ARENA OF CHANGING

CrossMark

The Ghost in the Machine

Being Human in the Age of AI and Machine Learning

Henrik Skaug Sætra [1] (ID)

Abstract

Human beings have used technology to improve their efficiency throughout history. We continue to do so today, but we are no longer only using technology to perform physical tasks. Today, we make computers that are smart enough to challenge, and even surpass, us in many areas. Artificial intelligence—embodied or not—now drive our cars, trade stocks, socialise with our children, keep the elderly company and the lonely warm. At the same time, we use technology to gather vast amounts of data on ourselves. This, in turn, we use to train intelligent computers that ease and customise ever more of our lives. The change that occurs in our relations to other people, and computers, change both how we act and how we *are*. What sort of challenges does this development pose for human beings? I argue that we are seeing an emerging challenge to the concept of what it means to be human, as (a) we struggle to define what makes us special and try to come to terms with being surpassed in various ways by computers, and (b) the way we use and interact with technology changes us in ways we do not yet fully understand.

Keywords Artificial intelligence · Machine learning · Social relations · Cognition · Human identity

Introduction

Our present age is characterised by rapid progress in the field of digital technology. Techno-logical change has always been an important part of human development, but today's change may be different. Throughout the ages, we have exchanged human labour with other sources of energy: animal power, steam power and then electrical power. We are in the midst of another great alleviation of human labour, but now, we are replacing the need for human *mental* powers, instead of just the physical. Parts of it involves combining artificial intelligence with

✉ Henrik Skaug Sætra
henrik.satra@hiof.no

[1] Faculty of Business, Languages, and Social Sciences, Østfold University College, P.O. Box 700, 1757 Halden, Norway

physical machines—furthering the automation of industrial tasks previously performed by human beings. Other parts, however, concern replacing the human intellect. Machines help us drive, trade stocks in the stock markets, decide who gets loans, help us choose movies, guide our missiles, and smart systems keep track of just about every move we make and makes us more effective. The game of chess was conquered by the machine long ago, and in 2017, one of the final frontiers of human superiority in games was broken. Google's artificial intelligence *AlphaGo* proved able to beat the best humans in the ancient game of Go. Furthermore, machines increasingly complement, and even replace, humans as social companions—even lovers. Our children bond with sociable robots, our elderly get therapy from robotic seals and our lonely ones can find companionship in robots designed for intimacy.

The last part of this equation is the connection between how we employ technology and the way we gather and use vast amounts of data. All kinds of data are collected, and our every move—yes, even the beats of our hearts—can be monitored and analysed. In modern society, it may at the same time be both impossible to hide and difficult to *really* connect to other people. Big data lets machines learn what we like, and even what we are likely to in the future. Many now point out that the way we consume both data and, in a sense, other people, are *changing* us. Not just change the way we view ourselves and struggle to find our place, but how we function. Our plastic brains adapt to how it is used, and the way we use it has changed, both in relation to information and other beings.

In this article, I argue that we are seeing an emerging challenge to the concept of what it means to be human, due to the developments just described. Firstly, what it means to be human, what defines us and makes us special, may change when we are surpassed in various ways by computers. Secondly, what it means to be human is by necessity influenced by the fact that we *are* changed by the technologies we employ. Understanding how technologies change us will be important both because we will need to understand the consequences of such changes, and we may want to prevent certain changes, or at least properly debate the desirability of them.

Artificial Intelligence in Today's Society

> The Promethean myth has acquired an ugly twist: the giant reaching out to steal the lightning from the gods is insane (Koestler 1967, p. 331).

First things first: what is *artificial intelligence* (AI)? In a report detailing the potential malicious use of AI, the authors define it as systems "capable of performing tasks commonly thought to require intelligence" (Brundage et al. 2018, p. 9). *Machine learning* refers to systems that can "improve their performance on a given task over time through experience" (Brundage et al. 2018, p. 9). A related concept is *big data*, which refers to the use of the vast amounts of data that we now have available and how we analyse it (McAfee et al. 2012).

AI has long existed, and the quest for intelligent machines started with Alan Turing's *Computing machinery and intelligence*, released in 1950 (Turing 2009). In this article, he describes a test to determine whether or not a machine is intelligent. The test is known as the *Turing test* and is passed by a computer that can make a human being think that it is *not* a computer when interacting with it (through a chat, for example). Turing himself worked on

intelligent machines. During the Second World War, he attempted to crack the German "Enigma"-code and succeeded, by using a computer (Copeland 2014).

Since Turing, machines have accomplished various daunting tasks. For the general public, a major milestone was when Deep Blue beat Garry Kasparov in chess in 1997 (Campbell et al. 2002). Deep Blue was a computer developed by IBM. Only 5 years earlier, Kasparov had "scorned the pathetic state of computer chess" (Kurzweil 2015, p. 148). Chess is widely considered a worthy intellectual pursuit, and when a computer now proved to be better than the best of humans, its intelligence could hardly be questioned. Its superiority to human beings, however, could still be challenged. After all, chess is chess, but we still had the game of Go! Go is an ancient board game, widespread in Asia in particular. When Deep Blue beat Kasparov, even a novice human Go player could beat the best Go computer. This game was far more complex, and the brute force approach of Deep Blue and early AI was no match for the more advanced cognitive approach of human beings. Then, in 2016, the comfort some may have taken in the fact that human intuition beat computer intelligence in Go was shattered. *AlphaGo* beat Lee Sedol, widely considered one of the best Go players of recent times (Chouard 2016). *AlphaGo* is Go-playing adaption of Google's *DeepMind* AI, the self-titled "world leader in artificial intelligence research and its application for positive impact" (Google 2018).

While these prominent peaks of AI achievements may be important, it is even more important to examine the widespread and low-key proliferation of the use of AI and machine learning in today's society. We face its applications every day, through "speech recognition, machine translation, spam filters, and search engines" (Brundage et al. 2018). Computers are "flying and landing airplanes, controlling the tactical decisions of automated weapons, making credit and financial decisions, and being given responsibility for many other tasks that used to require human intelligence" (Kurzweil 2015, p. 148). Cars driven by computers, machines helping doctors, phones that recognise us while drones can be programmed to plant trees, give aid, or deliver bombs. In many respects, the future is already here.

Our children are playing with computers from infancy, both on screen, and with robots like Furby, Zhu Zhu Pets, Tamagotchi and the like; even our elderly get care through robotic seals. When online, we often get help from chatbots that appear human, and people even talk to their watches and phones. *Siri*, they call her, or something else if they use another system. Even *love* is found in the computer. At least the more superficial kind, though sex robots like Roxxxy, in addition to providing physical love, also carries conversations (Scheutz and Arnold 2016). There are now brothels staffed by robots, looking to go global (Lockett 2017).

The possible applications of intelligent machines are practically limitless. What we decide to let the computers do, however, is still up to human beings. Thus far, it may be argued that we are not showing much restraint. In the future, "advanced AI holds out the promise of reducing the need for unwanted labor, greatly expediting scientific research, and improving the quality of governance" (Brundage et al. 2018). Reducing the need for labour, while possibly increasing the opportunities for love and companionship? What is not to love?

