Research Article

The Earthling's Secret Weapon: Cumulative Culture and the Singularity

Neil Levy

Macquarie University and University of Oxford.

Abstract | Many researchers believe that the singularity – roughly, explosive growth in machine intelligence, resulting in the appearance of AIs that are vastly superior to us cognitively – will or might occur in the relatively near future. Understandably, many are worried by this prospect. In this paper, I argue that we have less to fear from the singularity than many people think. Depending on how we define 'singularity', it is either less likely to occur, or it will likely occur in a form that is not threatening to us. I argue that the capacity for cumulative culture is central to our success as a species, and AIs that rely on processing power alone, without the scaffolding of culture, are unlikely to outcompete us intellectually. There is, however, nothing to prevent AIs from having, and taking advantage of, a capacity for culture. AIs with such a capacity may have intellectual powers greater than ours currently are, but the collective deliberation that underlies the power of cumulative culture is more powerful when there is sufficient diversity among the deliberators. This fact will give AIs a reason to value our continued flourishing, so that we are able to contribute to valuable epistemic diversity.

Editor | Gregg D. Caruso, Corning Community College, SUNY (USA).
Received | March 23, 2016; **Accepted** | July 01, 2016; **Published** | July 09, 2016
Correspondence | Neil Levy,, Department of Philosophy, Macquarie University, NSW 2109, Australia **Email:** neil.levy@mq.edu.au
Citation | Levy, N. (2016). The Earthling's secret weapon: Cumulative culture and the singularity. *Science, Religion and Culture*, 3(1): 19-30.
DOI | http://dx.doi.org/10.17582/journal.src/2016/3.1.19.30

Introduction

Human beings are remarkable animals. Nowhere is this fact more evident than in our technological achievements. Consider how the machines we have designed and built have literally changed the world. They have certainly changed our lives: the internal combustion engine, for instance, has transformed our relations to space, altering our conception of distance and enabling easy interchange between places once isolated from one another. It has also played a significant role in altering the climate. While there is much to celebrate (as well as much to lament) about our achievements, some people worry that we may soon prove too clever for our own good. We may be on the verge of creating technologies which will be more impressive than we are, in just the respects that make us such remarkable animals. We may soon construct genuine AIs; intelligent machines which are capable of feats of intellectual ingenuity every bit as impressive as we are. Given the fact that processing speed doubles every two years, though, these machines can be expected rapidly to outpace us. We will then have built machines that are more intelligent than we are. And that, some people fear, will place us on the edge of a very steep slippery slope.

Once we have machines that are more intelligent than us, our role as AI designers will be obsolete[1]. We will cede this role to the machines themselves. But if we are capable of designing machines more intelligent than us, then those machines will prove capable of designing (or evolving into) still more intelligent machines. They, in turn, will design yet more intelligent machines, and so on (call this process, whereby the development of an AI with an intelligence above a certain threshold leads inexorably to AIs of ever greater intelligence, the *intelligence ratchet*). Very soon after the appearance of the first genuinely intelligent machines, we will be confronted by AIs which are vastly, indeed unfathomably, more intelligent than we are. The event of an explosion in intelligence is known as the singularity[2]. The occurrence of the singularity is an event that many people fear.

There are two reasons for this fear. The fear which has received the most attention is the fear of human extinction. This fear is built on the belief that, given their vastly greater powers of comprehension, post-singularity AIs will have much greater powers of manipulating the world around themselves than we have, and our fates will be in their hands. But (some thinkers maintain), we have no reason to expect them to value our goals or our welfare. At worst, they might see us as a source of danger to be eradicated or a source of energy to be harvested; at best, their utter indifference to our fates will leave us in constant danger of being thoughtlessly swept aside. This fear has received powerful and influential expression in the writings of Yudkowsky (2015) and Bostrom (2014). Bostrom advocates proceeding slowly and cautiously with research that might lead to AIs, while other researchers put their hopes into designing morality into them (Wallach and Allen 2008). The aim is to produce a genuine AI which might value our welfare. Such an AI might help us to respond to the very significant challenges that we confront without adding a significant new challenge of their own.

