Ethics of brain emulations

Anders Sandberg

Future of Humanity Institute, Oxford University, Oxford, UK

Anders.sandberg@philosophy.ox.ac.uk

The Future of Humanity Institute, Faculty of Philosophy, University of Oxford, Suite 8, Littlegate House, 16/17 St Ebbe's Street, Oxford, OX1 1PT, UK

Ethics of brain emulations

Whole brain emulation attempts to achieve software intelligence by copying the function of biological nervous systems into software. This paper aims at giving an overview of the ethical issues of the brain emulation approach, and analyse how they should affect responsible policy for developing the field. Animal emulations have uncertain moral status, and a principle of analogy is proposed for judging treatment of virtual animals. Various considerations of developing and using human brain emulations are discussed.

Keywords: brain emulation; animal ethics; moral status; moral uncertainty; machine consciousness; computational neuroscience

Introduction

Whole brain emulation (WBE) is an approach to achieve software intelligence by copying the functional structure of biological nervous systems into software. Rather than attempting to understand the high-level processes underlying perception, action, emotions and intelligence, the approach assumes that they would emerge from a sufficiently close imitation of the low-level neural functions, even if this is done through a software process. (Merkle 1989, Sandberg & Bostrom 2008)

While the philosophy (Chalmers 2010), impact (Hanson 2008) and feasibility (Sandberg 2013) of brain emulations have been discussed, little analysis of the ethics of the project so far has been done. The main questions of this paper are to what extent brain emulations are moral patients, and what new ethical concerns are introduced as a result of brain emulation technology.

Brain emulation

The basic idea is to take a particular brain, scan its structure in detail at some resolution, construct a software model of the physiology that is so faithful to the original that, when run on appropriate hardware, it will have an internal causal structure that is essentially

the same as the original brain. All relevant functions on some level of description are present, and higher level functions supervene from these.

While at present an unfeasibly ambitious challenge, the necessary computing power and various scanning methods are rapidly developing. Large scale computational brain models are a very active research area, at present reaching the size of mammalian nervous systems. (Markram 2006, Djurfeldt et al. 2008, Eliasmith et al. 2012, Preissl et al. 2012) WBE can be viewed as the logical endpoint of current trends in computational neuroscience and systems biology. Obviously the eventual feasibility depends on a number of philosophical issues (physicalism, functionalism, non-organicism) and empirical facts (computability, scale separation, detectability, scanning and simulation tractability) that cannot be predicted beforehand; WBE can be viewed as a program trying to test them empirically. (Sandberg 2013)

Early projects are likely to merge data from multiple brains and studies, attempting to show that this can produce a sufficiently rich model to produce nontrivial behaviour but not attempting to emulate any particular individual. However, it is not clear that this can be carried on indefinitely: higher mammalian brains are organized and simultaneously individualized through experience, and linking parts of different brains is unlikely to produce functional behaviour. This means that the focus is likely to move to developing a "pipeline" from brain to executable model, where ideally an individually learned behaviour of the original animal is demonstrated by the resulting emulation.

Although WBE focuses on the brain, a realistic project will likely have to include a fairly complex body model in order to allow the emulated nervous system to interact with a simulated or real world, as well as the physiological feedback loops that influence neural activity.

At present the only known methods able to generate complete data at cellular and subcellular resolution are destructive, making the scanned brain non-viable. For a number of reasons it is unlikely that non-destructive methods will be developed any time soon (Sandberg & Bostrom 2008, appendix E).

In the following I will assume that WBE is doable, or at least doesn't suffer enough roadblocks to preclude attempting it, in order to examine the ethics of pursuing the project.

Virtual lab animals

The aim of brain emulation is to create systems that closely imitate real biological organisms in terms of behaviour and internal causal structure. While the ultimate ambitions may be grand, there are many practical uses of intermediate realistic organism simulations. In particular, emulations of animals could be used instead of real animals for experiments in education, science, medicine or engineering. Opponents of animal testing often argue that much of it is excessive and could be replaced with simulations. While the current situation is debatable, in a future where brain emulations are possible it would seem that this would be true: by definition emulations would produce the same kind of results as real animals.

However, there are three problems:

- Brain emulations might require significant use of test animals to develop the technology.
- Detecting that something is a perfect emulation might be impossible.
- An emulation might hold the same moral weight as a real animal by being sentient or a being with inherent value.

Need for experiments

Developing brain emulation is going to require the use of animals. They would be necessary not only for direct scanning into emulations, but in various experiments gathering the necessary understanding of neuroscience, testing scanning modalities and comparing the real and simulated animals. In order to achieve a useful simulation we need to understand at least one relevant level of the real system well enough to recreate it, otherwise the simulation will not produce correct data.

What kind of lab animals would be suitable for research in brain emulation and how would they be used? At present neuroscientists use nearly all model species, from nematode worms to primates. Typically there are few restrictions on research on invertebrates (with the exception of cephalopods). While early attempts are likely to aim at simple, well defined nervous systems like the nematode *Caenorhabditis elegans*, *Lymnaea Stagnalis* (British pond snail) or *Drosophila melanogaster* (fruit fly), much of the neuroscience and tool development will likely involve standard vertebrate lab animals such as mice, either for in vitro experiments with tissue pieces or in vivo experiments attempting to map neural function to properties that can be detected. The nervous system of invertebrates also differ in many ways from the mammalian nervous system; while they might make good test benches for small emulations it is likely that the research will tend to move towards small mammals, hoping that successes there can be scaled up to larger brains and bodies. The final stages in animal brain emulation before moving on to human emulation would likely involve primates, raising the strongest animal protection issues. In theory this stage might be avoidable if the scaling up from smaller animal brains towards humans seems smooth enough, but this would put a greater risk on the human test subjects.

Most "slice and dice" scanning (where the brain is removed, fixated and then analysed) avoids normal animal experimentation concerns since there is no experiment done on the living animal itself, just tissue extraction. This is essentially terminal anaesthesia ("unclassified" in UK classification of suffering). The only issue here is the pre-scanning treatment, whether there is any harm to the animal in its life coming to an end, and whether *in silico* suffering possible.

However, developing brain emulation techniques will likely also involve experiments on living animals, including testing whether an *in vivo* preparation behaves like an *in vitro* and an *in silico* model. This will necessitate using behaving animals in ways that could cause suffering. The amount of such research needed is at present hard to estimate. If the non-organicism assumption of WBE is correct, most data gathering and analysis will deal with low-level systems such as neuron physiology and connectivity rather than the whole organism; if all levels are needed, then the fundamental feasibility of WBE is cast into question (Sandberg 2013).

What can we learn from emulations?

The second problem is equivalent to the current issue of how well animal models map onto human conditions, or more generally how much models and simulations in science reflect anything about reality.

The aim is achieving structural validity (Zeigler, 1985, Zeigler, Praehofer, & Kim, 2000), that the emulation reflects how the real system operates. Unfortunately this might be impossible to prove: there could exist hidden properties that only very rarely come into play that are not represented. Even defining meaningful and observable measures of success is nontrivial when dealing with higher order systems (Sandberg 2013). Developing methods and criteria for validating neuroscience models is one of the key requirements for WBE.