Considering the rapid proliferation of the use of AI, more and more now highlight the potential malicious use of AI. Some of these dangers are threats to digital security, physical security and political security (Brundage et al. 2018). When our lives are thoroughly digitised, computer hackers can cause great damage. Physical security may be threatened by the use of autonomous weapons, or even the disruption of vital social services. A scary example of these kinds of threats was *Stuxnet*, a piece of malware used to attack Iranian nuclear facilities in 2007 (Langner 2011). The dangers to the political sphere can be seen in "privacy-eliminating

surveillance, profiling, and repression, or through automated and targeted disinformation campaigns" (Brundage et al. 2018).

Koestler (1967) wrote of the *Ghost in the machine*—an allusion to Descartes and the spirit within a physical body. Here, he discusses the possibility that humanity may be an evolutionary dead end. In light of the developments here discussed, it is interesting to see this in light of the possibility that the technological progress we praise may be part of some human "built-in error or deficiency which predisposes him towards self-destruction" (Koestler 1967, p. xi).

Bostrøm (2014) has written of one such possible error: the possibility that we build computers more powerful than ourselves, which we in turn become dependent on. This, he says, would be much like how gorillas—physically superior to us—are now more dependent on us than on themselves (Bostrøm 2014, p. 9). Müller and Bostrom (2014) interviews a group of AI experts on their take on future developments in the field. When asked what year they thought it might be a 50% chance that a "high-level machine intelligence" (HLMI) exists, the median results given are around 2040, with the mean significantly higher at between 2072 and 2097 for the various groups surveyed. A HLMI is an AI that "can carry out most human professions at least as well as a typical human" (Müller and Bostrom 2014).

Superintelligence and the *singularity* are often discussed together. The singularity is "a future period during which the pace of technological change will be so rapid, its impact so deep, that human life will be irreversibly transformed" (Kurzweil 2015, p. 146). *Ultraintelligence* is another term sometimes used to describe "a machine that can far surpass all the intellectual activities of any man however clever" (Good 1965, p. 33). Good (1965) goes on to describe how this machine would improve upon itself, leading to a situation like the *singularity*, in which ever smarter computers build ever smarter computers, leaving mankind behind in the dust.

Science fiction, you say? Let us return to AlphaGo. When this AI beat Lee Sedol in 2016, it had learnt what it knew by learning from human experts and reinforcement learning from self-play (Silver et al. 2017). In 2017, the team behind AlphaGo published an article called *Mastering the game of go without human knowledge*, where they reveal AlphaGo Zero (Silver et al. 2017). This new version of AlphaGo is given only the rules of the game, and learns everything from scratch, a true *tabula rasa* go player. The result? AlphaGo Zero spent just 3 days of teaching itself the game before it beat the version that the year before had made news worldwide by beating the human champion. Its record against the previous version? 100–0 (Silver et al. 2017). Unhampered by human "expert" judgement, the AI takes off. These kinds of developments make it hard to completely dismiss the notion of exponential growth and ultraintelligent machines making or training ever smarter machines.

Chalmers (2015) argues that we need not necessarily worry. We could *integrate* ourselves with computers and harness the benefits they give: uploading ourselves into the digital world and being one with the computers (Chalmers 2015, p. 200). Damasio comments on this possibility in his latest book *The Strange Order of Things,* and says that it is "an implausible scenario" that "reveals a limited notion of what life really is and also betrays a lack of understanding of the conditions under which real humans construct mental experiences" (Damasio 2018, loc 2936). Apart from integration, Chalmers mentions the other possible strategies we could take towards the coming of a singularity, namely *extinction, isolation* or *inferiority* (Chalmers 2015, p. 199). The first involves us going extinct, the second that we limit our interaction with computers as much as possible, while the third means that we interact and that humans are clearly the inferior party in the interaction (Chalmers 2015, p. 199). Chalmers sees *integration* as the best, or perhaps only, choice (Chalmers 2015, p. 200).

What Does It Mean to Be Human?

> Farm animals were a sort of inferior class, reassuring the humblest rural worker that he was not at the absolute bottom of the social scale, a consolation which his industrial successor was to lack (Thomas 1983:50).

As the "humblest rural worker" in days of yore needed consolation and reassurance, so may be the case of the workers of today. Kurzweil (2015) refers to people that are alarmed by the rise of intelligent machines and "have expressed the wish to remain 'unenhanced' while at the same time keeping their place at the top of the intellectual food chain" (Kurzweil 2015, p. 167). A member of the English House of Lords' Artificial Intelligence Committee, Steven Croft, pointed to the fact that more "mundane tasks" will be lost to automation, and that this might lead to an identity crisis (Campbell 2017). That personal identity is connected to jobs, and the feeling of value connected to not being "at the absolute bottom of the social scale", is not the end of this story. We may imagine even bigger challenges to our identity coming, as we are challenged by machines that rival us in some of the areas that we most value. An obvious example: should the process Chalmers (2015) described occur, where we *integrate* ourselves with machines and become uploaded selves in a digital universe, it is not very far-fetched to assume that we will experience uncertainty about the meaning of *being human*. The possibility of very smart machines, he says, "forces us to think hard about values and morality and about consciousness and personal identity", and "brings up some of the hardest traditional questions in philosophy and raises some new philosophical questions as well" (Chalmers 2015).

Human-computer interaction (HCI) is a field where the relationship between human and computes are studied. Norman (1991) is a proponent of the cognitive approach in the field, and he explains that there are two different view of artefacts—devices that "maintain, display or operate upon information" in order to, for example, assist us in cognitive tasks. The *system view* involves us seeing the actor, the task and the artifice as a whole, whereas the actor's capacity is *increased*. From the *personal view*, however, which is the view of the actor, his capacity is not enhanced by the artefact, but the task itself is *changed* (Norman 1991). If we take this approach to the introduction of artefacts in general, we see that the different perspectives give us different views on issues of automation and the introduction of AI in industry. From one perspective, humans are empowered, but from the other the tasks are changed, and the actor may even feel diminished. What is interesting is that the capacity of AI may be said to have gone beyond the role of the cognitive artefacts here discussed. AI systems are often less of a help, or tool, for human actors, but more of an autonomous replacement.

Activity theory is a different approach to HCI, in which both of Norman's views are considered *personal* (Kaptelinin 1996). In activity theory, tools are seen to empower, and even *change*, the actor. Kaptelinin (1996) refers to studies that show that we often go through three phases when tools are used to assist us in tasks. First, we cannot effectively use to tool, so performance of the task is the same with or without the tool. In the second phase we perform better with the tool than without. The third phase is the most interesting, and that is when we can perform the original task better than before, even *without* the tool (Kaptelinin 1996). Using the tools actually changes us and teaches us new things. It empowers us. I will return to this possibility later on.