Bostrom is influentially pessimistic about the prospects of designing morality into AIs. Other, researchers are much more optimistic about AIs, for a variety of reasons: because they reject Bostrom's claim that the goals pursued by agents and their intelligence are orthogonal (Goertzel 2016), because they think we are likely to have more control over the development of AIs than Bostrom thinks, and therefore will have a greater capacity to steer AIs toward moral goals (Agar 2016), because they think that the risks stemming from other problems (climate change, nanotechnology, and so on) that AIs might help us to confront outweigh the risks from AIs themselves (Goertzel 2015) and so on. However, even if AIs shared our values, or valued us, we might still face another threat from them: not to our existence but to our significance. Some have compared post-singularity AIs to gods (Stross 2005), so greatly will they surpass us (as we are now) in power and understanding. Unless we change dramatically ourselves, we will be less than children compared to them. This fact entails that we will be in many ways in a subordinate position to them: our science, our mathematics, even (very likely) our art and novels will pale in comparison to their output, and we will no longer be important source of creativity and knowledge. AIs who valued our welfare might protect us, but we will be in many ways pampered pets, lacking even the capacity to understand the ways in which they act to enhance our welfare. In the grand scheme of things, what we do might no longer much matter.

This prospect, too, is not inevitable of course: we might develop alongside or subsumed within the AIs, so that we keep pace with their development or their development is our development. Kurzweil (2005), for example, rejects the idea that the singularity must pit 'us' against 'them'; for him, the singularity is likely to result from the development of human intelligence in concert with machine intelligence. In this paper, I will join the optimists in arguing that we have less to fear from the singularity than many people have suggested. Depending on how 'singularity' is defined, either it is much less likely that the singularity will occur at all, or when it occurs we will be partners in it, just as Kurzweil suggests, though without us needing to transform our own brains. Worries about a singularity understood as an event after which our destinies are in the hands of machines rest, I will argue, on a conception of ourselves which is false, according to which it is our intelligence that is responsible for our success. In fact, I will suggest, our intelligence is only part of the explanation. While it is necessary, it is not sufficient on its own. We are such successful animals because we are *cultural* animals.

AIs which are not themselves culturally embedded may therefore be much less threatening to us than we tend to think. They may be very impressive, but intelligence alone won't give them power over us. Intelligence without culture is likely to be impoverished in its reach and its power. There is, however, nothing to stop AIs becoming acculturated, and thereby taking advantage of the actual processes that explain our success as a species. That might, indeed, lead to the development of AIs that are incredibly impressive in their capacity to adapt to and transform the world. But this, too, is likely less of a threat to us than it might seem: in extending their powers, such AIs would also extend ours.

Intelligence Versus Cumulative Culture

The development of AI has been slower than many people anticipated, and has undergone many setbacks. Nevertheless, many researchers believe that genuine artificial intelligence is imminent[3]. I will assume that they are right: some time in this century, perhaps in its first half, we will see the development of machines which will count as genuinely intelligent on defensible definitions of the term. Very likely, they will be able to pass the Turing test, because they possess the domain-*general* intelligence passing requires. Domain-*specific* intelligence is intelligence that utilizes a database of information pertaining to a particular problem and is sensitive to a narrow range of stimuli. It may enable impressive cognitive achievements, but it is limited and inflexible. Domain-specific intelligence is the kind of intelligence that evolution builds into minds when organisms face the same kind of problem repeatedly for many generations. Problems like converting visual input into a representation of the external world face all organisms with the right kind of transducers, and have solutions that generalize across most environments in evolutionary history. Less obviously, problems like mate choice and predator detection and avoidance occur repeatedly in similar forms for very many organisms. In both cases, the solutions we have developed depend (at least significantly) on domain-specific mechanisms, which we share with very many other organisms.

Domain-specific intelligence may enable the solution of complex problems, but it cannot enable a machine or organism to pass the Turing test. To pass that test, it is necessary to be able to respond flexibly, and that requires domain-general intelligence. Conversations may flit from topic to topic, and the Turing-capable AI must be able to remain relevant: that will require integrating the current topic (whatever it is) with conversational context. So *Star Wars* and Donald Trump, or carrots and computers, may need to be integrated. Right now, we do not have machines capable of anything like domain-general intelligence. Rather, our most impressive attempts at AI solve specific problems. Building an AI capable of genuine, flexible, intelligence remains an enormous challenge. Nevertheless, I see no reason to believe that it is a challenge that *cannot* be overcome. Since we are capable of domain-general intelligence, and we are entirely physical beings, it is in principle possible to build a physical machine that also exhibits such intelligence.