One of the peculiar things about the brain emulation program is that unlike many scientific projects the aim is not directly full understanding of the system that is being simulated. Rather, the simulation is used as a verification of our low-level understanding of neural systems and is intended as a useful tool. Once successful, emulations become very powerful tools for further investigations (or valuable in themselves). Before that stage the emulation does not contribute much knowledge about the full system. This might be seen as an argument against undertaking the WBE project: the cost and animals used are not outweighed by returns in the form of useful scientific knowledge. However, sometimes very risky projects are worth doing because they promise very large eventual returns (consider the Panama Canal) or might have unexpected but significant spin-offs (consider the Human Genome Project). Where the balance lies depends both on how the evaluations are made and the degree of long-term ambition.

What is the moral status of an emulation?

The question what moral consideration we should give to animals lies at the core of the debate about animal experimentation ethics. We can pose a similar question about what moral claims emulations have on us. Can they be wronged? Can they suffer?

Indirect theories argue that animals do not merit moral consideration, but the effect of human actions on them does matter. The classic example is Kantian theories, where animals lack moral autonomy and hence are not beings whose interests morally count. Our duties towards them are merely indirect duties towards humanity. Being cruel to animals harms our own humanity:

> "Our duties towards animals are merely indirect duties towards humanity. Animal nature has analogies to human nature, and by doing our duties to animals in respect of manifestations of human nature, we indirectly do our duty to humanity…. We

> can judge the heart of a man by his treatment of animals."
> (Regan and Singer, 1989: 23-24)

By this kind of indirect account the nature of the emulation does not matter: if it is cruel to pinch the tail of biological mice the same cruel impulse is present in pinching the simulated tail of an emulated mouse. It is like damaging an effigy: it is the intention behind doing damage that is morally bad, not the damage. Conversely, treating emulations well might be like treating dolls well: it might not be morally obligatory but its compassionate.

A different take on animal moral considerability come from social contract or feminist ethics, arguing against the individualist bias they perceive in the other theories. What matters is not intrinsic properties but the social relations we have with animals.

> "Moral considerability is not an intrinsic property of any creature, nor is it supervenient on only its intrinsic properties, such as its capacities. It depends, deeply, on the kind of relations they can have with us"
> (Anderson 2004)

If we have the same kind of relations to an emulated animal as a biological animal, they should presumably be treated similarly. Since successful emulations (by assumption) also have the same capacity to form reciprocal relations, this seems likely.

Another large set of theories argue that the interests of animals do count morally due to intrinsic properties. Typically they are based on the sentience of animals giving them moral status: experiences of pleasure or suffering are morally relevant states no matter what system experiences them. Whether animals are sentient or not is usually estimated from the Argument from Analogy, which supports claims of consciousness by looking at similarities between animals and human beings. Species membership is not a relevant factor. These theories differ on whether human interests still can trump animal interests or whether animals actually have the same moral status as human beings. For

the present purpose the important question is whether software emulations can have sentience, consciousness or the other properties these theories ground moral status on.

Animal rights can be argued on other grounds than sentience, such as animals having beliefs, desires and self-consciousness of their own and hence having inherent value and rights as subjects of a life that has inherent value. (Regan 1983) Successfully emulated animals would presumably behave in similar ways: the virtual mouse will avoid virtual pain; the isolated social animal will behave in a lonely fashion. Whether the mere behaviour of loneliness or pain-avoidance is an indication of a real moral interest even when we doubt it is associated with any inner experience is problematic: most accounts of moral patienthood take experience as fundamental, because that actually ties the state of affairs to a value, the welfare of something. But theories of value that ascribe value to non-agents can of course allow non-conscious software as a moral patient (for example, having value by virtue of its unique complexity).

To my knowledge nobody has yet voiced concern that existing computational neuroscience simulations could have aversive experiences. In fact, the assumption that simulations do not have phenomenal consciousness is often used to motivate such research:

> "Secondly, one of the more obvious features of mathematical modelling is that it is not invasive, and hence could be of great advantage in the study of chronic pain. There are major ethical problems with the experimental study of chronic pain in humans and animals. It is possible to use mathematical modelling to test some of the neurochemical and neurophysiological features of chronic pain without the use of methods which would be ethically prohibitive in the laboratory or clinic. Stembach has observed "Before inflicting pain on humans, can mathematical or statistical modelling provide answers to the questions being considered?" (p262) (53). We claim that mathematical modelling has the potential to add something unique to the armamentarium of the pain researcher."
> (Britton & Skevington 1996)

To some degree this view is natural because typical computational simulations contain just a handful of neurons. It is unlikely that so small systems could suffer[1] . However, the largest simulations have reached millions or even billions of neurons: we are reaching the numbers found in brains of small vertebrates that people do find morally relevant. The lack of meaningful internal structure in the network probably prevents any experience from occurring, but this is merely a conjecture.

Whether machines can be built to have consciousness or phenomenological states has been debated for a long time, often as a version of the strong AI hypothesis. At one extreme it has been suggested that even thermostats have simple conscious states (Chalmers 1996), making phenomenal states independent of higher level functions, while opponents of strong AI have commonly denied the possibility of any machine (or

[1] However, note the Small Network Argument (Herzog et al. 2007): "… for each model of consciousness there exists a minimal model, i.e., a small neural network, that fulfills the respective criteria, but to which one would not like to assign consciousness". Mere size of the model is not a solution: there is little reason to think that 10^{11} randomly connected neurons are conscious, and appeals to the right kind of complexity of interconnectedness runs into the argument again. One way out is to argue that fine-grained consciousness requires at least mid-sized systems: small networks only have rudimentary conscious contents (Taylor, 2007). Another one is to bite the bullet and accept, if not panpsychism, that consciousness might exist in exceedingly simple systems.

Assigning even a small probability to the possibility of suffering or moral importance to simple systems leads to far bigger consequences than just making neuroscience simulations suspect. The total number of insects in the world is so great that if they matter morally even to a tiny degree, their interests would likely overshadow humanity's interests. This is by no means a *reductio ad absurdum* of the idea: it could be that we are very seriously wrong about what truly matters in the world.

at least software) mental states. See (Gamez 2008) for a review of some current directions in machine consciousness.

It is worth noting that there are cognitive scientists who produce computational models they consider able to have consciousness (as per their own theories)[2].

Consider the case of Rodney Cotterill's CyberChild, a simulated infant controlled by a biologically inspired neural network and with a simulated body. (Cotterill 2003) Within the network different neuron populations corresponding to brain areas such as cerebellum, brainstem nuclei, motor cortices, sensory cortex, hippocampus and amygdala are connected according to an idealized mammalian brain architecture with learning, attention and efference copy signals. The body model has some simulated muscles and states such as levels of blood glucose, milk in the stomach, and urine in the bladder. If the glucose level drops too much it "expires". The simulated voice and motions allow it to interact with a user, trying to survive by getting enough milk. Leaving aside the extremely small neural network (20 neurons per area) it is an ambitious project. This simulation does attempt to implement a model of consciousness, and the originator was hopeful that there was no fundamental reason why consciousness could not ultimately develop in it.