Penrose (1999) has discussed the fact that having machines that challenge us mentally is something different than them challenging us physically. While being relieved of physical

labour is welcomed, being challenged intellectually is worse, as thinking "has been a very human prerogative" (Penrose 1999, p. 66). The question that follows is one of superiority, as if "machines can one day excel us in that one important quality in which we have believed ourselves to be superior, shall we not then have surrendered that unique superiority to our creations" (Penrose 1999, p. 67)? Gillies (1996) discusses these issues and relates his discussion to debates in the decades preceding his book *Artificial intelligence and scientific method*. He relates, among other things, opinions that placed superiority in chess well beyond the scope of any future advances in AI. Writing in 1996, he himself states that "if the progress of the last twenty or so years continues, then it is not unlikely that, within a few decades, computer chess programs will exist which *no* human is capable of beating" (Gillies 1996, p. 117). As we have seen, *Deep Blue* beat Garry Kasparov in 1997. Similarly, Kurzweil (2015) predicts that "within several decades information-based technologies will encompass all human knowledge and proficiency, ultimately including the pattern-recognition powers, problems-solving skills, and emotional and moral intelligence of the human brain itself" (Kurzweil 2015, p. 148).

Without getting too technical, let us, for our current purposes, assume that superhuman intelligence in machines is a future possibility. Intellectual inferiority, then, and a potentially very troubled relationship with what we have made. Much like in Mary Shelly's *Frankenstein*, the things we make may come back to haunt us (Shelley 2003). Will the superintelligence Bostrøm (2014) discusses be our Frankenstein? Will it cause us grief as we watch it surpass us, or will we simply welcome these new entities that can perform ever more impressive tasks for us? We may, according to activity theory, view ourselves as *empovered*, and simply in need of a reorientation with regard to our relationship with our tools (Kaptelinin 1996).

The subtitle of Shelley's book is *The Modern Prometheus* (Shelley 2003). This is a reference to the Greek myth of the Titan Prometheus, who is credited with creating man from clay. In doing so, Prometheus steals fire from the gods and gives it to humans. This fire brings not only life to man but "also the subtler fire of reason and wisdom from which all aspects of human civilization is derived" (Raggio 1958). The gods were angry, however, and sent "fever and disease to the human race" while condemning Prometheus to torture (Raggio 1958). There are many variations of this myth, and perhaps modern man is living one of them, making new forms of life and giving them reason and wisdom. While reaching for that divine fire, we must hope we are not overreaching.

I will not go into much detail relating to the nature of consciousness or the cognition and mental lives of robots in this paper. Instead, I focus on what they can *do* and how *we* see and relate to them. But, what does all this mean for what it means to be human? Whatever goes on inside their "minds", smart robots change the way we relate to robots, and they also change the way we see ourselves. According to Goldberg (2009) this is a two-way deal, where much of our knowledge of the brain has been presented through computer analogies, while the computers we build are modelled on what we know of neurology and biology (Goldberg 2009, p. 279). Koestler thought that science was "dizzy with its own success" and "forgot to ask the pertinent questions"—questions that could not be answered "so long as one's image of man is that of a conditioned reflex-automaton" (Koestler 1967, p. xi-xii). Turkle (2011) proposes that this way of seeing humans may be promulgated by smart robots—creatures that we somehow identify with or liken to ourselves or other biological creatures. Returning to Koestler, the behaviourism he so intently criticised would be a good science for conflating human-machine working (Koestler 1967). With behaviourism, we "did away with the concept of mind and put in its place the conditioned-reflex chain" (Koestler 1967, p. 6).

How, then, do we understand ourselves in relation to these new entities—already smart, but potentially *very* smart in the not so distant future?

Are We Really Special?

Throughout human history, many attempts at answering the question of what makes human beings special have been made. What is it, for example, that separates man from animals? This question has given rise to many answers, among them that man is "a political animal (Aristotle); a laughing animal (Thomas Willis); a tool-making animal (Benjamin Franklin); a religious animal (Edmund Burke); and a cooking animal (James Boswell, anticipating Lévi-Strauss)" (Thomas 1983:31). First, we will have a brief look at the philosophical debate, before moving on to some psychological perspectives.

The Philosophical Debate

This debate has gotten new prominence today, as it is clearly relevant to the question of what intelligent machines *are*. If reason, for example, is what defines man, a computer that satisfies our definitions of rationality may be considered a man? Or, if our distaste for this conclusion is too strong, do we revise our definitions to exclude machines, or perhaps search for other defining aspects of humanity? Turkle (2011) describes this as happening when primitive robots started showing what appeared to be emotions. Man was previously defined as a *rational* animal, according to the Aristotelian method of focusing on exclusive characteristics. Then, when computers could not be denied rationality, we suddenly became *emotional* animals. Cole (1990) proposed that what sets humans apart is our ability to use artefacts to modify our environment and to transmit these modifications to subsequent humans through language. These conditions, too, can surely be fulfilled by smart computers.

Regardless of how we approach this issue, intelligent machines are clearly a challenge to the belief that we are special, or in some way superior to (all) other things. This challenge is less of a problem for those that do not adhere to some form of anthropocentrism. Many environmental philosophers, like Arne Næss, adamantly dethroned man and put him squarely on the level of the rest of the natural world—promoting *biospherical egalitarianism* as an alternative to anthropocentrism (Næss 1999). The biosphere does not include machines, though, so these philosophies will also have to be rethought if we are to deal adequately with the possible moral value of intelligent machines. As cybernetics and biotechnology continues to evolve, it will be even harder to exclude objects that are not what we traditionally consider *biotic* from moral consideration.

The philosophical debates relating to the issues of technology are often linked to the terms *transhumanism* and *posthumanism.* Transhumanism is mainly concerned with the possibility of *enhancing* humanity through, for example, "regenerative medicine to nanotechnology, radical life extension, mind uploading and cryonics, among other fields" (Ferrando 2013). Posthumanism is about the "decentering" of man and "raises serious questions as to the very structures of our shared identity—as humans—amidst the complexity of contemporary science, politics and international relations (Braidotti 2013, p. 2).

We may forge new rules that excludes machines by definition, but that does not preclude a possible identity crisis for humanity. What are we, when we are not the most intelligent or most powerful beings we know of? What are we, when some of us wants to be with their robots instead of other people? Our inferior physical power was most likely acknowledged as soon as

we reflected on the strength of the beasts we came across. This was perhaps not that much of a defeat, since we also realised that we were mentally superior. We could use our cunning to overcome physical disadvantage, and soon we ruled the earth. Ingenuity, intelligence, language and creativity made humans a "huge, dangerous and powerful" force, and the Earth now suddenly appears to be both small and fragile (Randers 2012).

