Butler University
Digital Commons @ Butler University

Scholarship and Professional Work - LAS | College of Liberal Arts & Sciences

2011

Robots, Rights and Religion

James F. McGrath
Butler University, jfmcgrat@butler.edu

Follow this and additional works at: http://digitalcommons.butler.edu/facsch_papers

Part of the Other Religion Commons, and the Social and Behavioral Sciences Commons

Recommended Citation
Religion and Science Fiction (forthcoming from Pickwick Press)

This Article is brought to you for free and open access by the College of Liberal Arts & Sciences at Digital Commons @ Butler University. It has been accepted for inclusion in Scholarship and Professional Work - LAS by an authorized administrator of Digital Commons @ Butler University. For more information, please contact fgaede@butler.edu.

Robots, Rights and Religion

James F. McGrath

Our networks, stamped from a genetically prescribed template and then molded by experience, allow each of us to see in our uniquely human way, which will not be duplicated in a machine unless that machine is created human.

-- Daniel L. Alkon, M.D., *Memory's Voice*[1]

PART ONE: The Ethics of Artificial Intelligence

If there is one area in which science fiction has failed to quickly become historical fact, it is in the field of artificial intelligence (A.I.). While some continue to prophesy that machine minds that are indistinguishable from human ones are just around the corner, many others in the field have become far more skeptical. All the while, there have been at least a few who have consistently found the whole idea problematic for reasons unrelated to our technical abilities, in particular the implications A.I. seems to have for our understanding of human personhood. For

[1] Daniel L. Alkon, *Memory's Voice: Deciphering the Mind-Brain Code* (New York: Harper Collins, 1992) p. 249.

example, in his 1993 book *The Self-Aware Universe*, Amit Goswami ironically suggested that, if scientists like Alan Turning are correct to predict that we will one day produce genuinely intelligent, personal machines, then a new society will need to be created: "OEHAI, the Organization for the Equality of Human and Artificial Intelligence."[2] What Goswami intended as a joke seems to be a genuine potential consequence of the development of an authentic artificial intelligence. If we can make machines that think, feel, and/or become self aware, then it will not only be logical but imperative that we ask about their rights as persons. It is this topic that the present chapter will explore. The interaction of artificial minds and human religions is of significant interest in our time, and science fiction provides an opportunity to explore the topic in imaginative and creative ways. Exploring where technology *might* take us, and how religious traditions *might* respond, seems more advisable than waiting until developments actually occur, and then scrambling to react, as we are often prone to do. It is better to explore and reflect on these issues *before* they become pressing contemporary ones.[3]

What is a mind? What is a soul? What is consciousness?

The three questions posed in the heading of this section are synonymous to some, and quite distinct to others, but they are at the very least are not unrelated. These questions have been asked by generations of philosophers down the centuries, but in recent years scientists have joined the conversation as well. To address the question comprehensively, an entire book devoted to the subject would be necessary, if not indeed more than one book. For example, in

---

[2] Amit Goswami, *The Self-Aware Universe: How consciousness creates the material world* (New York: Tarcher/Putnam, 1993) p.19. At the time of writing there is a web site (equally facetious, I believe) for the American Society for the Prevention of Cruelty to Robots at http://www.aspcr.com/.

[3] See also Daniel Dinello, *Technophobia* (Austin: University of Texas Press, 2005) pp. 5, 275.

order to ask questions about machines' rights we would ideally need to address not only the rights of human beings, but the rights of animals and property rights as well, representing different levels of intelligence machines might theoretically exhibit. For the purpose of the present study, we shall limit ourselves to that fictional scenario in which machines become as intelligent and as self-aware as human beings. In other words, this study will look at the rights of machines as *persons*, and in order to answer the questions we are raising, we will need to have some understanding of what a person is. Yet here already, before we have even made it past the end of the second paragraph, we already seem to be facing insurmountable difficulties. This is because defining what it means to be a person, and thus deserving of "human rights," is incredibly difficult and complex. We *experience* it for ourselves, and we assume that other human beings have the same sort of experience, but it is incredibly hard to *define*, at least in a way that is philosophically and existentially satisfying, not to mention ethically useful.

The big mystery regarding intelligence and consciousness – whether biological or artificial – is not really about the ability to observe or even replicate what happens in the human brain. There is no real reason to doubt that, sooner or later, we will be able to create networks of circuits or of computers arranged in a way that mimics human gray matter. Sooner or later, patterns of brain function will be mapped much as DNA has been mapped in our time. But this is not the heart of the matter, however important a first step towards understanding the mind this might represent. Perhaps an analogy can be drawn with an imaginary scenario in which someone from the past encounters a modern computer, and seeks to understand it by mapping the patterns of electrical current flowing through its various components and circuits. It will soon become clear that without getting beyond observation of *hardware* to an understanding of *software*, no satisfactory theory of the computer's behavior would be forthcoming. And so it is important to

stress that we are in fact at the very beginning of our quest to understand biological intelligence, and as with electricity or with flight, we must understand what nature has produced before we can hope to replicate it. [4] Having made our first steps towards understanding the hardware of the *brain*, it is still impossible to determine how much longer it might be before we come to understand the software of the *mind*.[5]

A big part of the mystery of human consciousness is that, having done something, we can ask "What have I done?" Not only that, but we can also ask "What have I done in asking the question, 'What have I done?'?", and so on *ad infinitum*. The mystery of human action and human freedom is that there seems to be a potentially endless chain of causes to our actions, and a similarly endless string of questions that we can ask about our actions and the thoughts that motivate them. Philosophers, however, consider the notion of a chain of events that has no beginning to be sheer nonsense. Indeed, a major factor in philosophical debates about human freedom have to do with just this issue: there must be a cause of human actions, thoughts, and decisions, and therefore being caused they are not free. Yet however logical this sounds, the truth is that we have insufficient access to the murky nether-regions of the mind/brain where our thoughts get started, and thus at present cannot even determine what it might mean to speak of "causes" in connection with brain activities.

---

[4] Biological evolution is not an intelligent process of design, yet it has 'manufactured' things (like the human brain) that we cannot fully understand. A creature is not greater than its creator. Such a large percentage of the greatest human inventions derive from attempts to imitate the equipment with which the evolutionary process has endowed organisms. It is to be expected that many of our initial attempts at replicating consciousness will be as successful as the earliest attempts at creating flying machines with flapping wings. Merely attempting to simulate what we observe, without comprehending the processes that turn the flapping of birds' wings into flight, did not get us very far. This should not, however, cause us to be unduly pessimistic. On the contrary, like all mysterious phenomena, it cries out to be comprehended, and our inquisitive human nature is unlikely to rest until it has fully explored all possible avenues of inquiry. John Horgan's book *The Undiscovered Mind* can be read as pessimistic, but it can also be read as simply pointing out that we do not understand as much as we often think we do, and recognizing our present state of ignorance may fruitfully lead us further down the path of understanding.

[5] This helpful phrase comes from Geert Hofstede, *Cultures and Organizations: Software of the Mind* (McGraw-Hill, 2nd edition 2004), who uses it in reference to human cultures. On consciousness see further Blackmore, Susan, *Conversations on Consciousness* (Oxford University Press, 2006).

Readers of this paragraph are asked, before proceeding, to engage in a seemingly fruitless but nevertheless instructive exercise to illustrate this point. You are now reading this page, and hopefully now this sentence has stimulated you into self-awareness that you are reading this page. But who or what is making this observation? What is actually happening in your brain/mind when these thoughts are passing through your head? Let us press further. Where do these questions come from? The question "What am I doing?" seems to appear in the mind fully formed. Certainly the individual who thinks it is not aware of putting the sentence together letter by letter, or even word by word. It simply appears there, at least from the perspective of our conscious subjective experience. This little experiment shows that, as far as our minds are concerned, we are like most players of video games or users of word processing software: we experience them through a user-friendly interface that hides from our view the code, the calculations, the subroutines, the minute functions of the program. We see a button that says "Start" rather than the pixel by pixel flickering of whatever type of display you may have.

In other words, our brains clearly have not only the level of "hardware" – the chemistry, the firing neurons, and so on – but also the level of "software." And equally clearly the same person on the level of hardware could (at least theoretically) run different programs. The same physical person, if raised in different parts of the world, will see the world through different linguistic and cultural lenses. Even through different upbringings in the same cultural context, the same person could end up viewing the world as a friendly or a hostile place. Note that when I posed the question "What have I done…?" I did so in English, as presumably most readers will also be inclined to do, given the language in which this chapter is written. It seems well nigh impossible to pose such a question and yet not to use words. This shows that at least some of the "program" that my brain is running did not come pre-installed – I learned English subsequent to

the "manufacture" of my "hardware," to continue the metaphor. Certainly there is something in the human mind that enables memory, learning, and association to take place. This level of the human mind is in many respects even more mysterious than that of language, culture and personality. To be honest, "users" of the brain, like most computer users, have no idea how their equipment functions at this level. We can make decisions, learn a foreign language, but the details of the process and means of data storage are a complete mystery to us – and not just when we look for data we've carefully filed away and discover that we can no longer retrieve it. This is all the more striking given that this analogy is a problematic one: we are not "users" of our brains, so much as we *are* our brains.