However, were the CyberChild conscious, it would have a very impoverished existence. It would exist in a world of mainly visual perception, except for visceral inputs, 'pain' from full nappies, and hunger. Its only means of communication is crying and the only possible response is the appearance (or not) of a bottle that has to be manoeuvred to the mouth. Even if the perceptions did not have any aversive content there would be no prospect of growth or change.

---

[2] See for example the contributions in the theme issue of *Neural Networks* Volume 20, Issue 9, November 2007.

This is eerily similar to Metzinger's warning (Metzinger 2003, p. 621):

> "What would you say if someone came along and said, "Hey, we want to genetically engineer mentally retarded human infants! For reasons of scientific progress we need infants with certain cognitive and emotional deficits in order to study their postnatal psychological development—we urgently need some funding for this important and innovative kind of research!" You would certainly think this was not only an absurd and appalling but also a dangerous idea. It would hopefully not pass any ethics committee in the democratic world. However, what today's ethics committees *don't* see is how the first machines satisfying a minimally sufficient set of constraints for conscious experience could be just *like* such mentally retarded infants. They would suffer from all kinds of functional and representational deficits too. But they would now also subjectively experience those deficits. In addition, they would have no political lobby—no representatives in *any* ethics committee."

He goes on arguing that we should ban all attempts to create or even risk the creation artificial systems that have phenomenological self-models. While views on what the particular criterion for being able to suffer is might differ between different thinkers, it is clear that the potential for suffering software should be a normative concern. However, as discussed in mainstream animal rights ethics, other interests (such as human interests) can sometimes be strong enough to allow animal suffering. Presumably such interests (if these accounts of ethics are correct) would also allow for creating suffering software.

David Gamez (2005) suggests a probability scale for machine phenomenology, based on the intuition that machines built along the same lines as human beings are more likely to have conscious states than other kinds of machines. This scale aims to quantify how likely a machine is to be *ascribed* to be able to exhibit such states (and to some extent, address Metzinger's ethical concerns without stifling research). In the case

of WBE, the strong isomorphism with animal brains gives a fairly high score[3]. Arrabales, Ledezma, and Sanchis (2010) on the other hand suggests a scale for the estimation of the potential degree of consciousness based on architectural and behavioural features of an agent; again a successful or even partial WBE implementation of an animal would by definition score highly (with a score dependent on species). The actual validity and utility of such scales can be debated, but insofar they formalize intuitions about the Argument from Analogy about potential mental content they show that WBE at least has significant *apparent* potential of being a system that has states that might make it a moral patient. WBE is different from entirely artificial software in that it deliberately tries to be as similar as possible to morally considerable biological systems, and this should make us more ethically cautious than with other software.

Much to the point of this section, Dennett has argued that creating a machine able to feel pain is nontrivial, to a large extent in the incoherencies in our ordinary concept of pain. (Dennett 1978) However, he is not against the possibility in principle:

> "If and when a good physiological sub-personal theory of pain is developed, a robot could in principle be constructed to instantiate it. Such advances in science would probably bring in their train wide-scale changes in what we found intuitive about pain, so that the charge that our robot only suffered what we artificially called pain would lose its persuasiveness. In the meantime (if there were a cultural

[3] For an electrophysiological WBE model the factors are FW1, FM1, FN4, AD3, with rate, size and time slicing possibly ranging over the whole range. This produces a weighting ranging between 10^{-5} to 0.01, giving an ordinal ranking 170-39 out of 812. The highest weighting beats the neural controlled animat of DeMarse et al., a system containing real biological neurons controlling a robot.

> lag) thoughtful people would refrain from kicking such a robot."
> (Dennett 1978 p. 449)

From the eliminative materialist perspective we should hence be cautious about ascribing or not ascribing suffering to software, since we do not (yet) have a good understanding of what suffering is (or rather, what the actual underlying component that is morally relevant, is). In particular, successful WBE might indeed represent a physiological sub-personal theory of pain, but it might be as opaque to outside observers as real physiological pain.

The fact that at present there does not seem to be any idea of how to solve the hard problem of consciousness or how to detect phenomenal states seem to push us in the direction of suspending judgement:

> "Second, there are the arguments of Moor (1988) and Prinz (2003), who suggest that it may be indeterminable whether a machine is conscious or not. This could force us to acknowledge the possibility of consciousness in a machine, even if we cannot tell for certain whether this is the case by solving the hard problem of consciousness." (Gamez 2008)

While the problem of animal experience and status is contentious, the problem of emulated experience and status will by definition be even more contentious. Intuitions are likely to strongly diverge and there might not be any empirical observations that could settle the differences.

The principle of assuming the most

What to do in a situation of moral uncertainty about the status of emulations?[4] It seems that a safe strategy would be to make the most cautious assumption:

> **Principle of Assuming the Most (PAM)**: Assume that any emulated system could have the same mental properties as the original system and treat it correspondingly.

The mice should be treated the same in the real laboratory as in the virtual. It is better to treat a simulacrum as the real thing than to mistreat a sentient being. This has the advantage that many of the ethical principles, regulations and guidance in animal testing can be carried over directly to the pursuit of brain emulation.

This has some similarity to the Principle of Substrate Non-Discrimination ("If two beings have the same functionality and the same conscious experience, and differ only in the substrate of their implementation, then they have the same moral status.") (Bostrom & Yudkowsky 2011) but does not assume that the conscious experience is identical. On the other hand, if one were to reject the principle of substrate non-discrimination on some grounds, then it seems that one could also reject PAM since one does have a clear theory of what systems have moral status. However, this seems to be a presumptuous move given the uncertainty of the question.

Note that once the principle is applied, it makes sense to investigate in what ways the assumptions can be sharpened. If there are reasons to think that certain mental

---

[4] Strictly speaking, we are in a situation of moral uncertainty about what ethical system we ought to follow in general, and factual uncertainty about the experiential status of emulations. But being sure about one and not the other one still leads to a problematic moral choice. Given the divergent views of experts on both questions we should also not be overly confident about our ability to be certain in these matters.

properties are *not* present, they overrule the principle in that case. An emulated mouse that does not respond to sensory stimuli is clearly different from a normal mouse. It is also relevant to compare to the right system. For example, the CyberChild, despite its suggestive appearance, is not an emulation of a human infant but at most an etiolated subset of neurons in a generic mammalian nervous system.

It might be argued that this principle is too extreme, that it forecloses much of the useful pain research discussed by (Britton & Skevington, 1996). However, it is agnostic on whether there exist overruling human interests. That is left for the ethical theory of the user to determine, for example using cost-benefit methods. Also, as discussed below, it might be quite possible to investigate pain systems without phenomenal consciousness.

Ameliorating virtual suffering

PAM implies that unless there is evidence to the contrary, we should treat emulated animals with the same care as the original animal. This means in most cases that practices that would be impermissible in the physical lab are impermissible in the virtual lab.