Psychology and Humanity

The term *conditioned reflex-automaton* was rejected by Koestler as a proper description of man (Koestler 1967)? But is it suitable for AIs? Behaviourism can explain a lot, Koestler says, but perhaps not "scientific discovery and artistic originality" (Koestler 1967, p. 13). Watson and Skinner may have reduced us to little more than rats, but Goldberg (2009) points out that our frontal lobes was vastly different. Behaviour can be observed, whereas the soul, mind and consciousness are vague terms; "[c]onsciousness, Watson objected, is 'neither a definable nor a usable concept, it is merely another word for the "soul" of more ancient times. … No one has ever touched a soul or seen one in a test-tube. Consciousness is just as unprovable, as unapproachable as the old concept of the soul.'" (Koestler 1967, p. 15). These terms are rejected by those seeking an objective science of psychology. Behaviour and phsyiology are easier to measure and deal within a seemingly objective fashion. *Ratomorphy* is the term Koestler uses for denying humans any faculties not present in rats—the opposite of the anthropomorphic fallacy where we give animals human qualities (Koestler 1967, p. 17). Robotomorphy may be the applicable term in our context, if we end up viewing man as little more than an advanced computer. And if we do, a classical quote from Sir Cyril Burt may be appropriate:

> The result, as a cynical onlooker might be tempted to say, is that psychology, having first bargained away its soul and then gone out of its mind, seems now, as it faces an untimely end, to have lost all consciousness (Burt 1962, p. 229).

Antonio Damasio is of interest when discussing the differences between man and machine. In his famous book *Descartes' Error* (1994), he discusses the nature of rationality, and argues that it cannot be separated from *emotions*. Let us say that a person has "knowledge, attention, and memory", language and logic, and the ability to perform calculations. Damasio mentions a patient like this, seemingly well equipped to for rational thought, yet still somehow unable to achieve it. The patient had a brain damage relating to the processing of feelings, and this caused "flawed" reasoning. To Damasio, it became clear that "feeling was an integral component of the machinery of reason" (1994, p. xii). He acknowledges that emotions and feelings also can have *negative* influences on rational thinking but makes the point that their absence also has dramatic consequences (Damasio 1994, p. xii). In *Looking for Spinoza* (2003), he updates his theories with new evidence, and considers feelings themselves, not necessarily relating to decision-making. His view is that "feelings are the expression of human flourishing or human distress, as they occur in mind and body" (Damasio 2003, p. 6). Damasio is most famous for his *somatic-marker hypothesis*, discussing how rational deliberation activates "gut feelings" that guide us in the process of deliberation (Damasio 1994, p. 173). The somatic markers "forces our attention on the negative outcome to which a given action may lead, and functions as an automated alarm which says: Beware of danger ahead if you choose the option which leads to this outcome" (Damasio 1994, p. 173). In a recent paper, Maldonato and Valerio (2018) discuss the same topic, and says that "emotions are crucial in

moral decisions and that their understanding may help us to avoid mistakes in the construction of hybrid organisms capable of autonomous behavior" (Maldonato and Valerio 2018).

Is this how we are able to keep the barbed wire fence between man and other things intact, while keeping the criteria of *rationality?* Machines have just about everything required for reason, but not *emotion*s? Damasio discusses what is required for emotions, and starts with the requirement that "an entity capable of feelings must be an organism that not only has a body but also a means to represent that body inside itself"; plants have "bodies", but no way to mentally represent or understand their bodies, and thus no emotions (Damasio 2003, p. 109). Secondly, body states must be mapped, and the organism must be able to translate these body states into "mental patterns or images" (Damasio 2003, p. 110). Thirdly, he demands consciousness in return for emotions; "in plain terms, we are not able to feel if we are not conscious" (Damasio 2003, p. 110). Lastly, "the brain of an organism that feels creates the very body states that evoke feelings as it reacts to objects and events with emotions and appetites. In organisms capable of feeling, then, the brain is a double necessity" (Damasio 2003, p. 110). It must both create the "body maps" that relate to various feelings, and it must also "command or construct the particular emotional body state that ends up being mapped as a feeling" (Damasio 2003, p. 110). He concludes that "[m]ost animals with complex brains satisfy these conditions, in all probability" (Damasio 2003, p. 110).

Cominelli et al. (2018) picks up on Damasio's theory of mind and discusses the possibility of artificial intelligence with social and emotional intelligence. They present a "SEAI (Social Emotional Artificial Intelligence), a cognitive system specifically conceived for social and emotional robots. It is designed as a bio-inspired, highly modular, hybrid system with emotion modeling and high-level reasoning capabilities" (Cominelli et al. 2018, p. 1). They describe their model in detail, and while they conclude that there are "still some shortages", it seems AI cannot be expected to stay unable to approach some of the ways human beings work, and the way emotions guide our rationality. They claim to provide "clear demonstration of how SEAI and the chosen 'understanding by building' approach lead to an important confirmation: with SEAI, robots can benefit from their own artificial emotions for taking decisions and treasure their past interactions" (Cominelli 2018, p. 18). So, conscious and emotional robots are on their way, and as these scientists also conclude, "ethical issues will become extremely relevant and critical" (Cominelli 2018, p. 18). In a similar vein, Maldonato and Valerio (2018) claim that the "encounter between Humans and very powerful AI will lead, in the near future, to organisms capable of going over the simulation of brain functions: hybrids that will learn from their internal states, will interpret the facts of reality, establish their goals, talk with humans and, especially, will decide according to their own 'system of values'" (Maldonato and Valerio 2018). Scheutz (2016) echoes these sentiments and makes the case for the development of robots with "moral capabilities". This is necessary, he says, because "any ordinary decision-making situation from daily life can be turned into a morally charged decision-making situation" (Scheutz 2016, p. 515).

Damasio (2018) comments on these developments, but is sceptical of its prospects. We may have robots with something *like* feelings, "arising from an artificial substrate provided they would be reflections of 'homeostasis'", but "there is no reason to expect that such feelings would be comparable to those of humans or those of other species" (Damasio 2018, loc 2973). Turkle supports these sentiments, and states that robots "has no feelings, can have no feelings, and is really just a clever collection of 'as if' performances" (Turkle 2011, p. 6). *Authenticity* for her requires "the ability to put oneself in the place of another, to relate to the other because of a shared store of human experiences: we are born, have families, and know loss and the

reality of death" (Turkle 2011, p. 6). Damasio is more technical and names "the substrate that feelings actually use" as the crucial missing factor (Damasio 2018, loc 2973). These underlying substrates is also what makes Damasio reject the perspective that man is nothing but algorithms (Damasio 2018, loc 2951). Substrates, and the *living* body, as he explicitly states that "to build feeling, on the other hand, we require a living body" (Damasio 2018, loc 3974). Damasio also has another argument against the algorithmic view of man, in that that it "does not advance the human cause", something many others would argue that we should not require directly from science (Damasio 2018, loc 3974).

We saw Gillies (1996) predict the rise of a computer chess master within decades, and it arrived the very next year. Similarly, the timing of the arrival of intelligent computers that satisfy various conceptions of *consciousness* and *emotion* is hard to predict. What seems clear is that if our definitions revolve around description of *function*, and not biology of physiology, computers will get there. Probably sooner than we imagine, if we are to believe Kurzweil's (2015) description of the exponential progress we are in the midst of.

What is even more important to us than these technical aspects is that most human beings consider smart robots as *alive enough* to warrant both emotional reactions and ethical consideration (Turkle 2011, ch. 2). When we already consider them to be *some sort of life*, and people in experiments experience discomfort when a toy Furby appears to be hurt, we can assume that the problems of meaningfully separating ourselves from smart computers will be an increasingly tall order (Turkle 2011, p. 45).