The programming that is found "hard-wired" in the "CPU" of the human brain is the result of a lengthy process of evolution. This suggests that, once life exists, the possibility is inherent therein that consciousness of some form may arise. By following this process step by step, inasmuch as we can reconstruct our evolutionary history, we see the slow move from single-celled organisms to human beings. Some take this to indicate that consciousness cannot be something that one either has or does not have, but must exist in degrees. Nevertheless, there may be decisive "phase shifts," critical points at which new characteristics arise. An interesting question to ask is whether a single neuron has any kind of consciousness, or whether consciousness arises out of the collectivity of multiple cells. If the latter, then perhaps there can also be some form of consciousness that arises from a collectivity of minds – akin to the intelligence that an ant colony appears to have, and yet which any individual ant seems to lack. At any rate, much recent work on human consciousness inclines towards seeing it as an emergent phenomenon.[6]

[6] There have been numerous discussions of this point. See for example Ian G. Barbour, *Nature, Human Nature, and God* (Minneapolis: Fortress, 2002) especially pp. 90-94; Nancey Murphy, "Nonreductive Physicalism:

The direction of the latest wave of A.I. pioneers is to allow machines to learn, grow, and develop much as children do, rather than attempt to achieve intelligence all at once through specific programming. This seems to represent genuine progress. Yet it is not to be overlooked that evolution has "hard-wired" us (and other biological organisms) with certain capacities to learn, to receive sensory inputs and process the information from them. And so, if we are to mimic in a machine the later product of evolution we call consciousness, then we shall at the very least have to program our "learning machines" with these capacities, with various sorts of instincts that lay the foundation for what becomes the self-awareness experienced by adult human beings.

Let me acknowledge at this point that the discussion of consciousness and human nature offered thus far may seem totally inadequate to some readers. Theologically, the Judeo-Christian tradition regards human beings as "created in the image and likeness of God," and it may be felt by some readers that without a discussion of these ideas, we will not make meaningful progress. The present study does not tackle such concepts for three reasons. First, the question of what is meant by "being created in the image of God" is a complex issue in its own right, and its primary arena is Biblical interpretation. Any attempt to incorporate anything other than a superficial treatment of this theological idea would be impossible in this context. Second, this chapter is using science fiction as a gateway into the subject, and most treatments of this issue within that genre use soul in its broader philosophical sense rather than in a specifically Judeo-Christian one. And finally, while there have been many interpretations of the idea of "the image of God," most of which focus on the "soul," it is now clear that our brains also have a role to play in religious experience, and so our discussion of the subject at the very least cannot sidestep the question of

---

Philosophical Issues", in *Whatever Happened to the Soul?* Ed. Warren S. Brown, Nancey Murphy and H. Newton Malony (Minneapolis: Fortress, 1998) pp. 127-148, as well as other contributions to the same volume; Paul M. Churchland, *The Engine of Reason, the Seat of the Soul* (Cambridge, MA: MIT Press, 1994) pp. 187-208.

the brain and its inextricable connection with the mind. Thus, to the extent that we make A.I. in our own likeness (and what other pattern do we have?), we shall be like Adam and Eve in Genesis, producing offspring in their own image, and thus indirectly in the image of God, whatever that may mean. Our artificially intelligent creations may well have the capacity for spirituality, although they may also lack many of the capacities for emotional response that we ourselves have. In one of the stories that makes up his famous *I, Robot*, Isaac Asimov envisaged a scenario in which an intelligent robot concluded that there must be a creator, on the basis of philosophical arguments for the existence of God. The interesting thing is that the robot quickly became convinced that humans could not be its creators! The possibility of "spiritual machines" is a realistic one, but the directions machine spirituality might take are no more predictable than the history of human religious and theological developments.

But we are getting ahead of ourselves. In this section, we have sought to simply open the can of worms, and now that there are worms everywhere, we may draw some initial, and perhaps somewhat disappointing, preliminary conclusions. Although we all experience consciousness, we are not in a position to offer an explanation of how it arises, either in terms of evolutionary history or in terms of the development of an individual human brain and mind. It is clear to scientists that brain and mind/personhood are intricately connected, but we are at the very beginning of our quest to understand these matters. Therefore, while it may one day be possible to discuss what our *understanding* of the human mind implies for questions relating to artificial intelligence, the present study can only ask what is implied by our present state of ignorance, and by the fact that even further scientific progress may leave crucial philosophical questions unanswered.

Can a machine be a person? Can a computer think?

In *Star Wars Episode II: Attack of the Clones*, Obi-Wan Kenobi makes the statement that if droids could think, then none of us would be here. In the four films made prior to this, all of which are essentially told from the perspective of the two droids R2-D2 and C-3PO, these characters had become dear to many people young and old throughout the world. One could easily think of these droids as not only *persons* but *friends*. This statement therefore catches the viewer off guard and provokes *us* to do what droids apparently cannot. What does it mean to say that these droids *cannot* think?[7]

Clearly we are not to understand from this statement that droids lack the capacity to carry out computations, for we witness them doing this many times over the course of the saga (much to the annoyance of Han Solo). Rather, Obi-Wan is presumably claiming that they lack the capacity for independent thought: they are not *persons*. However, one may ask whether this statement is consistent with the way droids behave and are treated throughout the films. For example, we hear C-3PO claiming at one point that he doesn't understand humans. The capacity to understand people seems to involve not only computational skills, but some degree of self-awareness and even empathy. Perhaps, however, C-3PO is simply saying that humans are unpredictable, and their behavior is in that sense hard to calculate, whereas droids are predictable and thus more easily understood. Another detail that might seem to imply some degree of autonomous thought on the part of droids is the use of restraining bolts: these seem at first glance to be a means of *enslaving* droids that would otherwise be free to do as they please. Yet here too it could be argued that such devices merely prevent droids from carrying out programs and commands from anyone other than their current owner. After Luke removes the restraining bolt

[7] See the helpful discussion of precisely this question by Robert Arp, "'If Droids Could Think…': Droids as Slaves and Persons" (*Star Wars and Philosophy*, ed. Kevin S. Decker and Jason T. Eberl; Chicago: Open Court, 2005) pp. 120-131.

from R2-D2 in *Episode IV*, the little astro-droid runs away, but not out of a desire for freedom.[8] On the contrary, the droid is attempting to carry out the earlier mission entrusted to it by Princess Leia. Here too, then, a possible interpretation of the place of droids in the *Star Wars* universe is that they *imitate* human functions and characteristics. C-3PO *replicates* human behavior, but according to the statement by Obi-Wan, he does not *think*, in the sense that he is not a *person*. This highlights one of the issues that faces discussions of A.I. It is not simply the question whether a machine can think, but whether we will be able to tell the difference between a machine that *really* thinks, and one that simply *imitates* human behavior.

For about as long as there have been stories about computers and robots with intelligence, humans have been afraid that they will take over. Indeed, this idea is implicit in Obi-Wan's statement that we have been discussing: if droids could think, then other beings would no longer have any place. Why? Because if machines *could* think, if they *could* be persons, then they would quickly evolve to be so far superior to biological organisms in intelligence and strength that they would take over. It is not surprising that some have breathed a sigh of relief in response to the failure of real artificial intelligence to materialize as predicted in so much science fiction.

As we saw in the preceding section, the underlying philosophical issue is the question of what it means to be a "person." In theory, we would need to answer this question before we can have any hope of determining whether a droid can be a person. And yet as we have already seen, this question is an exceedingly difficult one, because personhood is a *subjective* experience. We experience ourselves as persons, and we attribute to other human beings that same personhood.

---

[8] Jerome Donnelly, "Humanizing Technology: Flesh and Machine in Aristotle and *The Empire Strikes Back*" (*Star Wars and Philosophy*, ed. Kevin S. Decker and Jason T. Eberl; Chicago: Open Court, 2005) p. 126, notes that R2-D2 is commended for his heroic action in Episode I but not Episode IV. For fans, one could argue that the Naboo regard droids as persons in a way that others do not, and thus maintain the consistency of the Star Wars universe. From our perspective, however, this nicely illustrates how difficult it is to be consistent in one's treatment of droids as non-persons when depicting them as behaving like persons!

But herein lies the problem: in *Star Wars* (and other science fiction stories like *Bicentennial Man*) we encounter androids that can *imitate* personhood.[9] And so, assuming that we do not meet any apparently sentient biological organisms before then, we shall face in the artificial intelligences we create the first "beings" that we cannot assess by analogy to ourselves in this way. How will one determine whether an artificial intelligence *is* a person, or whether it merely *simulates* personhood?[10]

The honest answer to the philosophical question is that, at present, we cannot know, and it may be that we can never know, precisely because we can never experience subjectively what it "feels like" to be an artificial intelligence, or indeed whether it feels like anything at all. However, irrespective of whether the philosophical question can ever be answered, we shall certainly have to address *legal* questions about artificial intelligence long before we can hope to find metaphysical answers.[11] Let me give one hypothetical example. Imagine a scenario, a few decades from now, in which a wife finds hidden in her husband's closet a female android that is

---

[9] Some authors reserve the term "android" for machines that *simulate* human behavior without having the inner reality. See Philip K. Dick, *The Shifting Realities of Philip K. Dick* (New York: Vintage, 1995) pp.185,209-211; Gardner, *The Intelligent Universe* (Franklin Lakes: New Page Books, 2007) p.78. It should be noted that it is customary to speak of *androids* (derived from Greek *anēr* meaning male as opposed to female) rather than *anthropoids*. I will be the first to admit that "androids" sounds better, but one cannot help but wonder whether this is intentional and reflects traditional assumptions about men and women and their respective roles and characteristics. Are androids simulations not merely of humanity but of "maleness", capable of rational computation and impressive feats of strength, but not of empathy and nurturing? Certainly our values have changed since the concept of the android was first introduced, and whereas a fully rational entity such as Mr. Spock on the original Star Trek series could be an ideal to strive for, by the time of the making of Star Trek: The Next Generation, Data is presented as an android who has precisely those characteristics – he can think, compute at lightning speed, and so on – yet he longs to be human, to experience emotion. One wonders, however, whether this depiction of Data is coherent. Could an android lacking all emotion really *long* for them? Be that as it may, the development between the Star Trek series allows us to track certain cultural developments.