Conversely, counterparts to practices that reduce suffering such as analgesic practices should be developed for use on emulated systems. Many of the best practices discussed in (Schofield 2002) can be readily implemented: brain emulation technology by definition allows parameters of the emulation can be changed to produce the same functional effect as the drugs have in the real nervous system. In addition, pain systems can in theory be perfectly controlled in emulations (for example by inhibiting their output), producing "perfect painkillers". However, this is all based on the assumption that we understand what is involved in the experience of pain: if there are undiscovered systems of suffering careless research can produce undetected distress.

It is also possible to run only part of an emulation, for example leaving out or blocking nociceptors, the spinal or central pain systems, or systems related to consciousness. This could be done more exactly (and reversibly) than in biological animals. Emulations can also be run for very brief spans of time, not allowing any time for a subjective experience. But organisms are organic wholes with densely interacting parts: just like in real animal ethics there will no doubt exist situations where experiments hinge upon the whole behaving organism, including its aversive experiences.

It is likely that early scans, models and simulations will often be flawed. Flawed scans would be equivalent to animals with local or global brain damage. Flawed models would introduce systemic distortions, ranging from the state of not having any brain to abnormal brain states. Flawed simulations (broken off because of software crashes) would correspond to premature death (possibly repeated, with no memory – see below). Viewed in analogy with animals it seems that the main worry should be flawed models producing hard-to-detect suffering.

Just like in animal research it is possible to develop best practices. We can approximate enough of the inner life of animals from empirical observations to make some inferences; the same process is in principle possible with emulations to detect problems peculiar to their state. In fact, the transparency of an emulation to data-gathering makes it easier to detect certain things like activation of pain systems or behavioural withdrawal, and backtrack their source.

Quality of life

An increasing emphasis is placed not just on lack of suffering among lab animals but on adequate quality of life. What constitutes adequate is itself a research issue. In the case of emulations the problem is that quality of life presumably requires both an adequate

body, and an adequate environment for the simulated body to exist in.

The VR world of an emulated nematode or snail is likely going to be very simple and crude even compared to their normal petri dish or aquarium, but the creatures are unlikely to consider that bad. But as we move up among the mammals, we will get to organisms that have a quality of life. A crude VR world might suffice for testing the methods, but would it be acceptable to keep a mouse, cat or monkey in an environment that is too bare for any extended time? Worse, can we know in what ways it is too bare? We have no way of estimating the importance rats place on smells, and whether the smell in the virtual cage are rich enough to be adequate. The intricacy of body simulations also matters: how realistic does a fur have to feel to simulated touch to be adequate?

I estimate that the computational demands of running a very realistic environment are possible to meet and not terribly costly compared to the basic simulation (Sandberg & Bostrom 2008, p. 76-78). However, modelling the *right* aspects requires a sensitive understanding of the lifeworlds of animals we might simply be unable to reliably meet. However, besides the ethical reasons to pursue this understanding there is also a practical need: it is unlikely emulations can be properly validated unless they are placed in realistic environments.

Euthanasia

Most regulations of animal testing see suffering as the central issue, and hence euthanasia as a way of reducing it. Some critics of animal experimentation however argue that an animal life holds intrinsic value, and ending it is wrong.

In the emulation case strange things can happen, since it is possible (due to the multiple realizability of software) to create multiple instances of the same emulation and to terminate them at different times. If the end of the identifiable life of an instance

is a wrong, then it might be possible to produce large number of wrongs by repeatedly running and deleting instances of an emulation even if the experiences during the run are neutral or identical.

Would it matter if the emulation was just run for a millisecond of subjective time? During this time there would not be enough time for any information transmission across the emulated brain, so presumably there could not be any subjective experience. Accounts of value of life built upon being a subject of a life would likely find this unproblematic: the brief emulations do not have a time to be subjects, the only loss might be to the original emulation if this form of future is against its interests.

Conversely, what about running an emulation for a certain time, making a backup copy of its state, and then deleting the running emulation only to have it replaced by the backup? In this case there would be a break in continuity of the emulation that is only observable on the outside, and a loss of experience that would depend on the interval between the backup and the replacement. It seems unclear that anything is lost if the interval is very short. Regan argues that the harm of death is a function of the opportunities of satisfaction it forecloses (Regan 1983); in this case it seems that it forecloses the opportunities envisioned by the instance while it is running, but it is balanced by whatever satisfaction can be achieved during that time.

Most concepts of the harmfulness of death deal with the irreversible and identity-changing aspects of the cessation of life. Typically, any reversible harm will be lesser than an irreversible harm. Since emulation makes several of the potential harms of death (suffering while dying, stopping experience, bodily destruction, changes of identity, cessation of existence) completely or partially reversible it actually reduces the sting of death.

In situations where there is a choice between the irreversible death of a biological being and an emulation counterpart, the PAM suggests we ought to play it safe: they might be morally equivalent. The fact that we might legitimately doubt whether the emulation is a moral patient doesn't mean it has a value intermediate between the biological being and nothing, but rather that the actual value is *either* full or none, we just do not know which. If the case is the conversion of the biological being into an emulation we are making a gamble that we are not destroying something of value (under the usual constraints in animal research of overriding interests, or perhaps human autonomy in the case of a human volunteer).

However, the reversibility of many forms of emulation death may make it cheaper. In a lifeboat case (Regan, 1983), should we sacrifice the software? If it can be restored from backup the real loss will be just the lost memories since last backup and possibly some freedom. Death forecloses fewer opportunities to emulations.

It might of course be argued that the problem is not ending emulations, but the fundamental lack of respect for a being. This is very similar to human dignity arguments, where humans are assumed to have intrinsic dignity that can never be removed, yet it can be gravely disrespected. The emulated mouse might not notice anything wrong, but we know it is treated in a disrespectful way.

There is a generally accepted view that animal life should not be taken wantonly. However, emulations might weaken this: it is easy and painless to end an emulation, and it might be restored with equal ease with no apparent harm done. If more animals are needed, they can be instantiated up to the limits set by available hardware. Could emulations hence lead to a reduction of the value of emulated life? Slippery slope arguments are rarely compelling; the relevant issues rather seem to be that the harm of death has been reduced and that animals have become (economically) cheap. The moral

value does not hinge on these factors but on the earlier discussed properties. That does not mean we should ignore risks of motivated cognition changing our moral views, but the problem lies in complacent moral practice rather than emulation.

Conclusion

Developing animal emulations would be a long-running, widely distributed project that would require significant animal use. This is not different from other major neuroscience undertakings. It might help achieve replacement and reduction in the long run, but could introduce a new morally relevant category of sentient software. Due to the uncertainty about this category I suggest a cautious approach: it should be treated as the corresponding animal system absent countervailing evidence. While this would impose some restrictions on modelling practice, these are not too onerous, especially given the possibility of better-than-real analgesia. However, questions of how to demonstrate scientific validity, quality of life and appropriate treatment of emulated animals over their "lifespan" remain.

Human emulations

Brain emulation of humans raise a host of extra ethical issues or sharpen the problems of proper animal experimentation.