We may, of course, choose to take the approach that computers are not a challenge to us, but somehow an *extension* of us—our tools that empower us. This can, for example, be based on activity theory (Kaptelinin 1996). Such an approach would not, however, change the fact that many struggle to perceive the meaningful difference between us and these tools—intellectually, emotionally and morally. A further challenge in this respect will be the merging of the technological and the biological, a topic I will not discuss in this paper.

Digital Technology and Human Relationships

> Technology is seductive when what it offers meets our human vulnerabilities. And as it turns out, we are very vulnerable indeed. We are lonely but fearful of intimacy (Turkle 2011, p. 1).

As computers become more intelligent, we interact with them in new ways. The interactions themselves are not necessarily new, but they were previously reserved for relations between humans and other living things. Things like family, friends, pets and other biological beings. This development is not entirely new, but as AI becomes increasingly powerful, the part that technology can play in changing our social lives dramatically increases.

In movies like *Her* from 2014, the love between a machine and a man is portrayed. Movies, however, are movies, and real life is real life? People surely do not form those kinds of bonds to artificial entities? David Levy would disagree, and he argues his case in *Love and Sex with Robots* (Levy 2008). Robots will teach us many things, he argues, and they are in many ways *better* than human mates as they are both faithful, caring and able to teach us and interact with us in ways human beings cannot (Levy 2008). Unlimited attention, patience and accommodation—what is not to love in such a partner? Turkle discusses the possibility of being against human-computer relations as equally reactionary in the future as today's current opponents to

same-sex marriage, etc. (Turkle 2011). When a reporter calls her about Levy's *Love and Sex with Robots* (Levy 2008), she discusses her reservations and is surprised when "the reporter suggested that I was no better than bigots who deny gays and lesbians the right to marry" (Turkle 2011, p. 7).

Less objectionable to most may be the idea of using AI for companionship and entertainment. Relationships that are non-intimate, that is. Less objectionable, because we already have this, and we have had them for a long time.

Fong et al. (2003) describes "socially interactive robots" as "robots for which social human-robot interaction is important" (Fong et al. 2003, p. 1). From the early beginning in the late 1940s, we have gotten very advanced robots, whose main function is to interact socially with humans, like MIT's *Kismet* and *Cog* (Fong et al. 2003, p. 4). A step back from these advanced robots, however, we have had several very popular social *computers* (not always embodied) in our societies. The *Tamagotchi* was perhaps the first, when it swept across the world in the 1997? The Tamagotchi had no body but lived on a screen in an egg. Its user had to feed it, clean up after it, and entertain it, and people everywhere were thrilled to see their Tamagotchis "grow up" with personalities shaped by their "parent's" actions. The thrill was real, and since people cared for them, they also grieved when they died (Turkle 2011, p. 34). Another big success was the *Furby,* released in 1998. The Furby *did* have a body, required care and attention, and even seemed to learn English from its owners. The Furbys and Tamagotchis were "the first computers that asked for love" (Turkle, 2011, p. 30). Furbys could not hear and did not "learn" in that way, but the kids that used them thought they did (Turkle 2011, p. 38). *My Real Baby,* also a Hasbro product, launched in 2000, was an intelligent baby doll, going through the development from 0 to 2 years in their owner's care (Turkle 2011, p. 67–8). It was not a huge success, but it showed some of the potential in these kinds of machines. The last example we should mention is *Aibo* and *Paro*.

Aibo takes us beyond the world of toys and into the world of pet replacement (Sony 2018). It is sold as "full of charm" with "[r]ound, alluring eyes with a powerful pull, a cute, roly-poly form, moving around with infectious energy, and an identity …" (Sony 2018). People bond with these "pets"—feeling guilty when not entertaining them and worrying about their health and well-being (Turkle, 2011, p. 57). If we are to believe the producers, Aibo *loves* humans, too, as "[b]eing with people is what aibo loves best" (Sony 2018).

Paro is perhaps the most interesting, as it is sold as a "therapeutic robot", and employed in many elder care facilities around the world (Paro Robots 2018). Paro provides the "documented benefits of animal therapy", even if it is not actually an *animal* (Paro Robots 2018). Paro remembers the name people give it, and even remembers the actions it did before being petted—trying to repeat these actions with the goal of behaving "the way the user prefers" (Paro Robots 2018). Stress and loneliness is relieved in the elderly and demented, time is freed for the caretakers, and the family members' conscience is a bit clearer when leaving their elders with a smile and a seal on their lap. Win, win, win?

The problem is, however, that while these robots are not designed for that more superficial kind of intimacy, they do create bonds that can easily be construed as *intimate*. Turkle (2011) tells many stories of how even simple computers create emotional responses in us, and how we grow to care for them. We even grieve when they die, and are reluctant to let them go, or leave them. Even if we *know* that they are not real and have no proper feelings, they look at us, ask for nurture and provide companionship. This makes them "alive enough" for people to bond, which is the main point Turkle (2011) emphasises as she asks us to think about what we are doing with these machines. The question, then, is what happens to us? What happens to our relationships when we bond with computers, and even replace human contact with contact with machines?

Turkle is concerned, as a psychoanalytically trained psychologist valuing "relationships of intimacy and authenticity", about the prospects of intimate relationships with computers. Computers "have no feelings, can have no feelings, and is really just a clever collection of "as if" performances, behaving as if it cared, as if it understood us" (Turkle 2011, p. 6). Damasio supports this view, when stating that "lifelike, humanlike" are "an absurd and non-existing myth", and that "[h]umans have life and have feelings, and such robots have neither" (Damasio 2018, loc 3058). This goes back to the fundamental question of what machines really are, and what makes for *authenticity*. After all, what are feelings, apart from chemical reactions in our bodies? Is this really fundamentally different from the "feelings" or actions of an AI, influenced by impulses and reacting on them? Learning, evolving, and developing new reactions as it gathers experience? If we see humans this way, there is little danger in interacting with other automata. If, however, we see ourselves as *more*, as Turkle (2011) does, the interactions we increasingly have with computer may *diminish* us. Perhaps it will even make us more like automata, just by interacting in such a way and accepting that our companions do.

For Turkle, an authentic relationship "involves coming to savor the surprises and the rough patches of looking at the world from another's point of view, shaped by history, biology, trauma, and joy" (Turkle, 2011, p. 6). She concludes that robots cannot do this. We, however, may argue that this conclusion is premature, Metzler et al. (2016) does when they comment on Turkle's position and ends up saying that robots *might* become authentic companions. The progress on this field is so rapid that it is difficult not to imagine a future where some will argue that robots *are* shaped by their experiences, and we have already seen that computers are now built with the intimations of *emotions* (Cominelli 2018). Furthermore, as we have seen, it is also possible to view human beings as little more than machines reacting to external impulses—much the same as robots.

Turkle (2011) calls our acceptance of these social machines "the robotic moment" and that "we don't seem to care what these artificial intelligences 'knows' or 'understands'" (Turkle 2011, p. 9). While children previously projected themselves upon inanimate objects like dolls, we treat these machines as subjects—"Rorschach gives way to relationship" (Turkle 2011, p. 95). It is tempting to propose the argument that we never really *know* if other people really know or understand us, either. Following Descartes, we could not say that all I *really* know is that I myself exist?