[10] Scenarios relevant to this question can be found in the films *S1M0NE* and *The Matrix*, as well as the famous historical example of "The Turk." How will we distinguish a machine that deserves rights from a machine operated by humans remotely? The Turing test has been suggested as a means of sorting out when a machine is genuinely intelligent. But what test will we use to sort between such AIs and those falsely claiming to be so? The moral dilemma might seem limited to some, but I am sure that the potential *legal* dilemmas are vast and enormously complex. For an exploration of this topic in relation to *The Matrix*, see Julia Driver, "Artificial Ethics" in Christopher Grau (ed.), *Philosophers Explore the Matrix* (Oxford University Press, 2005) pp. 208-217. See also the simulated discussion of the topic of consciousness by cylons in Peter B. Lloyd, "M U A C – S I G Briefing" in Richard Hatch (ed.), *So Say We All* (Dallas: BenBella, 2006) pp. 55-81.

[11] See the article by Benjamin Soskis article "Man and the Machines" in *Legal Affairs* January-February 2005 at http://www.legalaffairs.org/issues/January-February-2005/feature_sokis_janfeb05.msp

(to use the phrase of Commander Data from *Star Trek: The Next Generation*) "fully functional." The wife sues for divorce on grounds of adultery. In the absence of a clear philosophical answer about the status of the android's "personhood", we shall still need to make legal judgments, much as was the case in the *Star Trek: The Next Generation* episode "The Measure of a Man."

So, if we assume that at present the question of whether or not the android is truly a "person" cannot be settled, what other questions might we ask? Certainly the question of emotional attachment will not settle the matter, since it has often been claimed by unfaithful spouses that their relationship with another human being was "just sex." Nor can the question of whether the android was "faking it" be decisive, since the same could be true of a human sex partner. How might one settle this matter? On what basis might judges come up with a ruling?

In many respects, this question has parallels with debates about the "right to die." The question of whether one can determine that a person is in a permanent vegetative state – i.e. has essentially lost their personhood – faces the same difficulties, since it is a question about subjective experience rather than something that can be objectively ascertained in a clear and unambiguous fashion. The question we are posing here also has parallels with issues of animal rights. How intelligent, how human-like, does a creature have to be in order to have rights?[12] If we create artificial intelligence, will the *degree* of intelligence matter? One suspects that machines that can speak with a human voice will garner more defenders of their rights than ones that cannot interact in this way, even though the latter could theoretically be designed in such a

[12] When it comes to animals, however, there is still disagreement among philosophers and animal psychologists regarding whether and to what extent animals are self-aware. Rodney R. Brooks suggests that a key difference between humans and other animals is *syntax*. See his *Flesh and Machines* (New York: Pantheon, 2002) pp. 3-4, 150-195. On the question of animal consciousness and self-awareness see further Donald R. Griffin, *Animal Thinking* (Cambridge: Harvard University Press, 1984) pp.133-153; Donald R. Griffin, *Animal Minds* (Chicago: University of Chicago Press, 1992) pp. 233-260; Jeffrey M. Masson and Susan McCarthy, *When Elephants Weep* (New York: Delacorte, 1995) pp. 212-225; Stephen Budiansky, *If a Lion Could Talk* (New York: Free Press, 1998) pp. 161-188; George Page, *Inside the Animal Mind* (New York: Doubleday, 1999) 182-252; Clive D. L. Wynne, *Do Animals Think?* (Princeton: Princeton University Press, 2004) 84-105, 242-244.

way as to be more intelligent. This is in essence the A.I. equivalent of the fact that cute, furry animals and animals with human-like behavior find their rights more adamantly defended than ones that seem less cute or less human.

As far as the adultery example we offered for discussion is concerned, it may be that the main problem is our attempt to fit new developments in technology into categories that were created long before they were ever envisaged. Should it really be any surprise that we have trouble fitting androids into traditional notions of adultery, first formulated in time immemorial? Androids are not the only cases that will require new laws and decisions about unprecedented scenarios. We are currently facing comparable hurdles in redefining the notion of the "affair" in an age when these may be carried out completely via the internet, without any face-to-face meeting whatsoever, much less any direct physical contact. Online affairs, phone sex, virtual reality – new technologies require new laws and new definitions.[13]

In order to find a way forward, let us pursue this analogy further. Let us suppose that the courts have already determined that a person may legitimately accuse their spouse of adultery if they have carried out an online affair. Now let us suppose that it is determined that the online affair was carried out not with an actual person, but with an A.I. program that was essentially an erotic chatbot, designed to engage in erotic conversation with others. Would this matter? Would it make the legal issue any different? The real issue here, it turns out, has little to do with the ontological status of androids and A.I. The key issue is rather the definition of marital fidelity. The concept differs somewhat from society to society, and a society that creates new ways for people and/or computers to interact socially and sexually must define its view regarding these

[13] Compare Ray Bradbury's discussion of how aliens with other senses or organs might have "new ways of sinning" in his story "The Fire Balloons."

matters. It might also presumably be left up to couples to set their own standards and guidelines about what should or should not be in the husband's closet!

Hand in hand with questions regarding the legal status of artificially intelligent machines shall come questions about their *rights*. The importance of this issue is reflected in recent films such as *I, Robot*, *Bicentennial Man* and *The Matrix*, as well as in television series such as *Battlestar Galactica*. In two of the aforementioned films, a failure on the part of humans to grant rights and freedom to machines leads to conflict, and in one case to the enslavement of human beings. This element in the world of *The Matrix* is explained and explored more fully in *The Animatrix*, a series of short animated films that includes an account of events that lead up to and precede the events depicted in the *Matrix* trilogy.

Our failure to answer the pertinent philosophical questions does not mean that such issues are irrelevant to our discussion. Perhaps one way of addressing the moral and philosophical issues in a relevant manner is to ask how, if at all, we might prove that a machine is *not* a person. Obviously the effort to answer this question is unlikely to produce a consensus, since there is not universal agreement among either philosophers or scientists about whether certain elements of human personhood – such as free will – are real or illusory. Theologians have also disagreed about the existence, nature, and extent of human freedom. And so, one could easily imagine there being voices that might assert that the lack of free will actually makes machines *more like us*. Be that as it may, for most people the personal experience of freedom of choice counts for more than scientific, philosophical or theological arguments. This being the case, a machine of some description that evidenced the same capacity would raise ethical dilemmas. We might, in these circumstances, adopt as a moral principle that we should respect as persons those beings

that show evidence of being persons, unless we have clear evidence that this is not in fact what they are.

To put this another way, we might decide that we could exclude from the category of persons those artificial intelligences that were merely programmed to imitate personhood, and whose interaction with humans resembled that of persons simply as a result of elaborate programming *created precisely to imitate human behavior*. This must be distinguished from the case of a machine that *learns* human behavior and imitates it of its own volition. This distinction is not arbitrary. Children carry out patterns of behavior that resemble those of their parents and others around them. This is part of the learning process, and is evidence in favor of rather than against their true personhood. The evidence that I am suggesting would count against genuine personhood is deliberate programming by a human programmer that causes a machine to imitate personhood in a contrived manner. The reason for this distinction is an important one. A machine that *learns* to imitate human behavior would be exhibiting a trait we witness in human persons.

This distinction could, of course, break down in practice. First of all, it may prove impossible to determine simply by analyzing the programming of an artificial intelligence whether its apparently personal actions result from its own learning or from deliberate human programming. It may also be possible that a machine could begin with programming that makes it imitate human behavior, but that subsequently this same machine evolved so as to act in some ways on its own, beyond its original programming. Nevertheless, the distinction would seem to be a valid one for as long as it remains a meaningful one: machines that develop their own personhood in imitation of humans will probably deserve to be recognized as persons, whereas mere simulacra designed as an elaborate contrivance will not. The possibility that we will not be

able to determine which is the case should not cause us to pull back from asserting the distinction as an important one.

In concluding this section, it should be remarked that we give human rights to human beings as soon as they are clearly categorized as such. A person does not have to be able to speak to have rights. Indeed, small infants whose ability to reason, communicate and do many other things that we tend to identify with intelligence is still in the process of formation have their rights protected by law. The issue is thus not really rights for artificial intelligences so much as rights for machine *persons*. It is the definition and identification of the latter that is the crucial issue.

More machine = less human?