Moral status

The question of moral status is easier to handle in the case of human emulations than in the animal case since they can report back about their state. If a person who is sceptical of brain emulations being conscious or having free will is emulated and, after due introspection and consideration, changes their mind, then that would seem to be some evidence in favour of emulations actually having an inner life. It would actually not prove anything stronger than that the processes where a person changes their mind are

correctly emulated and that there would be some disconfirming evidence in the emulation. It could still be a lacking consciousness and be a functional philosophical zombie (assuming this concept is even coherent).

If philosophical zombies existed, it seems likely that they would be regarded persons (at least in the social sense). They would behave like persons, they would vote, they would complain and demand human rights if mistreated, and in most scenarios there would not be any way of distinguishing the zombies from the humans. Hence, if emulations of human brains work well enough to exhibit human-like behaviour rather than mere human-like neuroscience, legal personhood is likely to eventually follow, despite misgivings of sceptical philosophers[5].

Volunteers and emulation rights

An obvious question is volunteer selection. Is it possible to give informed consent to brain emulation? The most likely scanning methods are going to be destructive, meaning that they end the biological life of the volunteer or would be applied to post-mortem brains.

In the first case, given the uncertainty about the mental state of software there is no way of guaranteeing that there will anything "after", even if the scanning and

---

[5] Francis Fukuyama, for example, argues that emulations would lack consciousness or true emotions, and hence lack moral standing. It would hence be morally acceptable to turn them off at will. (Fukyama, 2002, p.167-170) In the light of his larger argument about creeping threats to human dignity, he would presumably see working human WBE as an insidious threat to dignity by reducing us to mere computation. However, exactly what factor to base dignity claims on is if in anything more contested than what to base moral status on; see for example (Bostrom, 2009) for a very different take on the concept.

emulation are successful (and of course the issues of personal identity and continuity). Hence volunteering while alive is essentially equivalent to assisted suicide with an unknown probability of "failure". It is unlikely that this will be legal on its own for quite some time even in liberal jurisdictions: suicide is increasingly accepted as a way of escaping pain, but suicide for science is not regarded as an acceptable reason[6]. Some terminal patients might yet argue that they wish to use this particular form of "suicide" rather than a guaranteed death, and would seem to have autonomy on their side. An analogy can be made to the use of experimental therapies by the terminally ill, where concerns about harm must be weighed against uncertainty about the therapy, and where the vulnerability of the patient makes them exploitable (Falit & Gross 2008).

In the second case, post-mortem brain scanning, the legal and ethical situation appears easier. There is no legal or ethical person in existence, just the preferences of a past person and the rules for handling anatomical donations. However, this also means that a successful brain emulation based on a person would exist in a legal limbo. Current views would hold it to be a possession of whatever institution performed the experiment rather than a person[7].

---

[6] The Nuremberg code states: "No experiment should be conducted, where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects." But self-experimentation is unlikely to make high risk studies that would otherwise be unethical ethical. Some experiments may produce so lasting harm that they cannot be justified for any social value of the research (Miller and Rosenstein, 2008).

[7] People involved in attempts at preserving legally dead but hopefully recoverable patients are trying to promote recognition of some rights of stored individuals. See for instance the Bill of Brain Preservation Rights (Brain Preservation Foundation 2013). Specifically, it argues

Presumably a sufficiently successful human brain emulation (especially if it followed a series of increasingly plausible animal emulations) would be able to convince society that it was a thinking, feeling being with moral agency and hence entitled to various rights. The PAM would support this: even if one were sceptical of whether the being was "real", the moral risk of not treating a potential moral agent well would be worse than the risk of treating non-moral agents better than needed. Whether this would be convincing enough to have the order of death nullified and the emulation regarded as the same *legal* person as the donor is another matter, as is issues of property ownership.

The risk of ending up a non-person and possibly being used against one's will for someone's purposes, ending up in a brain damaged state, or ending up in an alien future, might not deter volunteers. It certainly doesn't deter people signing contract for cryonic preservation today, although they are fully aware that they will be stored as non-person anatomical donations and might be revived in a future with divergent moral and social views. Given that the alternative is certain death, it appears to be a rational choice for many.

---

that persons in storage should be afforded similar rights to living humans in temporally unconscious states. This includes ensuring quality medical treatment and long term storage, but also revival rights ("The revival wishes of the individual undergoing brain preservation should be respected, when technically feasible. This includes the right to partial revival (memory donation instead of identity or self-awareness revival), and the right to refuse revival under a list of circumstances provided by the individual before preservation.") and legal rights allowing stored persons to retain some monetary or other assets in trust form that could be retrieved upon successful revival. This bill of rights would seem suggests similar right for stored stored emulations.

Handling of flawed, distressed versions

While this problem is troublesome for experimental animals, it becomes worse for attempted human emulations. The reason is that unless the emulation is so damaged that it cannot be said to be a mind with any rights, the process might produce distressed minds that are rightsholders yet have existences not worth living, or lack the capacity to form or express their wishes. For example, they could exist in analogues to persistent vegetative states, dementia, schizophrenia, aphasia, or have on-going very aversive experience. Many of these ethical problems are identical to current cases in medical ethics.

One view would be that if we are ethically forbidden from pulling the plug of a counterpart biological human, we are forbidden from doing the same to the emulation. This might lead to a situation where we have a large number of emulation "patients" requiring significant resources, yet not contributing anything to refining the technology nor having any realistic chance of a "cure".

However, brain emulation allows a separation of cessation of experience from permanent death. A running emulation can be stopped and its state stored, for possible future reinstantiation. This leads to a situation where at least the aversive or meaningless experience is stopped (and computational resources freed up), but which poses questions about the rights of the now frozen emulations to eventual revival. What if they were left on a shelf forever, without ever restarting? That would be the same as if they had been deleted. But do they in that case have a right to be run at least occasionally, despite lacking any benefit from the experience?

Obviously methods of detecting distress and agreed on criteria for termination and storage will have to be developed well in advance of human brain emulation, likely based on existing precedents in medicine, law and ethics.

Persons might write advance directives about the treatment of their emulations. This appears equivalent to normal advance directives, although the reversibility of local termination makes pulling the plug less problematic. It is less clear how to handle wishes to have more subtly deranged instances terminated. While a person might not wish to have a version of themselves with a personality disorder become their successor, at the point where the emulation comes into being it will potentially be a moral subject with a right to its life, and might regard its changed personality as the right one.

Identity

Personal identity is likely going to be a major issue, both because of the transition from an original unproblematic human identity to successor identity/identities that might or might not be the same, and because software minds can potentially have multiple realisability. The discussion about how personal identity relates to successor identities on different substrates is already extensive, and will be foregone here. See for instance (Chalmers, 2010).

Instantiating multiple copies of an emulation and running them as separate computations is obviously as feasible as running a single one. If they have different inputs (or simulated neuronal noise) they will over time diverge into different persons, who have not just a shared past but at least initially very similar outlooks and mental states.