The main question Turkle seems to ask is this: "What if a robot companion makes us feel good but leaves us somehow diminished" (Turkle 2011, p. 6)? Companionship feels good, and we do not seem to object to robots replacing loneliness – or even human beings. "People are lonely", Turkle says, but in always finding something to let us escape that loneliness, "we may deny ourselves the rewards of solitude" (Turkle 2011, p. 3). What rewards are these? One is to be able "to reflect on one's emotions in private" (Turkle 2011, p. 176). Erich Fromm (1994) describes the need for relatedness to the outside world, and the fact that "mental disintegration" follows loneliness and feelings of isolation (Fromm 1994, p. 17). But what if "loneliness is just failed solitude" (Turkle 2011, p. 288)?

There are others that agree with Turkle that solitude has its rewards, both for the individual and society. Solitude has been a topic in philosophy since early modernity, especially since population growth lead more and more people to live in densely populated areas (Thomas 1983, p. 268). For Thomas Hobbes, for example, solitude was horrible, while philosophers coming after him, like Rousseau and J. G. Zimmermann, wrote extensively on the benefits of

solitude (Thomas 1983, p. 268). One clear statement on the benefits of solitude came from John Stuart Mill:

> A world from which solitude is extirpated is a very poor ideal. Solitude, in the sense of being often alone, is essential to any depth of meditation or of character; and solitude in the presence of natural beauty and grandeur, is the cradle of thoughts and aspirations which are not only good for the individual, but which society could ill do without (Mill 2004, p. 692).

Solitude gives us the "space to think our own thought" and it "refreshes and restores" (Turkle 2011, p. 204, 288). According to Anthony Storr "by far the greater number of new ideas occur" during states akin to solitude (Storr 2005, p. 198). But solitude seems hard to come by these days. Young people experience *discomfort* when their phones are not around, and that many modern people "trained by the Net" cannot find solitude even in nature (Turkle 2011, p. 289). If "[l]oneliness is failed solitude" people seem to be failing in record numbers if one looks at reported loneliness combined with the purported difficulty of finding solitude (Turkle 2011, p. 288).

On a final note, it is interesting to see that there are developments in Turkle's own discipline of psychology towards robot therapists. By robot therapists I am not talking about therapeutic robots like Paro, but robots that communicate with *words*. While communication technology lets traditional therapy be provided over great distance, and without synchronicity, we now go even further and remove the human therapist altogether (Ebert et al. 2015).

Practice Makes Perfect, *or* Machine Evolved/Man Devolved

> We have to love our technology enough to describe it accurately. And we have to love ourselves enough to confront technology's true effects on us (Turkle 2011, p. 243).

Machine learning refers to the fact that computers learn and evolve through *training* and *exposure*. What if human beings actually learn in a similar fashion? What if we learn by being exposed to intellectual challenges. If this is so, what are the implications of letting more and more of the most challenging intellectual work being done by computers?

Once upon a time, pupils were required to learn the multiplication table. They were even able to do manual calculations. Perhaps they still do, but nowadays our young may be the only ones skilled at manual calculation, as they are the only ones forced to perform it on a regular basis. The rest of us have mostly lost the skill, due to the ubiquitous calculator—ever present on our computer, phone or watch. In fact, I can just ask Siri to perform the calculations I need. Research shows that children's mathematical abilities are *not* negatively affected by calculators, but manual calculation abilities would be expected to suffer (Barton 2000). This goes well with research on the effects of "brain training", where there is much evidence on the possibility of improving performance on certain tasks through training, but that *transfer* to other tasks is hard to prove (Nouchi et al. 2013; Simons et al. 2016).

If this process is indicative of how our brain works in other ways as well, we may be poised for a somewhat sad future. If we leave the advanced intellectual challenges and thinking to computers, what will happen to our cognitive abilities? We may, according to activity theory, become empowered and learn to perform the tasks *better* after a period of using the tools (Kaptelinin 1996). This does not seem likely to occur if we use computers to autonomously

replace us, instead of merely assisting us with the tasks performed. Norman's (1991) idea from the *personal view* of cognitive artefacts, that tasks *change* and that the capacities humans must then use is affected, seems more relevant in this case. This means that we will practice other abilities than before, and the net effect may of course be both positive and negative. What is important here is to make sure that we understand *how* we are changed, and that we do not let ourselves loose, or diminish, abilities that we value and consider important.

A different but related topic is what happens to our brains if we change our social lives and replace humans with robots? The neuropsychologist and cognitive neuroscientist Goldberg (2009) asks whether "social stimulation is to the development of the frontal cortex what visual stimulation is to the development of the occipital cortex?" (Goldberg 2009, p. 174). The question being: will our brain develop differently if we change the way we socialise? Here, Turkle's statement that when "you practice sharing 'feelings' with robot 'creatures,' you become accustomed to the reduced 'emotional' range that machines can offer", becomes relevant (Turkle 2011, p. 125). The frontal cortex is also called the *executive brain* and is linked with issues of self-control and "intentionality, purposefulness, and complex decisions making" (Goldberg 2009, p. 4). Furthermore, the frontal lobes are said to be what "make us human" and "the organ of civilization", as no other species develop these lobes like us humans (Goldberg 2009, p. 4). It would be paradoxical, then, if we used our evolved executive brains to develop robot companions, whose companionship in turn prevents the development of this part of the brain. This, again, could by the very same line of reasoning lead us to the outrageous proposition that socialising with robots may lead to unintelligence and *decadence*. Even *morality*, Goldberg (2009) argues, is linked with the frontal lobes (Goldberg 2009, p. 17).

Does this lead to a situation in which we retard our executive functions, much like the degeneration of our ability to do manual calculations after the arrival of calculators? Carr (2011) has written on the effect the Internet on our brains. He cites the sceptics who say that our new way of connecting and consuming media leads to a new "dark age of mediocrity and narcissism" (Carr 2011, loc 116). Turkle (2011) echoes these sentiments. She describes a *self* that is developed "in the cacophony of online spaces" and in this process tempted "into narcissistic ways of relating to the world" (Turkle 2011, p. 179).

Carr (2011) describes his own feelings of change:

> Now my concentration starts to drift after a page or two. I get fidgety, lose the thread, begin looking for something else to do. I feel like I'm always dragging my wayward brain back to the text. The deep reading that used to come naturally has become a struggle (Carr 2011, loc 158).