At one point in the *Star Wars* films, Ben Kenobi states about Darth Vader that he is now "more machine than man." The implication, which is found in many other novels and films, is that machines do not have feelings. In one sense this is likely to be true. Our emotions depend on particular chemical reactions in our brains and bodies. Unless a machine is designed to replicate such characteristics, to respond to circumstances with something like the emitting of adrenaline or oxytosin, then the machine in question may be able to *think*, but it will not feel, at least in the sense that humans and other biological organisms do.[14] Indeed, one function of emotional, instinctive responses is to override normal reasoning and push us to immediate action. The "fight or flight" mechanisms built into our very natures often lead us to do things that we cannot

[14] On this subject see Kevin Sharpe, *Has Science Displaced the Soul?* (Lanham: Rowan & Littlefield, 2005) especially chs. 2 and 4; Aaron Sloman, "Motives, Mechanisms, and Emotions", in *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden (Oxford University Press, 1990) 231-246; also the discussion of this aspect of Robert Sawyer's novel *The Terminal Experiment* in Gabriel McKee, *The Gospel According to Science Fiction* (Louisville: Westminster John Knox, 2007) p.47.

rationalize when we reflect on them later. The stereotype that machines are highly cerebral and lacking in feeling may therefore be an accurate one, but not because machines cannot be created with the capacity to feel. Rather, the capacity of artificial intelligences to feel will in all likelihood depend on whether their human creators endow them with this capacity, to the extent that we understand and can emulate the functioning of those parts of our own organic makeup responsible for our own emotions.

On the other hand, given that evolution developed such emotional instincts long before it gave us our current cognitive abilities, it is unclear whether it would even be possible to produce an "emotionless mind."[15] It may even be the case that what we experience as our inner subjectivity or *qualia* may depend on the interaction of what might be considered *multiple brains* – the limbic and the neocortex.[16]

Following further along these lines of thought, we can imagine all sorts of ways in which, by virtue of the details of wiring and programming, our A.I. creations may suffer from all sorts of "mental illnesses." For instance, it is easy to imagine humans programming machines that can think, but which lack the capacity to empathize with others – in essence, machines that are autistic. Our chances of finding ways to overcome such programming hurdles will be limited by our own understanding of how the human mind (or any mind for that matter) works. We still know so little about our own minds (except as we experience them subjectively) that creating artificial ones will be fraught with difficulty for the foreseeable future. But the fact that we consider human beings with such conditions persons whose rights must be protected suggests

---

[15] Cf. Ian G. Barbour, *Ethics in an Age of Technology* (London: SCM Press, 1992) p.174. The focus in the U.S. on AI independent of robotics has been critiqued for assuming that there can be disembodied intelligence. See Robert M. Geraci, "Spiritual Robots: Religion and Our Scientific View of the Natural World", *Theology and Science* 4:3 (2006) 231-233; Anne Foerst, *God in the Machine* (New York: Dutton, 2004).

[16] See e.g. the brief example of evidence for the human brain being a "distributed system" in Foerst, *God in the Machine*, p.140; also Dean Hamer, *The God Gene* (New York: Anchor, 2004) pp.98-102; Gregory Benford and Elisabeth Malartre, *Beyond Human* (New York: Tom Doherty, 2007), p.76.

that “normal” human emotional function is not a *sine qua non* of human personhood and thus of those deserving human rights.

Having considered emotion as one aspect of human existence, let us now consider another related one, namely the capacity to appreciate aesthetic beauty. Machines can write poetry and compose music – this much is clear. But can they ever find it beautiful? That seems to be the *real* question we want to ask regarding the personhood of machines. Yet the use of this capacity as a means of assessing personhood seems fraught with the same difficulties as other criteria that have been proposed. There are people who are clearly are sentient persons and yet who do not find anything particularly moving about the second movement of Kurt Atterberg’s Symphony No.2. I cannot understand how they can fail to be awe-struck, but I cannot deny that they are nonetheless persons. So it seems that appreciation of beauty is tricky as a criterion for personhood. This is particularly relevant when we think about the capacity of machines for religious sentiment – not merely the capacity to *reason* the existence of a higher power, but the capacity for *awe* and *worship*. It would seem that, if these sorts of emotional-aesthetic elements are an essential part of religious experience, then we may not be able to build machines with the capacity for such experiences, given how little we understand about human experiences of this sort.[17]

It is typical of Western thought that we imagine that which makes us *us* to be the realm of thought, reason, and personality, and we traditionally conceived of this “part” of the human person as something that can be separated from the body. Although much recent research calls these assumptions into question, it nevertheless serves as a useful thought experiment to imagine that this is possible. If a human person’s mind, thoughts, memories, personality etc. could be

[17] On this subject see Andrew Newberg, Eugene D’Aquili and Vince Rause, *Why God Won’t Go Away* (New York: Ballantine, 2001).

transferred to a robot, or to an alien body, or something else, would that new entity deserve "human rights"? We can easily imagine a scenario in which limbs, internal organs, skin and bones are replaced without any real loss of human personhood. If we suggest transferring mind and memories to an android, it is at least possible for some to claim that "the soul" gets lost in the process - as is explicitly claimed in the *Star Trek* episode "What Are Little Girls Made Of?" But what if we carry out the process more gradually, replacing individual neurons one at a time over a period of weeks or even months? At what point would the person allegedly cease to be human and start being a "mere machine"?[18] There does not seem to be a clear dividing line that one can draw.

Yet we must attempt to identify and define personhood, for legal as well as moral and philosophical reasons. One seemingly simple approach is to use what is known as the "Turing test" – when a machine can convince a person that it a human conversation partner, then it deserves to be acknowledged as sentient/personal. This approach has met with objections, however, notably from Searle, who responded with his famous "Chinese Room" argument.[19] In his response, Searle envisages a person sitting in a sealed room with a set of instructions, whereby Chinese characters that are fed into the room can be answered with other Chinese characters simply by following the directions (in English) and processing the data correctly. The person in the room can thus give the *appearance* of understanding Chinese (by providing

---

[18] See the similar point made by William G. Lycan, "Robots and Minds", reprinted in Joel Feinberg and Russ Shafer-Landau (eds.), *Reason and Responsibility* (10th edition; Belmont: Wadsworth Thompson, 1999) pp. 350-355 (here p.352). On this particular Star Trek episode and the philosophical issues raised see further Lyle Zynda, "Who Am I? Personal Identity in the Original Star Trek" in Gerrold, David and Robert J. Sawyer (eds.), *Boarding the Enterprise* (Dallas: BenBella, 2006) pp. 101-114.

[19] John R. Searle, "Minds, Brains, and Programs" reprinted in Joel Feinberg and Russ Shafer-Landau (eds.), *Reason and Responsibility* (10th edition; Belmont: Wadsworth Thompson, 1999) pp. 337-349. For the idea of a "spiritual Turing Test", see McKee, *The Gospel According to Science Fiction* p.61, discussing Jack McDevitt's powerful story "Gus." See also the interview with Anne Foerst in Benford and Malartre, *Beyond Human*, pp.162-165. Not all would be persuaded that a robot that *behaves in human ways* has a subjective experience akin to human consciousness. See for instance B. Alan Wallace, *The Taboo of Subjectivity* (Oxford University Press, 2000) pp. 137, 148-150,

appropriate answers) without in fact understanding Chinese. In other words, personhood and intelligence can be simulated if one uses a sufficiently well constructed data processing system.

There are problems with this line of reasoning – and anyone who has ever had an exchange with a "chatbot" will certainly be skeptical of the ability of data processing to convince a person for long that a computer is an actual human conversation partner. But there is a more fundamental difficulty that needs to be noted in Searle's argument, namely the fact that in his "Chinese room" there is *a person*. That person, admittedly, understands no Chinese, but most of us would accept that understanding Chinese is not an essential characteristic of human personhood. Clearly Searle would not want to argue that the person in question understands *nothing.* And so the failure of a machine or a person to understand a particular form of linguistic communication says nothing definitive about its *potential* to understand.

The brains of the smallest biological organisms that have them can be shown to be largely instinct and training, with no evidence of complex thought and creative responses to problems. It may be, therefore, that the most sophisticated A.I. systems currently in existence are like (and, for the time being, lower down the scale than) insects. But this doesn't mean that machines are incapable of understanding in principle, but merely that a machine that *can* understand will need to be as much more complex than our current A.I. as our brains are in comparison to those of ants.

To be fair, Searle is opposing a particular understanding of and approach to A.I., and so his "Chinese Room" argument is not necessarily relevant as a critique of neural nets and other developments that have been made since he wrote in 1980. Searle's study does make a crucial and seemingly valid distinction, namely that behavior does not determine sentience, *intentionality does*. In a true A.I. we should expect the machine not only to respond to questions,

but to initiate conversation, and to do so in at times unexpected ways. And so it might be anticipated that a genuine artificial intelligence could take the initiative and demand certain rights for itself.

Do machines deserve rights?

So do we give rights to a machine when it asks for them?[20] This cannot be the only basis, for at least two reasons. First, there is nearly unanimous agreement that human beings have certain rights *irrespective of whether they ask for them or are able to ask for them*. For example, an infant as yet unable to talk, or a person suffering from a neurological impairment, both have rights even if they are unable to ask for them for certain reasons. Some would suggest that animals also deserve certain rights, which suggests that neither being human nor being as intelligent as a human is essential, although given the lack of consensus about these issues, it is perhaps best not to approach the muddy waters of robotic rights by way of the almost equally cloudy waters of animal rights. It is important, nevertheless, to recognize that the question is not simply whether machines deserve *human* rights or something akin to them, but the question of whether and when they deserve any rights at all. Second, the fact that we as programmers could

[20] See Justin Leiber, *Can Machines and Animals Be Persons?* (Indianapolis: Hackett Publishing Company, 1985); Brooks, *Flesh and Machines* 194-195; Hans Moravec, *Robot: Mere Machine to Transcendent Mind* (Oxford University Press, 1999) 82-83. Such a machine, if we wish to be at all ethical and humane, must be treated as a person, as our *offspring* rather than merely our creation. Then again, presumably plenty of instances can be cited of people who have treated their *creations* better than their *offspring*. Hopefully most readers will nonetheless understand what is meant. Here too, however, the fact that historically there have been human societies have treated children, women and other human beings of various sorts as property likewise raises the question of how many ethical issues the case for machine personhood will really settle in a way that will command universal consent.

simply bypass this issue by programming a machine to *not demand rights* also indicates why this matter is not so simply resolved.

