems externally.

As an example, returning to the thought experiment from the start of the article, you can be in one of three systems, A, B or C, and the probability of each is  $\frac{1}{3}$ . That is consistent with what an external observer would see: he/she would see three boxes and each box, if analyzed closely, would appear to be simulating a single version of you. If there are three

versions of you, each in a different situation, then, all else being equal, any one of these versions should assume a  $\frac{1}{3}$  chance of being in each situation. When we put Computer A and Computer B in the same box, making Computer AB, things could still be easily seen to make sense. There is a  $\frac{2}{3}$  probability that you are in Computer AB and a  $\frac{1}{3}$  probability that you are in Computer C. The  $\frac{2}{3}$  probability for Computer AB may seem strange to an external observer, who only sees one computer, but on looking inside Computer AB he/she can see that it is made of 2 computers, A and B, each running a version of you, and this could account for the  $\frac{2}{3}$  probability. Things get difficult, however, with two computers that have different probabilities merely by having different substrates, as when AB is made by merging A and B together more extremely. We could have a situation where there is a  $\frac{2}{3}$  probability that you are in Computer AB and a  $\frac{1}{3}$  probability that you are in Computer C, but when you look inside AB and C you do not see any “multiple versions”, but merely a different substrate to that provided by Computer C. This substrate may be equivalent to multiple versions, but it does not change the fact that when you look at Computer AB or Computer C you do not see them: you see a box behaving in a particular way.

This is the *apparent* paradox: when we are considering these systems from “inside”, when they are possible candidates for our situation, each system seems to correspond in a sense to many possible situations for us, yet to an outside observer there is just the system exhibiting whatever behaviour it is exhibiting.

This issue is most obvious when we mix value with it. Suppose we have some thinking entity running on two computers, A and B, which are both under some kind of threat. We can spend resources on trying to protect Computer A, Computer B or both. The entity being simulated in each case appears scared and loud pleas to be protected are coming from loudspeakers on both computers. What if the substrate provided by Computer A supported the entity with much more redundancy than that provided by Computer B? According to the argument given so far, we should give Computer A greater value than Computer B, and we should be prepared to spend more of our resources on trying protect Computer A, because Computer A and Computer B can be considered equivalent to multiple computers running versions of the thinking entity and the greater redundancy provided by Computer A makes it equivalent to a computer running *more* versions. On looking inside Computer A’s case, however, we do not see all these separate computers laid out in front of us, doing redundant processing: all we see is an architecture that is less efficient, in one sense, than Computer B’s. For example, we might see that Computer A has thicker wires, or uses a different technology entirely to Computer B. The collection of computers to which A is equivalent appears in some sense to be a philosophical device, introduced to deal with this issue of what happens when we move two computers progressively closer together, yet now it is forcing our ethical decisions. One fact is obvious: despite all our argument about probability, substrate and value, we do not hear more pleas to be protected from Computer A.

Observation and conventional intuition would suggest that there is only one entity in each computer. In fact, some people would say that the entity “in” each computer is just an abstracted way of describing the computer itself, yet the argument about substrate seems

to suggest that when an entity runs on some substrate there is really a number of entities there – that it has some kind of *measure* dependent on the nature of the substrate itself. How do we resolve this apparent paradox?

It may seem tempting to discard the earlier argument about substrate as an aberration or something only meaningful in some abstract sense, but I will not do that. We can get meaningful results for three separate computers. We can get meaningful results when we enclose two computers in the case. There is no reason for things abruptly to become an aberration, or abstract, when we start to move two of the computers closer together, ultimately merging them. The substrate argument will not go away.

We have a situation where the *measure* of a mind in a substrate seems to mean something. With entities getting multiplied like this we are being pushed in the direction of many-worlds cosmologies. We need to think about what it means to say that minds, or any other objects, “exist”. This is a matter for the next article.

## Conclusion

A thought experiment has shown that that the substrate on which a mind exists is not completely irrelevant. When you could be in any one of a number of different situations, on different substrates, the nature of each substrate influences the probability of you being on it. While the substrate may not have a *qualitative* effect on whether or not a mind can be supported by it, it does, therefore, have a *statistical* effect when considering substrates that could be supporting you. It follows that this does not just apply to you, but also to others and that an entity existing on a substrate needs regarding, somehow, as multiple entities existing on that substrate, the number of entities, or “measure”, depending on the nature of the substrate.

This has possible ethical implications as it could influence the relative value that we place on systems supporting thinking entities. It could also have implications for our future expectations in various situations.

The thought experiment indicates that we need to consider minds and substrates in terms of multiple minds or measure, but our experience of observing thinking entities is that we just see one thinking system: we do not have any experience of this measure other than inferring it from a thought experiment. This apparent paradox will be resolved in the next article in this series, using the many-worlds type view implied by the thought experiment.

The third article will discuss the implications of this. The fourth will discuss the implications for strong AI and Searle’s argument, which is invalid, in particular.

## References

[1] Searle, J. R. (1997). *The Mystery of Consciousness*. New York: The New York Review of Books.

- [2] Searle, J. R. (1980). Minds, brains and computers. *The Behavioral and Brain Sciences* 3:417-457.
- [3] Web Reference: Strout, J. *Mind Uploading Home Page*. (2002). Retrieved 22 June 2003 from <http://www.ibiblio.org/jstrout/uploading/MUHomePage.html>.
- [4] Web Reference: *Mind Uploading Research Group*. (2002). Retrieved 22 June 2003 from <http://minduploading.org/>.
- [5] Bostrom, N. (2003). Are you living in a computer simulation? *Philosophical Quarterly*, 2003, Vol. 53, No. 211, pp 243-255. (Bostrom circulated a draft of this paper in 2001).
- [6] Web Reference: Bostrom, N. (2003). *Are you living in a computer simulation?* (An online version of the article in reference [5]) Retrieved 8 September 2007 from <http://www.simulation-argument.com/simulation.html>. (Further information about this subject by Bostrom and others is at <http://www.simulation-argument.com>.)
- [8] Egan, G. (1994). *Permutation City*. London: Millennium. (Fiction).
- [9] Drexler, K.E. (1986). *Engines of Creation*. New York: Anchor Books.
- [10] Drexler, K.E., Peterson, C., Pergamit, G. (2000). *Unbounding the Future: the Nanotechnology Revolution*. New York: William Morrow.
- [11] Drexler, K.E. (1992). *Nanosystems*. New York: John Wiley and Sons Inc.
- [12] *The Foresight Institute*. (1986-2007). Retrieved June 22, 2003 from <http://www.foresight.org/>.
- [13] Web Reference: Zyvex. (n.d.) Retrieved June 22, 2003 from <http://www.zyvex.com>.
- [14] Reynolds, A. (2000). *Revelation Space*. London: Gollancz. (Fiction). Chapters 11-12.