He then goes on to describe that this feeling of change is echoed by a lot of people in today's society. One of his main arguments is that our brains are plastic (adult ones too, even if they are *less* plastic), and that the way we use them changes the way they function: "Our ways of thinking, perceiving, and acting, we now know, are not entirely determined by our genes. Nor are they entirely determined by our childhood experiences. We change them through the way we live—and, as Nietzsche sensed, through the tools we use" (Carr 2011, loc 595). Cab drivers' brains change, because they continually store and use so much geographical information; when they start using GPS this does not happen, and they "lose the distinctive mental benefits of that training" (Carr 2011, loc 3594). Another interesting study shows that people that photographed objects in a museum actually *remembered* less of what they saw—as if the brain understood that it need not bother about the things we can use other tools for (Henkel 2014). van Nimwegen (2008) calls this *externalisation*, and refers to

situations in which "the necessity to remember certain characteristics of operators is eliminated or lessened" (van Nimwegen 2008, p. 123). He tested these effects in an experiment where two groups were given different interfaces for a task—one which provided various forms of assistance and one that did not. The users of the interface with less assistance "imprinted relevant task and rule knowledge better and where not affected by a severe interruption in the workflow" (van Nimwegen 2008, p. 125). Professionals are also affected, as a study shows that doctor's use of clinical guidelines and electronic medical records leads to a perceived and experienced "deskilling", involving "decreased clinical knowledge, decreased patient trust, increased stereotyping of patients, and decreased confidence in making clinical decisions" (Hoff 2011). So, research shows that brain training is not an effective means to improving general cognitive abilities, but that they are clearly effective for the specific skills we train (Daniel et al., 2016). In a 2010 article Owen et al. (2010) states the same, and that while "improvements were observed in every one of the cognitive tasks that were trained, no evidence was found for transfer effects to untrained tasks, even when those tasks were cognitively closely related" (Owen et al. 2010, p. 775).

The reasoning is that these sorts of mechanisms apply to most areas of our lives. Furthermore, Carr (2011) states that electronic media is particularly effective at changing our nervous systems, much because they function in similar ways (Carr 2011, loc 3618). It may seem that practice, if it does not necessarily make perfect, at least improves us. The use of computers is *some* kind of practice, but its effects are not necessarily on balance beneficial, as "our ability to learn can be severely compromised when our brains become overloaded with diverse stimuli online" (Carr 2011, loc 3635).

This is also described by Turkle (2011), who describes our ever better multitasking abilities. Once hailed as a great benefit of the digital revolution, we now "know that multitasking degrades performance on everything we try to accomplish" (Turkle, 2011, p. 242). She also describes how the digital consumption of news and other media "always invites you elsewhere" (Turkle, 2011, p. 242). The old-fashioned books invited "daydreams and personal associations" with readers looking into themselves, while the new media is usually fleeting and constantly disrupted by things like messages and Facebook (Turkle, 2011, p. 242). Longing for a past that never existed, you may say, but her musings about how we consume texts echoes Carr's findings, that media "aren't just channels of information" but somehow "shape the process of thought" (Carr 2011, loc 174). Carr dislikes the reshaping the Net has been doing to him, as he sees it as "chipping away at my concentration for concentration and contemplation" (Carr 2011, loc 174). As Turkle puts it, "[a]s we try to reclaim our concentration, we are literally at war with ourselves" (Turkle 2011, p. 296). In the words of Culkin: "We shape our tools and thereafter they shape us" (Culkin 1967, p. 70). Carr rewrites this to become "we program our computers and thereafter they program us" (Carr 2011, loc 3633). While activity theory may see us as *empowered* when we use our new tools to perform better actions*,* we must keep in mind that we are also changed. It seems that we are only at the second stage that Kaptelinin (1996) described—the stage where we perform tasks more effectively with the tools. The third stage, where we achieve increased performance of the tasks even without the tools, may be harder to achieve once the tools we use become as powerful as they are today.

Conclusion: a World for the Brave

We have seen that AI permeates all aspects of modern society and that the development has been rapid. We also know that much of the applications of AI to new fields are done by

corporations and industrious individuals, often in the absence of any law of regulation to control or prohibit it. Why *would* we prohibit it, when it grants so many benefits?

We have seen that there are several reasons to think this through, as we are rapidly changing both what it means to be human, and how human beings function.

The challenge to human identity is a challenge from two sides: automation and displacement and a challenge for mental superiority. I have not focused on jobs lost to smart computers, but chosen to discuss in depth how the progress of artificial intelligence challenges the very notion of what it means to be human. What separates us from these smart computers?

First, it is possible to view human beings as machines reacting to their environments—emotions and all being nothing but chemical reactions. This view portrays man as some sort of machine, and the firing of synapses in the human brain is little more than the electrical impulses that works its magic in a robot. Secondly, we see that they even make artificial synapses now, and emotional robots satisfying pretty strict conditions concerning what is required for something to be considered conscious (Schneider et al. 2018). With this development, it is becoming increasingly harder to set the boundaries between man and smart machine, unless one happens to be religious or spiritual. The *soul* is something computer scientists have not yet claimed to have produced. For the rest of us, however, we might have to come to terms with our creations, which now rival us intellectually in many ways, being considered life of some sort.

One interesting fact is that we *already* consider social robots of various kinds "life of some sort"—meaningful enough to love, to mourn, to consider as moral beings that should not be needlessly tortured, etc. When this happens with Tamagothis and Furbys, there is little to suggest that we would not have increasingly strong relationships with computers—even relations of the most intimate kind.

In the final part of this paper discussed how we must be aware of the fact that training actually matters, also when it comes to matters of the brain. The brain develops as it is used, which means that it will develop differently if we set away various tasks to computers.

Firstly, the development of our brain is influenced by how we interact socially. The latter part was discussed in relation to how we exchange humans for robots in social settings, and how Goldberg (2009) shows that these social relations are important for the development of our "executive brain"—the frontal lobes. Impaired development of our executive brains would negatively influence not only most high-level complex cognitive functions but also moral evaluations.

Secondly, our cognitive functions change according to how we employ them. We have seen that training of specific cognitive skills improves them, but that there is little to suggest much transfer or general effects from such training. These mechanisms point towards increased specialisation, as computer tools help us with many of the general skills that everyone previously had to know, letting us train and become proficient at more expert skills. If so, we are in fact empowered by these tools, in accordance both with activity theory and the system view of the cognitive approach in HCI.

The final point I have discussed is the general development towards better multitasking abilities, at the cost of performance on single tasks. This would seem to go against some of the benefits of specialisation, and might put a question mark after the idea that we are empowered, at least if Carr (2011) and others are right about our decreased attention spans and abilities to focus on complex and challenging tasks.

It might also be worthwhile to consider the long-term effects on human intelligence if we leave the more challenging intellectual tasks to computers. Being scared of some evolutionary retardation down the road may seem far-fetched, but combined with the cognitive changes

described, we are already seeing that the path we were on might not be heading exactly where we would like it too.

The story of the creator that gets into trouble with his creations has been told many times. It us up to us to be aware of what we make, how we employ it, and not least: how it affects us. Should we manage to avoid the pitfalls of Prometheus and Frankenstein's creator, going a little bit slower on our way to the future seems like a small price to pay.

References

Barton, S. (2000). What does the research say about achievement of students who use calculator technologies and those who do not. In Fife (Ed.), *Electronic Proceedings of the Thirteenth Annual International Conference on Technology in Collegiate Mathematics.* Atlanta: Addison Wesley.

Bostrøm, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

Braidotti, R. (2013). *The posthuman*. Cambridge: Polity Press.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B. & Anderson, H. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:*1802.07228.