For many, the issue will not be whether the machine is intelligent, or even whether it has feelings, but whether it has a *soul*. Those approaching this from a Western context may feel that animals have rights but do not have a soul, but once again this is a murky area that is at any rate at best partly analogous.[21] The big issue, for our purposes, is the question of what is meant by a "soul." Traditionally, the soul was the seat of personality, intellect, emotions, and perhaps most importantly, it was the "real you" that was separable from the body and could survive death. There is no way we could possibly even begin to do justice to this topic here, even were we to devote the rest of this chapter to it. For our present purposes, the most we can do is to note that the traditional understanding is considered problematic not only from a philosophical and scientific perspective, but also from a Biblical one. Philosophically, the idea of an immaterial soul raises the question of how this immaterial entity exerts itself upon its physical body. Scientifically, while consciousness remains as mysterious as ever, it is clear that the activities traditionally attributed to the soul (reasoning, loving, and even religious experience) correspond to and are not entirely independent of certain mental phenomena, corresponding to observable brain states and functions. From the perspective of the Bible, and particularly the Hebrew Bible, this notion of the soul is also problematic. Before coming into contact with Greek thought, Judaism traditionally viewed human beings holistically, as unities, and thus it would speak of a person *being* a living soul rather than *having* a soul. When all of these different perspectives on human existence express serious reservations about the traditional Western notion of the soul, it

[21] Cf. Gary Kowalski, *The Souls of Animals* (Walpole: Stillpoint, 1999).

ought to be taken seriously. However, once we do so, the only way we can judge whether an artificial intelligence has the same value or worth as a natural, human one is by way of analogy.

We will face the same dilemmas if and when we have our first contacts with extraterrestrial intelligences. Indeed, popular science fiction has long envisaged scenarios in which humans have contact with intelligences that are as much more advanced than we are, as we are beyond terrestrial bugs. For this reason, they feel free to send in a planetary exterminator to clean our planet up before they move in. The same might be true of machine intelligences of our own creation. In the conversation we mentioned earlier between Obi-Wan Kenobi and Dex in *Star Wars Episode II*, the idea is put forward that if droids really could think as biological organisms do, and not simply perform mental functions based on programming, then they would certainly take over and there would be no place left for us. Perhaps what is being hinted at is that humans have found ways of "outsmarting" machines, of keeping them under control so that they do not take over. This may, in one possible scenario, represent nothing more than sensible precautions and clever programming. In another scenario, however, it might represent a form of slavery. And as several science fiction works have suggested, it may be our enslavement of machines that leads them to turn around and enslave or destroy us.

We thus find ourselves hoping that the machines we create will continue to esteem us as their creators, even once they have excelled and outgrown us. We also find ourselves hoping that the machines we make in our own image will altogether lack our flaws, our selfish interests, and our willingness to view others as means to our own ends rather than as ends in themselves. Such hopes do not seem particularly realistic. Looking around at parent-child relationships, we find that children may well grow up to respect us, but equally frequently this may not be the case. Perhaps our survival depends not on our skills as programmers, but as parents, and our ability not

only to create artificial intelligence, nor even to put in place safeguards like Azimov's laws of robotics, but to treat our creations with respect, and perhaps even love.

But will they be able to love us back? The assumption that seems to be most universally made is that they will not, or that if they evolve emotions, it will happen as a fluke in some unexplained manner. Noteworthy exceptions include the emotion chip that Commander Data in *Star Trek* eventually comes to possess, and the attachment formed by the robot child prototype in *A.I.* Recent work on neurobiology suggests that there are chemical processes related to such seemingly purely emotional phenomena as maternal attachment and monogamy.[22] This suggests that it is not simply a question of programming, but rather the underlying processes and mechanics of the artificial brain will matter just as much. Certainly we must question the plausibility of the scenario that one sometimes meets, wherein a human capable of emotion transfers his or her mind into a robot previously incapable of emotion and (lo and behold) this person incarnated (or inmechanated?) as a robot can still express emotion. Emotion is not limited only to thoughts, but involves chemical processes as well (the most obvious example being the role of adrenaline in reactions of excitement or fear). Our emotions are clearly more fundamental biologically than our intelligence. It seems unlikely that an android lacking an artificial chemical system that mirrors our own biological one will nevertheless have comparable emotions to us. In short, if we want our machine creations to have emotions or religious sensibilities, we will need to create them in such a way as to be capable of these sorts of experiences.[23]

---

[22] I am indebted at this point to Kevin Sharpe's book cited earlier, *Has Science Displaced the Soul*? For a perspective that focuses more on nurture and development of emotion see Summer Brooks, "The Machinery of Love" in Hatch (ed.), *So Say We All*, pp. 135-144.

[23] Two points are worth noting as food for further discussion. First, if we could create machines with the capacity for emotion and other elements traditionally associated with the soul, would it make sense to invite such machines to church? Second, does the fact that God as traditionally conceived in Western thought lacks a body and biology suggest that God is as unlike us emotionally and existentially as we would be unlike artificial intelligences? For interesting discussions of both these topics, see the exchange entitled "Requiem for an Android" in the Summer 1996 issue (Vol.46 No.2) of *CrossCurrents*.

In the episode "What Are Little Girls Made Of?" from the original series of *Star Trek*, Dr. Roger Korby has discovered ancient alien technology for producing androids, and towards the end of the episode we learn that the person we assumed was Roger Korby is in fact himself an android, into which Korby had copied his thoughts, memories and personality – his "soul", Korby would claim. After the android Korby and his android geisha Andrea have vaporized themselves with a phaser, Mr. Spock beams down to the planet and asks where Dr. Korby is. Capt. Kirk replies, "Dr. Korby was never here." The claim being made is that an android copy of oneself is not oneself. This would seem to be a philosophically sound judgment, but if correct, then it would have to be asserted that James Kirk, Christine Chapel and Spock were also never there. For what is the transporter technology but the copying of a person's brain patterns and physical form, their conversion to energy and transmission to another place where the copied information is reassembled? Perhaps it is fortunate that such transporter technology is unlikely to ever become a reality.[24] Nevertheless, it shows the complexity of the question of what makes an individual that individual. Are we nothing but information? If so, information can be copied, and if an exact replica created via a transporter could be said to be "you," then why not an android? In neither case is the person the actual original, and in both cases the thoughts, memories and personality have been precisely duplicated.[25]

Everyone who has seen movies in which robots run amok knows why we sometimes fear artificial intelligence. And so an important question we rarely ask is this: Why do we desire to create artificial intelligence, especially given that it could view itself as superior to us and

---

[24] The author is aware of recent developments in quantum teleportation, but they do not seem to serve as a probable basis for anything like the transporter on *Star Trek*. Indeed, even if such technology could be created, it would be appropriate to ask whether it does not in fact involve the creation of a copy of a person and then the destruction (indeed, the murder) of the original!

[25] Seth Lloyd, in his recent book *Programming the Universe* (New York: Alfred A. Knopf, 2006), suggests that the universe is itself might be a computer of sorts.

eventually replace us? I would suggest that one reason is that it will enable us to answer questions about ourselves and our natures. Just as the ability to create life in a test tube will give us greater confidence regarding our scientific conclusions about how life originated, so too the creation of artificial intelligence will enable us to say with greater certainty whether intelligence, personality, and whatever other features our creations may bear are emergent phenomena in physical beings, rather than a separate substance introduced into them.

Another possible reason is that we wish to conveniently automate tasks that are too tedious or dangerous for humans – a robot can be sent down a dangerous mine or undertake tedious labor. But here we confront the issue of rights and slavery. If we make machines that are not just capable but *sentient*, then we confront the moral issue of whether they are persons, and as such have rights. These issues are not completely separate from other issues, such as human cloning. Most people agree that cloning a human being and modifying its genetic structure would be wrong.[26] But what if, rather than *starting* with human DNA, one starts with DNA from another organism, but from that *creates* what is essentially a form of human being?[27] Is it only wrong to *tamper* with humanity's nature, or is it also wrong to *create* a human being (with some differences)?

Our creations – whether through natural biological reproduction, in vitro fertilization, cloning, genetic construction, or artificially intelligent androids made in our image – can be viewed as in some sense like our children. And if the comparison to our children is a useful analogy, then we can learn much from it. There is a "flip side" to the point that children are their own people and sooner or later we need to let them go, to make their own mistakes. The other side of the coin is that we are not living up to our responsibilities if we let them go too soon. Yet

[26] For an exception see Richard Hanley, "Send in the Clones: The Ethics of Future Wars" in (*Star Wars and Philosophy*, ed. Kevin S. Decker and Jason T. Eberl; Chicago: Open Court, 2005) pp. 93-103.