Obviously multiple copies of the same original person pose intriguing legal challenges. For example, contract law would need to be updated to handle contracts where one of the parties is copied – does the contract now apply to both? What about marriages? Are all copies descended from a person legally culpable of past deeds occurring before the copying? To what extent does the privileged understanding copies

have of each other affect their suitability as witnesses against each other? How should votes be allocated if copying is relatively cheap and persons can do "ballot box stuffing" with copies? Do copies start out with equal shares of the original's property? If so, what about inactive backup copies? And so on. These issues are entertaining to speculate upon and will no doubt lead to major legal, social and political changes if they become relevant.

From an ethical standpoint, if all instances are moral agents, then the key question is how obligations, rights and other properties carry over from originals to copies and whether the existence of copies change some of these. For example, making a promise "I will do X" is typically meant to signify that the future instance of the person making the promise will do X. If there are two instances it might be enough that one of them does X (while promising *not* to do X might morally oblige both instances to abstain). But this assumes the future instances acknowledges the person doing the promise as their past self – a perhaps reasonable assumption, but one which could be called into question if there is an identity affecting transition to brain emulation in between (or any other radical change in self-identity).

Would it be moral to voluntarily undergo very painful and/or lethal experiments given that at the end that suffering copy would be deleted and replaced with a backup made just after making the (voluntarily and fully informed) decision to participate? It seems that current views on scientific self-experimentation do not allow such behaviour on the grounds that there are certain things it is never acceptable to do for science. It might be regarded as a combination of the excessive suffering argument (there are suffering so great that no possible advance in knowledge can outweigh its evil) and a human dignity argument (it would be a practice that degrade the dignity of humans). However, the views on what constitutes unacceptable suffering and risk has changed

historically and is not consistent across domains. Performance artists sometimes perform acts that would be clearly disallowed as scientific acts, yet the benefit of their art is entirely subjective (Goodall, 1999). It might be that as the technology becomes available boundaries will be adjusted to reflect updated estimates of what is excessive or lacks dignity, just as we have done in many other areas (e.g. reproductive medicine, transplant medicine).

Time and communication

Emulations will presumably have experience and behave on a timescale set by the speed of their software. The speed a simulation is run relative to the outside world can be changed, depending on available hardware and software. Current large-scale neural simulations are commonly run with slowdown factors on the order of a thousand, but there does not seem to be any reason precluding emulations running faster than realtime biological brains[8].

Nick Bostrom and Eliezer Yudkowsky have argued for a Principle of Subjective Rate of Time: "In cases where the duration of an experience is of basic normative significance, it is the experience's subjective duration that counts." (Bostrom & Yudkowsky 2011). By this account frozen states does not count at all. Conversely, very fast emulations can rapidly produce a large amount of positive or negative value if they are in extreme states: they might count for more in utilitarian calculations.

[8] Axons typically have conduction speeds between 1-100 m/s, producing delays between a few and a hundred milliseconds in the brain. Neurons fire at less than 100 Hz. Modern CPUs are many orders of magnitude faster (in the gigaHerz range) and transmit signals at least 10% of the speed of light. A millionfold speed increase does not seem implausible.

Does human emulation have a right to real-time? Being run at a far faster or slower rate does not matter as long as an emulation is only interacting with a virtual world and other emulations updating at the same speed. But when interacting with the outside world, speed matters. A divergent clockspeed would make communication with people troublesome or impossible. Participation in social activities and meaningful relationships depend on interaction and might be made impossible if they speed past faster than the emulation can handle. A very fast emulation would be isolated from the outside world by lightspeed lags and from biological humans by their glacial slowness. It hence seems that insofar emulated persons are to enjoy human rights (which typically hinge on interactions with other persons and institutions) they need to have access to real-time interaction, or at least "disability support" if they cannot run fast enough (for example very early emulations with speed limited by available computer power).

By the same token, this may mean emulated humans have a right to contact with the world outside their simulation. As Nozick's experience machine demonstrates, most people seem to want to interact with the "real world", although that might just mean the shared social reality of meaningful activity rather than the outside physical world. At the very least emulated people would need some "I/O rights" for communication within their community. But since the virtual world is contingent upon the physical world and asymmetrically affected by it, restricting access only to the virtual is not enough if the emulated people are to be equal citizens of their wider society.

Vulnerability

Brain emulations are extremely vulnerable by default: the software and data constituting them and their mental states can be erased or changed by anybody with access to the system on which they are running. Their bodies are not self-contained and their survival dependent upon hardware they might not have causal control over. They

can also be subjected to undetectable violations such as illicit copying. From an emulation perspective software security is identical to personal security.

Emulations also have a problematic privacy situation, since not only their behaviour can be perfectly documented by the very system they are running on, but also their complete brain states are open for inspection. Whether that information can be interpreted in a meaningful way depends on future advances in cognitive neuroscience, but it is not unreasonable to think that by the time human emulations exist many neural correlates of private mental states will be known. This would put them in a precarious situation.

These considerations suggest that the ethical way of handling brain emulations would be to require strict privacy protection of the emulations and that the emulated persons had legal protection or ownership of the hardware on which they are running, since it is in a sense their physical bodies. Some technological solutions such as encrypted simulation or tamper-resistant special purpose hardware might help. How this can be squared with actual technological praxis (for example, running emulations as distributed processes on rented computers in the cloud) and economic considerations (what if an emulation ran out of funds to pay for its upkeep?) remains to be seen.

Self-Ownership

Even if emulations are given personhood they might still find the ownership of parts of themselves to be complicated. It is not obvious that an emulation can claim to own the brain scan that produced it: it was made at a point in time where the person did not legally exist. The process might also produce valuable intellectual property, for example useful neural networks that can be integrated in non-emulation software to solve problems, in which case the matter of who has a right to the property and proceeds from it emerge.

This is not just an academic question: ownership is often important for developing technologies. Investors want to have returns on their investment, innovators want to retain control over their innovations. This was apparently what partially motivated the ruling in *Moore v. Regents of the University of California* that a patient did not have property rights to cells extracted from his body and turned into lucrative products. (Gold, 1998).This might produce pressures that work against eventual self-ownership for the brain emulations.

Conversely essential subsystems of the emulation software or hardware may be licenced or outright owned by other parties. Does right to life trump or self-ownership property ownership? At least in the case of the first emulations it is unlikely they would have been able to sign any legal contracts, and they might have a claim. However, the owners might still want fair compensation. Would it be acceptable for owners of computing facilities to slow down or freeze non-paying emulations? It seems that the exact answer depends on how emulation self-ownership is framed ethically and legally.

Global issues and existential risk

The preliminary work that has been done on the economics and social impact of brain emulation suggest they could be a massively disruptive force. In particular, simple economic models predict that copyable human capital produces explosive economic growth and (emulated) population increase but also wages decreasing towards Malthusian levels. (Hanson 1994, 2008). Economies that can harness emulation technology well might have a huge strategic advantage over latecomers.

WBE could introduce numerous destabilizing effects, such as increasing inequality, groups that become marginalized, disruption of existing social power relationships and the creation of opportunities to establish new kinds of power, the creation of situations in which the scope of human rights and property rights are poorly

defined and subject to dispute, and particularly strong triggers for racist and xenophobic prejudices, or vigorous religious objections. While all these factors are speculative and depend on details of the world and WBE emergence scenario, they are a cause for concern.