Burt, C. (1962). The concept of consciousness. *British Journal of Psychology, 53*(3), 229–242.

Campbell, C. (2017). *They will take our jobs! Rising AI to threaten human jobs and cause major identity crisis.* Sunday Express. Retrieved from https://www.express.co.uk/news/uk/846434/Robot-AI-bbc-sunday-Artificial-Intelligence-Committee-Steven-Croft-Edward-Stourton-musk.

Campbell, M., Hoane Jr., A. J., & Hsu, F. H. (2002). Deep blue. *Artificial Intelligence, 134*(1–2), 57–83.

Carr, N. (2011). *The shallows: What the Internet is doing to our brains*. New York City: WW Norton & Company.

Chalmers, D. J. (2015). The singularity: A philosophical analysis. In S. Schneider (Ed.), *Science fiction and philosophy: From time travel to superintelligence* (pp. 171–224). Chichester: Wiley-Blackwell.

Chouard, T. (2016). The Go Files: AI computer wraps up 4-1 victory against human champion. *Nature News.* Retrieved from https://www.nature.com/news/the-go-files-ai-computer-wraps-up-4-1-victory-against-human-champion-1.19575.

Cole, M. (1990). Cultural psychology: A once and future discipline?. In *Nebraska Symposium on Motivation 1989: Cross-cultural perspectives*. University of Nebraska Press.

Cominelli, L., Mazzei, D., & De Rossi, D. E. (2018). SEAI: Social emotional artificial intelligence based on Damasio's theory of mind. *Frontiers in Robotics and AI, 5*, 6.

Copeland, B. J. (2014). *Turing: Pioneer of the information age*. Oxford: Oxford University Press.

Culkin, J.M. (1967). A schoolman's guide to Marshall McLuhan. *Saturday Review*, 51-53, 70–72.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Quill.

Damasio, A. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Orlando: Harcourt Inc.

Damasio, A. (2018). *The strange order of things*. New York: Pantheon Books.

Ebert, D. D., Zarski, A. C., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M., & Riper, H. (2015). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: A meta-analysis of randomized controlled outcome trials. *PLoS One, 10*(3), e0119895.

Ferrando, F. (2013). Posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms. *Existenz, 8*(2), 26–32.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3–4), 143–166.

Fromm, E. (1994). *Escape from freedom*. New York: Henry Holt and Company.

Gillies, D. (1996). *Artificial intelligence and scientific method*. Oxford: Oxford University press.

Goldberg, E. (2009). *The new executive brain: Frontal lobes in a complex world*. Oxford: Oxford University Press.

Good, I. J. (1965). Speculations concerning the first ultraintelligent machine. *Advances in Computers, 6*, 31–88 Elsevier.

Google. (2018). Solve intelligence. Use it to make the world a better place. Retrieved from https://deepmind.com/about/.

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science, 25*(2), 396–402.

Hoff, T. (2011). Deskilling and adaptation among primary care physicians using two work innovations. *Health Care Management Review, 36*(4), 338–348.

Kaptelinin, V. (1996). Computer-mediated activity: Functional organs in social and developmental contexts. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 45–68). Cambridge: MIT Press.

Koestler, A. (1967). *The ghost in the machine*. New York: The Macmillan Company.

Kurzweil, R. (2015). Superintelligence and singularity. In S. Schneider (Ed.), *Science fiction and philosophy: From time travel to superintelligence* (pp. 146–170). Chichester: Wiley-Blackwell.

Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security and Privacy, 9*(3), 49–51.

Levy, D. (2008). *Love and sex with robots: The evolution of human-sex relationships*. New York: Harper Perennial.

Lockett, J. (2017). *World's first brothel staffed entirely by robot sex workers now looking for investors to go global*. The Sun. Retrieved from https://www.thesun.co.uk/news/4131258/worlds-first-brothel-staffed-entirely-by-robot-sex-workers-now-looking-for-investors-to-go-global/.

Maldonato, M., & Valerio, P. (2018). Artificial entities or moral agents? How AI is changing human evolution. In *Multidisciplinary approaches to neural computing* (pp. 379–388). Cham: Springer.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60–68.

Metzler, T. A., Lewis, L. M., & Pope, L. C. (2016). Could robots become authentic companions in nursing care? *Nursing Philosophy, 17*(1), 36–48.

Mill, J. S. (2004). *Principles of political economy*. New York: Prometheus Books.

Müller, V. C., & Bostrom, N. (2014). Future progress in artificial intelligence: A poll among experts. *AI Matters, 1*(1), 9–11.

Næss, A. (1999). *Økologi, samfunn og livsstil*. Oslo: Universitetsforlaget.

Norman, D. A. (1991). Cognitive artifacts. In *Designing interaction: Psychology at the human-computer interface* (Vol. 1, pp. 17–38).

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Nozawa, T., Kambara, T., Sekiguchi, A., Miyauchi, C. M., Kotozaki, Y., Nouchi, H., & Kawashima, R. (2013). Brain training game boosts executive functions, working memory and processing speed in the young adults: A randomized controlled trial. *PLoS One, 8*(2), e55518.

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., Howard, R. J., & Ballard, C. G. (2010). Putting brain training to the test. *Nature, 465*(7299), 775.

Paro Robots. (2018). *Paro Therapeutic Robot*. Retrieved from http://www.parorobots.com.

Penrose, R. (1999). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford: Oxford Paperbacks.

Randers, J. (2012). 2052: A global forecast for the next forty years. *In The Future in Practice: The State of Sustainability Leadership*. University of Cambridge.

Raggio, O. (1958). The myth of Prometheus: Its survival and metamorphoses up to the eighteenth century. *Journal of the Warburg and Courtauld Institutes, 21*(1/2), 44–62.

Scheutz, M. (2016). The need for moral competency in autonomous agent architectures. In *Fundamental Issues of Artificial Intelligence* (pp. 517–527). Cham: Springer.

Scheutz, M., & Arnold, T. (2016). Are we ready for sex robots?. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 351–358). IEEE Press.

Schneider, M. L., Donnelly, C. A., Russek, S. E., Baek, B., Pufall, M. R., Hopkins, P. F., Dresselhaus, P. D., Benz, S. P., & Rippard, W. H. (2018). Ultralow power artificial synapses using nanotextured magnetic Josephson junctions. *Science Advances, 4*(1), e1701329.

Shelley, M. W. (2003). *Frankenstein*. New York: Barnes & Noble Classics.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., & Chen, Y. (2017). Mastering the game of go without human knowledge. *Nature, 550*(7676), 354.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest, 17*(3), 103–186.

Sony. (2018). *aibo*. Retrieved from https://aibo.sony.jp/en/.

Storr, A. (2005). *Solitude: A return to the self*. New York: Free Press.

Thomas, K. (1983). *Man and the natural world: Changing attitudes in England 1500–1800*. Middlesex: Penguin Books.

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test* (pp. 23–65). Dordrecht: Springer.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

van Nimwegen, C. (2008). The paradox of the guided user: Assistance can be counter-effective. Utrecht University.