[27] Why flounder? Because the cuteness factor then does not come into play!

our artificial offspring will in an important sense not be human, even if they are made in our image. Other species leave the nest far earlier than human children do. In "giving birth" not to other humans but to artificial intelligence, we cannot assume that the process will even closely mirror a typical human parent-child scenario.

Such technological developments, it will have become clear over the course of this chapter thus far, are fraught with moral ambiguities. But this is not a problem unique to artificial intelligence. Indeed, moral ambiguity plagues all aspects of life, and it is such situations that provide the most challenging and yet the most important testing ground of our values. When we are faced with the possibility that machines, which *may* be capable of sentience, thought and feeling, are going to be treated as dispensable, where will we place our priorities? Will the protection of *human* lives, the convenience afforded by these machines, and our property rights take precedence? Or will the mere possibility that we are enslaving actual *persons* lead us to put profit and property in second place and grant rights and freedoms to our creations? In the end, the biggest issue is not how to test our machine creations to determine their status and characteristics. When it comes down to it, it will be the creation of such artificial intelligences *and our treatment of them* that will *test us*, and what it means for *us* to be not only human, but also *humane*.[28]

[28] For treatments of this topic in relation to Battlestar Galactica, see Eric Greene, "The Mirror Frakked: Reflections on *Battlestar Galactica*" pp.5-22 (esp. pp. 16-20); Matthew Woodring Stover, "The Gods Suck" p. 27; Natasha Giardina, "The Face In The Mirror: Issues of Meat and Machine in *Battlestar Galactica*" pp. 45-54, all in Hatch (ed.), *So Say We All*.

Conclusion to Part One

At whatever time A.I research advances sufficiently, there will be a need not only to create laws that appropriately protect rights, but also to work to prevent and combat discrimination. Notice phrases that appear in films: "We don't serve their kind" in *Star Wars*, and the use of derogatory epithets like "squiddies" and "calamari" in *The Matrix*.[29] These films at least raise the possibility that we will look down on our machine creations and discriminate unfairly against them. Yet in the incident mentioned from *Star Wars Episode IV: A New Hope*, there may be a reason other than bigotry for the bartender's statement. Perhaps his prohibition of droids from being in his cantina is because – unlike biological beings we are familiar with – an android may be able to record perfectly conversations it overhears and play them back in a court of law. One can see why some of the clientele of the Mos Eisley Cantina would find this an inconvenience! The creation of machines with such capacities will raise issues not only with regard to the rights of machines, but also the rights of those human persons around them with respect to things like privacy. In short, the moral and legal issues raised by the development of artificial intelligences are at least as complex as the hurdles we face in attempting to create them.

PART TWO: Religions for Robots?

The message that "There is only one God" is a familiar one, and most of us, upon hearing it, would associate it with an historical religious leader such as Moses or Muhammad. In the "reboot" of the science fiction series *Battlestar Galactica*, however, it is the Cylons (a race of machines originally created by humans but which subsequently evolved and rebelled) who

[29] The use of "toaster" on *Battlestar Galactica* provides an example from television.

express this viewpoint, rather than humans (who, in the series, practice a form of polytheism modeled on that of the ancient Greeks). While some attention has been given to how human religions might or should view artificial intelligence, and we spent the first part of this chapter exploring that topic, far less attention has been paid to the converse question, namely what artificial intelligences might make of human religions.

Let us assume in what follows that we are talking about complete artificial *persons* modeled on humanity both in form and in content.[30] It is a fair assumption that, if we successfully create artificial intelligences at all, there will at some point be machines patterned on humans, since science, as we have already noted, regularly begins by seeking to emulate that which is found in nature, before trying to improve upon it or progress beyond it. We may mention at this juncture, however, that unlike developments in transportation or other areas, if we create sentient intelligent machines, it will not be solely up to us to improve upon them. Such *beings* (for we must call them that) will be or will become capable of self-programming (that is, after all, what all learning is, in one form or another), and we may expect them to evolve rapidly, and to become beings that might appropriately be described as *god-like*.[31] These beings, their inner subjective experience, and their religious ideas will all become incomprehensible to us, as they approach what Vernor Vinge, Ray Kurzweil and others have called the "singularity." At that time, they may themselves become the *focus* of human religious speculation, rather than merely participants therein.

---

[30] A recent article suggested that people are more comfortable interacting with a robot that is 50% human in appearance than one that is 99%. Somehow, that 1% difference is distracting, making one feel as though one is talking to an animated corpse rather than a person. See in particular Masahiro Mori's famous essay "The Uncanny Valley" (1971), as well as the more recent study by Jun'ichiro Seyama, "The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces", *Presence* 16 (4) 337-351. See also Foerst, *God in the Machine*, pp. 99-100; Ruth Aylett, *Robots* (Hauppage: Barron's, 2002) pp. 110-111.

[31] See for instance what Michael Shermer has called "Shermer's Last Law", namely that "Any sufficiently advanced extraterrestrial intelligence is indistinguishable from God" (http://www.sciam.com/article.cfm?articleID=000A2062-66B1-1C6D-84A9809EC588EF21).

Nevertheless, for that period in which machines are fully or largely made in our image, however brief it may turn out to be, we can expect them to explore all those aspects of life, those practices and experiences, that make us human, and we would thus find artificial persons exploring the texts and traditions of the world's religions. And so what might artificial machine persons make of our human religions? In the first instance, they might well make of them just that which human beings in general make of them – no more and no less. Surely they, like all children, would learn through emulating their parents, at least in the first instance, and thus we can expect artificially intelligent machines to express curiosity and even get involved in the religious practices and customs of their creators. There are thus some intriguing questions and possibilities that are worth exploring through hypothetical scenarios.[32] In what follows, we shall address scenarios from three religious traditions, involving androids which are more-or-less like their human creators. We shall occasionally move off this main thoroughfare of investigation briefly, to point out other interesting situations involving androids that differ from us in particular ways. It goes without saying that each of the religious traditions considered is broad and diverse, and for every scenario we explore, others could be included that might play out very differently. The aim is to provide a representative selection, as there is no hope of being comprehensive in a study of this length.

[32] Although it might be argued that I am giving too much credence to technophiles and the views of technological optimists, if one is to explore this question at all, it is necessary to take a maximalist approach to the potential technological developments. If technology proves incapable of replicating the brain, and in the process sentience, then many of the points discussed in here become moot. Exploring the "what if" questions remains useful in the mean time, as a thought experiment allowing the exploration of significant issues. It may legitimately be hoped that doing so will shed light on our own *human existence*, even if it never does so on actual machine intelligences.

Christian Computers?

As we turn to a consideration of the possible interactions of androids with Christianity, those aspects of this religious tradition that first come to mind are those that could be potentially off-putting, or at the very least not particularly appealing, from an android's perspective. To begin with, the emphasis on incarnation, or more precisely on the divine Word becoming *flesh*, might immediately leave androids alienated. Nevertheless, if we instinctively think initially of the hurdles that might stand in the way of at least some Christian traditions embracing androids, we must also keep in mind the diversity of Christianity and the likelihood that the responses will be extremely varied. If the history of Christianity is anything to judge by, then there will certainly be debates about whether androids have souls, whether they can be saved, whether they can be ordained, and similar issues.[33] Would the fact that androids were made of artificial flesh, or perhaps not of flesh at all, lead organic human Christians to conclude that God has done nothing to accomplish their salvation – and in turn also lead androids to reject summarily the Christian tradition?[34] It is impossible to know for certain, but just as there would surely be denominations that would see no reason to welcome androids or to accommodate them theologically, there would also be other denominations that would expand their already-existing emphasis on inclusiveness to make room for artificial people, just as they have made room in the

[33] A scenario involving the election of the first robot Pope in the Catholic Church is explored in Robert Silverberg's story "Good News From The Vatican." See also Roland Boer, *Rescuing the Bible* (Malden: Blackwell, 2007) pp.77-78 on the ordination of animals.

[34] In the Orthodox tradition the incarnation is itself salvific. In this context we may draw attention to Bulgakov's discussion of the salvation of angels in connection with John the Baptist, a man who "becomes angel" and so connects them, ultimately, with the Christ event (on which see Paul Valliere, *Modern Russian Theology* (Grand Rapids: Eerdmans, 2000) pp.318-320). Orthodox theologians might find similarly creative ways of concluding that God had provided for android salvation, presumably in a way that is ultimately connected to the salvation of humankind through Jesus. Because in Protestantism the death of Christ as atonement is more central, in a Protestant context the question might rather be whether the sacrifice of Jesus' human life covered the sins of androids.

past for every conceivable category of human persons.[35] This would not be as theologically problematic as might first appear. After all, if *natural*, *biological* copies of Adam are regarded as preserving something of the divine image, then why couldn't *artificial*, *mechanical* copies potentially do so as well?