An often underappreciated problem is existential risk: the risk that humanity and all Earth-derived life goes extinct (or suffers a global, permanent reduction in potential or experience) (Bostrom 2002). Ethical analysis of the issue shows that reducing existential risk tends to take strong precedence over many other considerations (Bostrom, 2003, 2013). Brain emulations have a problematic role in this regard. On one hand they might lead to various dystopian scenarios, on the other hand they might enable some very good outcomes.

As discussed above, the lead-up to human brain emulation might be very turbulent because of arms races between different groups pursuing this potentially strategic technology, fearful other groups would reach the goal ahead of them. This might continue after the breakthrough, now in the form of wild economic or other competition. Although the technology itself is not doing much, the sheer scope of what it *could* do leads to potential war.

It could also be that competitive pressures or social drift in a society with brain emulation leads to outcomes where value is lost. For example, wage competition between copyable minds may drive wages down to Malthusian levels, produce beings only optimized for work, spending all available resources on replication, or gradual improvements in emulation efficiency lose axiologically valuable traits (Bostrom 2004). If emulations are zombies humanity, tempted by cybernetic immortality, may gradually trade away its consciousness. These may be evolutionary attractors that may prove

inescapable without central coordination: each step towards the negative outcome is individually rational.

On the other hand, there at least four major ways emulations might lower the risks of Earth-originating intelligence going extinct:

First, the existence of nonbiological humans would ensure at least partial protection from some threats: there is no biological pandemic that can wipe out software. Of course, it is easy to imagine a digital disaster, for example an outbreak of computer viruses that wipe out the brain emulations. But that threat would not affect the biological humans. By splitting the human species into two the joint risks are significantly reduced. Clearly threats to the shared essential infrastructure remain, but the new system is more resilient.

Second, brain emulations are ideally suited for colonizing space and many other environments where biological humans require extensive life support. Avoiding carrying all eggs in one planetary basket is an obvious strategy for strongly reducing existential risk. Besides existing in a substrate-independent manner where they could be run on computers hardened for local conditions, emulations could be transmitted digitally across interplanetary distances. One of the largest obstacles of space colonisation is the enormous cost in time, energy and reaction mass needed for space travel: emulation technology would reduce this.

Third, another set of species risks accrue from the emergence of machine superintelligence. It has been argued that successful artificial intelligence is potentially extremely dangerous because it would have radical potential for self-improvement yet possibly deeply flawed goals or motivation systems. If intelligence is defined as the ability to achieve one’s goals in general environments, then superintelligent systems would be significantly better than humans at achieving their goals – even at the expense

of human goals. Intelligence does not strongly prescribe the nature of goals (especially in systems that might have been given top-level goals by imperfect programmers).

Brain emulations gets around part of this risk by replacing the de novo machine intelligence with a copy of the relatively well understood human intelligence. Instead of getting potentially very rapidly upgradeable software minds with non-human motivation systems we get messy emulations that have human motivations. This slows the "hard takeoff" into superintelligence, and allows existing, well-tested forms of control over behaviour – norms, police, economic incentives, political institutions – to act on the software. This is by no means a guarantee: emulations might prove to be far more upgradeable than we currently expect, motivations might shift from human norms, speed differences and socioeconomial factors may create turbulence, and the development of emulations might also create spin-off artificial intelligence.

Four, emulations allows exploration of another part of the space of possible minds, which might encompass states of very high value (Bostrom, 2008).

Unfortunately, these considerations do not lend themselves to easy comparison. They all depend on somewhat speculative possibilities, and their probabilities and magnitude cannot easily be compared. Rather than giving a rationale for going ahead or for stopping WBE they give reasons for assuming WBE will – were it to succeed – matter enormously. The value of information helping determining the correct course of action is hence significant.

Discussion: Speculation or just being proactive?

Turning back from these long-range visions, we get to the mundane but essential issues of research ethics and the ethics of ethical discourse.

Ethics matters because we want to do good. In order to do that we need to have some ideas of what the good is and how to pursue it in the right way. It is also necessary for establishing trust with the rest of society – not just as PR or a way of avoiding backlashes, but in order to reap the benefit of greater cooperation and useful criticism.

There is a real risk of both overselling and dismissing brain emulation. It has been a mainstay of philosophical thought experiments and science fiction for a long time. The potential impact for humanity (and to currently living individuals hoping for immortality) could be enormous. Unlike de novo artificial intelligence it appears possible to benchmark progress towards brain emulation, promising a more concrete (if arduous) path towards software intelligence. It is a very concrete research goal that can be visualised, and it will likely have a multitude of useful spin-off technologies and scientific findings no matter its eventual success.

Yet these stimulating factors also make us ignore the very real gaps in our knowledge, the massive difference between current technology and the technology we can conjecture we need, and the foundational uncertainty about whether the project is even feasible. This lack of knowledge easily leads to a split into a camp of enthusiasts who assume that the eventual answers will prove positive, and a camp of sceptics who dismiss the whole endeavour. In both camps motivated cognition will filter evidence to suit their interpretation, producing biased claims and preventing actual epistemic progress.

There is also a risk that ethicists work hard on inventing problems that are not there. After all, institutional rewards go to ethicists that find high-impact topics to pursue, and hence it makes sense arguing that whatever topic one is studying is of higher impact than commonly perceived. Alfred Nordmann has argued that much debate about human enhancement is based on "if … then .." ethical thought experiments

where some radical technology is first assumed, and then the ethical impact explored in this far-fetched scenario. He argued that this wastes limited ethical resources on flights of fantasy rather than the very real ethical problems we have today. (Nordmann 2007)

Nevertheless, considering potential risks and their ethical impacts is an important aspect of research ethics, even when dealing with merely possible future radical technologies. Low-probability, high impact risks do matter, especially if we can reduce them by taking proactive steps in the present. In many cases the steps are simply to gather better information and have a few preliminary guidelines ready if the future arrives surprisingly early. While we have little information in the present, we have great leverage over the future. When the future arrives we may know far more, but we will have less ability to change it.

In the case of WBE the main conclusion of this paper is the need for computational modellers to safeguard against software suffering. At present this would merely consist of being aware of the possibility, monitor the progress of the field, and consider what animal protection practices can be imported into the research models when needed.

Acknowledgments: I would like to thank Håkan Andersson, Peter Eckersly, Toby Ord, Catarina Lamm, Stuart Armstrong, Vincent Müller, Gaverick Matheny, and Randal Koene for many stimulating discussions helping to shape this paper.

References

Anderson, E. (2004). Animal Rights and the Values of Nonhuman Life. In Sunstein, C. R. & Nussbaum, M. (eds.), *Animal Rights: Current Debates and New Directions*. p. 289. Oxford: Oxford University Press.