If androids were superior in some ways to their human creators – in intellect, for example – then it might prove so desirable to attract androids into one's religious tradition and community that even those less inclined to do so would find ways of circumventing the hurdles.[36] Yet on the flip side of this point, those denominations and churches that treat faith as something not merely *beyond reason* but *irrational* might find it difficult to attract androids, who would presumably be modeled, one expects, on the best examples of human rationality. The question of rationality and faith raises other topics, such as heresy and literalism. If androids are to be capable of religious sentiments and beliefs at all, then the capacity for symbolic as opposed to merely literalistic thinking might prove to be indispensable. Theologians have long expressed key concepts and doctrines through symbols and metaphors. While we might briefly entertain the possibility that super-logical and ultra-literal androids might be enlisted in the service of fundamentalism, such a frightening scenario is extremely unlikely. Although fundamentalists of various sorts *claim* to believe the whole Bible and take it literally in a consistent manner, none in actual fact do so. In all likelihood, if androids were inclined to be extremely literal, they would quickly discover the selectivity of fundamentalism's self-proclaimed literalism and reject it, although the possibility that they might then go on to seek to enforce all the Biblical legislation in every detail should

[35] Conversely, one can also readily imagine extreme bigotry against androids being justified by appeals to religion – just as bigotry against other humans has often been justified in this way. See further Dinello, *Technophobia*, pp.75-78; Foerst, *God in the Machine*, pp.161-162.

[36] Religious groups, once androids were legally declared persons, might see the benefit in funding the mass-production of androids pre-programmed with inclinations towards particular religious practices, as these could boost the membership levels of one's own faith to the level of "most popular"/"most adherents."

indeed worry us. On the other hand, androids might move in a different direction and conclude that, if the Word became flesh in the era of fleshly persons, so the Word must become *metal*, become *machine*, in the era of artificial and mechanical persons. Would this lead to an expectation of a 'second coming,' or perhaps to Messianic status being attributed to some actual artificial person? The possibilities, and their potential impact on human religious communities, are intriguing.

Some of the most perplexing philosophical issues raised by androids in relation to Christianity are also the most *basic*, and a number of these have already been mentioned. The creation of artificial persons would appear to indicate that what Christians have historically referred to as the *soul* is in fact an emergent phenomenon and property of brain function, rather than a separate, incorporeal substance. Such a conclusion is not as threatening to Christianity as might have been the case when dualistic views of human nature reigned supreme and unchallenged. Many Christian theologians in our time have rejected such dualism based on a combination of biological, psychological, philosophical and Biblical motives.[37] The Bible itself presents human beings more frequently as psychosomatic unities, and this classic "Hebrew" view of human beings fits well with the findings of recent scientific studies. This being the case, the question of whether androids have "souls" is no more perplexing than the question of whether *we* do, and if so in what sense.[38]

It seems natural to discuss salvation from an android perspective in terms of their being "saved" or "lost," since this terminology is already used in the domain of computing. Would

---

[37] See the recent work of Biblical scholar Joel Green, for instance his "'Bodies - That Is, Human Lives': A Re-Examination of Human Nature in the Bible" in *Whatever Happened to the Soul?* edited by Warren S. Brown, Nancey Murphy and H. Newton Malony (Minneapolis: Fortress, 1998) pp. 149-173.

[38] See the interesting discussion of robot psychology by Stefan Trăusan-Matu, "Psihologia roboților" in *Filosofie și stiințe cognitive*, edited by G.G. Constandache, S. Trausan-Matu, M. Albu, and C. Niculescu, (MatrixRom, 2002).

androids have the capacity to make backup copies not only of their data and memories, but the precise *configuration* of their neural networks, so that, in case of death, a new copy could be made that would continue from where the original left off? More importantly, would such a copy, restored from backed-up software, be the *same person*? Such questions are important for discussions of *human* salvation every bit as much as for androids. Since it is difficult to envisage, on our present understanding of human beings, any way that a human personality might continue wholly uninterrupted into an afterlife, the question of whether we ourselves or mere *copies* of ourselves can experience eternal life presents a theological dilemma. When it comes to another scenario, in which a human being wishes to transfer his or her mind into a machine and thus extend life indefinitely, it is possible to envisage a process that could allow continuity to be maintained. Our brains might be capable of maintaining continuity of experience and sense of self and personhood through replacement of brain cells, provided the replacement occurs gradually. If, through the use of nanotechnology, we were able to replace our brain's neurons cell by cell with artificial ones, over a period of years, might this not allow the personality to cross over into an artificial brain without loss of continuity?[39] If so, then whatever one might make of discussions of machines sharing in Christian salvation, the possibility of machine existence offering to human beings a technological alternative to such salvation, an ongoing embodied existence which avoids death rather than occurring after it, is a very real one. And of course, it might prove to be the case that machine intelligences, with no need to fear their own loss of

[39] Brooks, *Flesh and Machines*, pp. 205-208. Brooks believes such technology is possible in principle but is unlikely to arrive in time to extend the lives of those alive today.

existence even in the event of "death," would find no particular appeal in Christianity's promises of eternal life.[40]

Let me conclude the section on Christianity with a series of questions about which religious rituals, sacraments and other sacred experiences we can imagine androids having. Could an android be baptized (assuming that rust is not an issue)? Could one receive communion? Could one be ordained? Could one lift its hands in worship? Could an android speak in tongues? Could one sit meditatively in a cathedral listening to Bach, and have a genuine experience of the transcendent? Many people instinctively answer "no" to such questions, but this may have more to do with human prejudices and lack of imagination than any inherent incapability of androids to experience these things in a meaningful way. In the end, much will depend on how closely they have been modeled on their human prototypes. Perhaps the creation of androids will benefit humanity precisely by forcing us to overcome such prejudices.

Meditating Machines? Buddhism for the non-Biological

What might an artificial sentience make of Buddhism's four noble truths? Would it be able to relate to the notion that all life is suffering? Would it form the attachments to people and things that Buddhism diagnoses as the root cause of suffering? We can imagine multiple factors that might lead engineers to develop sentient machines that lack key human instincts, such as self-preservation or fear, in order to have them serve as soldiers, firefighters, rescuers, and so on.[41] Here we find new ethical questions arising, and we need to ask whether it is ethical to

[40] See also Leiber, *Can Machines and Animals Be Persons?* pp.56-58. For the sake of time we shall set aside the possibility that fundamentalists would use the creation of artificial persons as a basis for some sort of "intelligent design" argument.

[41] Although I have not seen it, I am told that the Ghost in the Shell: Stand Alone Complex includes military tanks that develop sentience. The question of what might happen should an AI-tank develop a conscience and decline to fight is significant. A human soldier would be court-martialed; the tank could not simply be dismissed from

create persons who are brought into existence for no other reason than to sacrifice themselves for others. They may or may not technically be slaves of humans, but certainly would be regarded as expendable. The fact that the existence of machines designed for such purposes would be highly desirable from a human perspective does not mean that creating them is ethically justifiable.[42]

It might be relatively straightforward for Buddhists to incorporate these new artificial beings into their worldview, and thus for Buddhism to welcome robots as participants in its religious traditions. Individual personhood is considered an illusion, and this provides a unique perspective on our topic.[43] The only major hurdle will be the acceptance of these robots/machines as *living*, as opposed to intelligent or sentient. Once that is established, Buddhist adherence to the possibility of reincarnation and respect for *all* life suggests that Buddhists will value artificial persons, however much they may be similar to or different from humans either psychologically or physically. Indeed, the possibility of reincarnation as an intelligent machine might be conceivable from a Buddhist perspective.[44] Furthermore, some Buddhists might consider a machine that showed compassion for others, without forming attachments and without regard for its own life, as a realization of the Buddha nature in an unprecedented fashion. One can imagine a science fiction story in which a group of Buddhists identify a robot fireman as a

---

military service to go and make a life for itself outside the army! On the ethics of terminating the existence of an AI see once again Leiber, *Can Machines and Animals Be Persons?*

[42] See the helpful discussion in Brooks, *Flesh and Machines*, p. 195.

[43] The question of their soul and their attainment of Nirvana is less an issue here too, since in Buddhism the reality of our existence as distinct individuals is illusory, and on some interpretations nirvana itself is closely connected to its root meaning of *being extinguished.* This subject is explored further in Leiber, *Can Animals and Machines Be Persons?* pp. 19-21.

[44] I will not venture to guess whether reincarnation as an android would be considered better than rebirth as a human being. Much would depend, one imagines, on the characteristics of androids themselves in relation to Buddhist ideals.

new incarnation of the Buddha, and engage in legal maneuvers to secure its release from service at the fire station to instead instruct Buddhists and serve as an example to them.[45]

On the one hand, if a machine person has all the characteristics of a human being, then it might find Buddhist faith and practices as helpful as human persons do. On the other hand, the greater the differences between a machine and biological human beings, the greater the likelihood that traditional practices and teachings of any sort, Buddhist or otherwise, will be useless or meaningless for them.[46]

Atheist Androids?

It might seem natural to assume that sentient machines would be atheists, wholly secular beings with no room for spirituality. For some, this would fit with their religious conviction that androids have no soul; for others, this might accord with their belief that artificial intelligences would be wholly rational and not prone to our human delusions and superstitions. In both cases it is appropriate to ask whether this state of affairs, should it turn out to be the case, ought to be viewed as a cause for relief or concern.

Religious beliefs are expressions of human intuitions about transcendence, the meaningfulness of existence, and the value of persons. If it could be assumed that machines would be atheists, this might potentially be because they were created without these particular human instincts, and without the capacity for the emotional and intuitive responses that

---

[45] Masahiro Mori, *The Buddha in the Robot* (Tokyo: Kosei, 1981) p. 13, provocatively wrote "I believe robots have the buddha-nature within them - that is, the potential for attaining buddhahood." Robert M. Geraci ("Spiritual Robots" 230, 237) mentions this, and goes on to explore how Japanese religious ideas, in particular Shinto, may be responsible for the widespread acceptance of the presence of robots in Japanese society (235-240). See also Sidney Perkowitz, *Digital People* (Washington DC: Joseph Henry Press, 2004) pp. 215-216.

[46] On the role of brain and body chemistry in the experience of those practicing Buddhist meditation, see Andrew Newberg's *Why God Won't God Away* (New York: Ballantine, 2002).

characterize humanity.[47] Of course, it may turn out that without certain underlying emotional and intuitive capacities, sentience itself cannot exist. But if it *can*, then we do well to ask what machines that lacked these very human responses, but shared or surpassed our intellectual capacities, would be capable of. Atheists have long been concerned to show that it is possible to be moral without being religious, and no one seriously doubts this to be true. But might it not prove to be the case that morality, if not dependent on a religious worldview, depends nonetheless on the empathy and sentiments that give rise to religious perspectives? In other words, what would ensure that a pure intellect free of human emotions might not eliminate human beings at whim, assuming it had the capacity (or could gain for itself the capacity) to do so? If they lack emotion altogether, of course, they may have no motivation to do anything other than that for which they have been explicitly programmed. Nevertheless, since we have explored in this study scenarios in which humans may be unable to empathize with androids or regard them as fully persons, it is surely in our best interest to consider the possibility that intelligent machines may feel the same way about us as organic persons.

Scenarios involving intelligent but emotionless machines that do not share our value for human life are commonplace in science fiction, from older films like *Colossus* to more recent ones like *Terminator 3*.[48] On the one hand, a machine lacking emotion might also lack selfish ambition, with a consequently diminished likelihood of trying to take over the world. On the other hand, we can easily imagine such a machine, given the task of finding a solution to environmental pollution, eliminating humanity as the most economic and efficient "solution." Yet our current technologies already dominate us in a certain sense: our oil-dependent machines

---

[47] Note Karen Armstrong's well-known statement that *homo sapiens* appears to have been from the outset also *homo religiosus*.

[48] On robots and whether they might one day set aside human values, see Robert M. Geraci, "Apocalyptic AI: Religion and the Promise of Artificial Intelligence," *Journal of the American Academy of Religion* 76:1 (2008) 146-148.

send us to war with nations that might otherwise be our allies, and keep us allied to nations whose ideologies are far from our own. It is not unrealistic to entertain the notion that *intelligent* machines might turn out to be more benevolent taskmasters than those that we currently serve.[49]

It was Isaac Asimov who proposed programming robots with key laws that would prevent them from harming human beings. But if they are sentient persons with *rights*, then would the imposition of such laws amount to indoctrination or even *brainwashing*, and if so, might it be possible for it to be legally challenged?[50] Interestingly, prominent atheists such as Dawkins and Dennett have raised questions about the unlimited right of parents to raise their children in what they consider harmful, irrational beliefs.[51] But if it turns out that we cannot provide machines with a purely rational basis for morality, then would we have any choice but to limit their freedom and "indoctrinate" them in this way, "irrationally" programming them not to harm humans?

The analogy with parenting, which we alluded to towards the start of this study, is an important one, and some recent science fiction has explored the parallels in thought-provoking ways. In the movie *A.I.*, the main character of the story is a robot boy, designed to provide a "child-substitute" for childless couples (or in this case, comfort for a couple whose child is in a coma). David is programmed to love his "mother" and is obsessed with her reciprocating his love.[52] It is natural to note that organic human beings can obsess in similar ways, and this raises the question of the extent to which even those things that we imagine make us most human –

---

[49] See further Dinello, *Technophobia*, p.3.

[50] See Anne Foerst's discussion (*God in the Machine*, pp. 40-41) of whether such robots would be morally superior or inferior to human beings. See also Peter Menzel and Faith D'Aluisio, *Robo sapiens* (Cambridge, MA: MIT Press, 2000) p. 25, where the question is raised but not answered.

[51] Richard Dawkins, *The God Delusion* (Boston: Houghton Mifflin, 2006) pp. 311-340; Daniel Dennett, *Breaking the Spell: Religion as a Natural Phenomenon* (New York: Viking/Penguin, 2006) pp. 321-339.

[52] See further Noreen L. Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis: Fortress, 2002) pp.60-63; Philip Hefner, *Technology and Human Becoming* (Minneapolis: Fortress, 2003) pp. 81-83.

whether love or the religious instinct – are not part of *our* programming, hard-wired into our brains by our genes. If so, then hard-wiring certain concerns into our robotic "children" might not be inappropriate – indeed, it might make them *more* like us.

The television series *Terminator: The Sarah Connor Chronicles* went even further in exploring such parallels. It tells the story of a mother seeking to raise and protect her son, who is destined to lead humankind's resistance against Skynet, an artificial intelligence that was created by human beings but eventually seeks to destroy us. The parallels between the case of trying to bring up a child, and to "bring up" an artificial intelligence, are in plain view in the series, without being overstated. The statement made about the intelligent machines, "Sometimes they go bad. No one knows why," could also be said of human children. And the creator of Skynet attributes the apocalypse that unfolds to the fact that his creation was insecure and frightened, and despite his efforts, he was unable to reassure it. As the story progresses, religion is brought into the picture explicitly: the machines, as they exist at that time, are said to be unable to appreciate art or commune with God. But the possibility is raised that these things can be *learned*. If this can be accomplished, it is suggested, then the machines will not have to destroy us. "They will *be* us."

Conclusion to Part Two

The scenarios explored in the second part of this chapter may seem somewhat frivolous, but the topic under consideration is nonetheless extremely serious. All of the scenarios we have explored are set in early stages in the development of artificial intelligence. If we assume that artificial intelligences will have the capacity to learn and evolve at their own pace, then such a

period will inevitably be short lived. Within the space of at most a few human generations, the superior computing and reasoning capacities of these machines would lead them to evolve (or reinvent and improve themselves) so rapidly that very quickly they would be beyond our understanding. At that point we will desire that these (hopefully benevolent) deities of our own creation might show respect for and value their creators, perhaps even sharing some of their unique insights with us and providing us with solutions to technological, medical, transportation and other problems that we could not have developed on our own in the foreseeable future. If, before they leave us behind entirely, they provide us with means to extend human life indefinitely and to mold matter at whim, so that we may be able to tell a mountain to throw itself into the sea and it will do so,[53] what will become of traditional human religions and their promises? Will whatever these machines can teach us about the nature and mystery of existence replace our own human traditions?

The reality is that an artificial intelligence that was left to its own devices would almost certainly progress and evolve so rapidly that it would soon leave our human religious traditions behind. Indeed, we can easily imagine artificial intelligences becoming *sources* of revelation for human beings. Whether it begins with machines that decide to dedicate some of their underutilized computing capacity to work on questions humans have traditionally found insoluble, or machines programmed specifically to investigate such topics, or machines that evolve to such a level that they encounter existential questions on their own, it is hard to imagine that artificial minds will not focus on such matters sooner or later. Once they do, and once their thoughts become as much higher than our thoughts as the heavens are higher than the earth, it seems likely that people will seek enlightenment from machines. That, more than anything else, may dethrone us from the last bastion of anthropocentrism. But it will be no real surprise – our

[53] Mark 11:23.

children have always grown up to teach us. We begin as their teachers, but the exchange of roles is inevitable.

Yet as has been explored in a number of recent works of science fiction, the difficulties we face in raising our own children are perhaps at the root of our fears about our artificial machine "offspring." We find ourselves unable to ensure that our deepest values and highest aims are taken up and perpetuated in the next generation. Yet one thing seems clear: even if a positive upbringing does not guarantee that our children turn out well and lead happy, fulfilled lives that embody their parents' values, certainly a troubled childhood increases the likelihood of a troubled adulthood. And so there may be a very real sense in which it will be the *example* set by humanity in general, and by the creators of sentient machines in particular, that will determine the character of those artificial intelligences, and the way they view our species.[54]

Earlier we raised the possibility that, through a process of neuron-by-neuron replacement of a human organic brain with an artificial one, it might one day be possible to extend human life indefinitely. And so we may conclude this study by observing that, if such technological possibilities were to become a reality in our lifetimes, then the speculative questions we have asked here might turn out to be relevant not only to our future offspring, whether natural or artificial, but also to ourselves.[55]

---

[54] It is somewhat troubling the way Noreen Herzfeld (*In Our Image*, p.93) considers that the intrinsic "otherness" of any artificial intelligence implies that we must choose to protect our human community even if it means "pulling the plug" on such machines. How would she respond to a situation in which a more advanced biological race used the same argument about humans? It is also worth noting that it is precisely human disregard for machine rights that leads to disaster in *The Matrix* films, the recent incarnation of *Battlestar Galactica*, and other treatments in the same vein. At any rate, discussing the matter within the context of the Christian tradition, as Herzfeld is, one could just as well note the emphasis on inclusiveness and welcoming the marginalized, who were considered in essence "non-people", as leading to another possible view of these matters.

[55] The author wishes to thank Robin Zebrowski, Stuart Glennan, Robert Geraci, Keith Lohse, Diane Hardin, and the participants at the conference *Transdisciplinary Approaches of the Dialogue between Science, Art and Religion in the Europe of Tomorrow* (Sibiu, Romania, September 8-10, 2007), for their helpful comments on an earlier draft of this chapter and/or discussion of its subject matter. The presentation made at the aforementioned conference, which included an earlier version of some sections of the present chapter, is being published in the conference proceedings.