Arrabales, Raul; Ledezma, A.; Sanchis, A. ConsScale: A Pragmatic Scale for Measuring the Level of Consciousness in Artificial Agents. *Journal of Consciousness Studies*, Volume 17, Numbers 3-4, 2010 , pp. 131-164(34)

Nick Bostrom, Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards. *Journal of Evolution and Technology*, Vol. 9, No. 1 (2002)

Nick Bostrom, Astronomical Waste: The Opportunity Cost of Delayed Technological Development, *Utilitas* Vol. 15, No. 3 (2003): pp. 308-314

Nick Bostrom, The future of human evolution, *in Death and Anti-Death: Two Hundred Years After Kant, Fifty Years After Turing*, ed. Charles Tandy (Ria University Press: Palo Alto , California , 2004): pp. 339-371. http://www.nickbostrom.com/fut/evolution.html

Nick Bostrom, Why I Want to be a Posthuman When I Grow Up. In *Medical Enhancement and Posthumanity*, eds. Bert Gordijn and Ruth Chadwick (Springer, 2008): pp. 107-137.

Bostrom, Nick. "Dignity and enhancement." *Contemporary Readings in Law and Social Justice* 2 (2009): 84

Bostrom, Nick (2013) Existential Risk Prevention as Global Priority, Global Policy (2013), forthcoming. http://www.existential-risk.org/concept.html

Nick Bostrom, Eliezer Yudkowsky, The ethics of artificial intelligence, In the *Cambridge Handbook of Artificial Intelligence*, eds. William Ramsey and Keith Frankish. Cambridge University Press, 2011

Brain Preservation Foundation, Bill of Preservation Rights, http://www.brainpreservation.org/content/preservation-rights (Downloaded on March 1 2013)

Nicholas F. Britton & Suzanne M. Skevington, On the Mathematical Modelling of Pain, *Neurochemical Research*, Vol. 21, No. 9, 1996, pp. 1133-1140

David Chalmers. *The Conscious Mind: In Search of a Fundamental Theory* (1996). Oxford University Press

Chalmers, D. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17(9-10), 9-10. pp. 7-65(59)

Cotterill, Rodney (2003). CyberChild: A Simulation Test-Bed for Consciousness Studies. In Owen Holland (ed.), *Machine Consciousness*. Exeter: Imprint Academic.

Daniel C. Dennett, Why you can't make a computer that feels pain, *Synthese*, 38 (1978) 415-456

M Djurfeldt, M Lundqvist, C Johansson, M Rehn, Ö Ekeberg, A Lansner. Brain-scale simulation of the neocortex on the IBM Blue Gene/L supercomputer. *IBM Journal of Research and Development* 52:1.2, p. 31-41 2008

Chris Eliasmith, Terrence C. Stewart, Xuan Choo, Trevor Bekolay, Travis DeWolf, Yichuan Tang, Daniel Rasmussen. A Large-Scale Model of the Functioning Brain. *Science* 30 November 2012: Vol. 338 no. 6111 pp. 1202-1205

Falit, B. P., & Gross, C. P. (2008). Access to experimental drugs for terminally ill patients. *JAMA: The Journal of the American Medical Association*, 300(23), 2793-2795.

Francis Fukyama, *Our Posthuman Future*, Farrar Straus & Giroux, 2002

Gamez, David (2005). An Ordinal Probability Scale for Synthetic Phenomenology. In R. Chrisley, R. Clowes and S. Torrance (eds.), *Proceedings of the AISB05 Symposium on Next Generation approaches to Machine Consciousness*, Hatfield, UK, pp. 85-94.

Gamez D. Progress in machine consciousness. *Conscious Cogn*. 2008 Sep;17(3):887-910. Epub 2007 Jun 14

E. Richard Gold, *Body Parts: Property Rights and the Ownership of Human Biological Materials*. Georgetown University Press, 1 Mar 1998

Goodall J. An order of pure decision: Un-natural selection in the work of Stelarc and Orlan. *Body & Society*, 5: 149-170, 1999

Hanson, R. (1994). If uploads come first: The crack of a future dawn. *Extropy*, 6

Hanson, R. (2008). Economics of the singularity. *IEEE Spectrum*, 37-42.

Michael H. Herzog, Michael Esfeld, Wulfram Gerstner. Consciousness & the small network argument. *Neural Networks* Volume 20, Issue 9, November 2007, Pages 1054–1056

Henry Markram, The Blue Brain Project. *Nature Reviews Neuroscience*. Volume 7, 2006 p. 153-160

Merkle, R. (1989). *Large scale analysis of neural structures*. CSL-89-10 November 1989 Palo Alto, California: Xerox Palo Alto Research Center. http://www.merkle.com/merkleDir/brainAnalysis.html

Metzinger, Thomas (2003). *Being No One*. Cambridge Massachusetts: The MIT Press

Miller FG and Rosenstein DL. Challenge experiments. In Emanuel E.J., Grady C., Crouch R.A.et al (Eds.), *The oxford textbook of clinical research ethics*. Oxford: Oxford University press, 2008, pp. 273-279.

Moor, J.H. (1988). Testing robots for qualia. In H.R. Otto and J.A. Tuedio (eds), *Perspectives on Mind*. Dordrecht/ Boston/Lancaster/ Tokyo: D. Reidel Publishing Company

Alfred Nordmann, If and Then: A Critique of Speculative NanoEthics, *Nanoethics* (2007) 1:31–46

Robert Preissl, Theodore M. Wong, Pallab Datta, Myron D. Flickner, Raghavendra Singh, Steven K. Esser, Emmett McQuinn, Rathinakumar Appuswamy, William P. Risk, Horst D. Simon, Dharmendra S. Modha. Compass: A scalable simulator for an architecture for Cognitive Computing. In proceedings of Supercomputing 2012, Salt Lake City, November 10-16 2012 http://www.modha.org/blog/SC12/SC2012_Compass.pdf

Prinz, Jesse J. (2003). Level-Headed Mysterianism and Artificial Experience. In Owen Holland (ed.), *Machine Consciousness*. Exeter: Imprint Academic

Regan, Tom. *The Case for Animal Rights* (Berkeley: The University of California Press, 1983).

Regan, T. and P. Singer, eds. *Animal Rights and Human Obligations* 2/e (Englewood Cliffs, NJ: Prentice Hall, 1989).

Sandberg, Anders (2013), 'Feasibility of whole brain emulation', in Vincent C. Müller (ed.), *Theory and Philosophy of Artificial Intelligence* (SAPERE; Berlin: Springer), 251-64.

Sandberg, A., & Bostrom, N. (2008). *Whole brain emulation: a roadmap*. Oxford: Future of Humanity Institute, Oxford University.

John C. Schofield, *Analgesic Best Practice for the Use of Animals in Research and Teaching - An Interpretative International Literature Review*. Food and Agriculture Organization of the United Nations (FAO). 2002

J.G. Taylor. Commentary on the 'small network' argument. *Neural Networks* Volume 20, Issue 9, November 2007, Pages 1059–1060

Zeigler, B. (1985). *Theory of Modelling and Simulation*. Malabar: Krieger.

Zeigler, P. B., Praehofer, H., & Kim, T. (2000). *Theory of modeling and simulation: integrating discrete event and continuous complex dynamics systems*. Academic Press