Once we have cracked that (admittedly, very hard) problem, we might quite rapidly be able to increase machine intelligence. Intelligence in human beings is limited by processing power: there are bottlenecks in information processing and trade-offs between processes. Conscious processing, which seems essential to domain general intelligence (Levy 2014), is a limited resource. Sheer speed of processing is also a limitation. These limitations seem to be contingent features of our minds; that is, they do not represent trade-offs made because increases in these parameters would interfere with other processes required for cognition (as we shall see, such trade-offs are common in the mind). They can therefore be overcome, given the availability of a sufficient number of sufficiently fast processors.

If we define the singularity as consisting in the event that occurs when machines become very much more intelligent than us, then I think we can reasonably expect the singularity to occur, sometime in the current century. However, if we define 'singularity' a little more restrictively, to refer to the event of our fate coming to be in the hands of the AIs (in the same way the fate of gorillas rests more in our hands than in the hands of gorillas, as Bostrom (2014) puts it), then we may have to wait much longer for its occurrence[4]. Even when they are more intelligent than us, they won't be able to outcompete us intellectually, I suspect.

Obviously this claim depends on there being a distinction between the kinds of capacities that might allow them to win the intellectual competition with us, on the one hand, and intelligence, on the other. Let's take human intelligence to refer to the kinds of problem solving powers that human beings possess in virtue of our clever brains. How impressive is this capacity? It is, of course, quite impressive, but it is not very much more impressive than the problem solving capacity of, say, gorillas or baboons. It is not our intelligence that explains our intellectual capacities. It is our culture.

This claim is likely to strike most of us as unlikely. Our success as a species is due to our intelligence, we typically think, and if AIs possess more of it than we do then they are in a position to exercise power over us (given the right effectors, of course). In fact, the claim that our success as a species is due to our intelligence is not merely intuitive; it is one that has been defended by evolutionary psychologists. For Pinker (2010), for instance, homo sapiens' spectacular success at colonizing almost every environment on the planet is due to the fact that we, alone, occupy the "cognitive niche". Whereas other animals have only domain-specific intelligence, we alone have domain-general intelligence, and domain-general intelligence allows us to cope with the novel problems that novel environments pose. Call this the *Martian model*, after the film (and novel) in which someone's ability to 'science the shit' out of things enable them to cope with an adverse environment, and call intuitions that accord with this model *Martian intuitions*[5].

We are remarkably intelligent animals, and there is no doubt whatsoever that this intelligence plays a very significant role in explaining our success. But our intelligence is not sufficient to explain our success. Consider adaptation to a harsh environment, and how many innovations it requires (Richerson and Boyd 2005; Boyd, Richerson and Henrich 2011). Survival in the Arctic Circle requires a large range of complex technologies. Staying warm requires specialized clothes, and these have to be designed out of the impoverished locally available material. The design of these clothes is elaborate. First, caribou must be hunted and their skins stretched, scraped, moistened and stretched again. Then the skin must be sewn into parkas designed to capture heat while allowing moisture to escape. Footwear consists of multiple layers: three different layers of stockings, each with a different design, then two different kinds of boots. But this specialist clothing is not enough to keep the person warm: he or she needs shelter as well. The central Inuit lacked building materials other than snow. Their snow houses were designed to maintain an inside temperature above 10° Celsius, when outside temperatures were below -25°, while at the same time keeping the walls around 0 degrees so that they don't melt. Finding food is a major challenge in the Arctic winter, requiring specialized tools and a great deal of training.

How difficult is to acquire the necessary skills to live in the Arctic Circle? One piece of evidence bearing on this question comes from natural experiments provided by non-native explorers who have found themselves in these environments. In 1846, two British ships became stuck in sea ice at King William Island. The island is regarded by the Netsilik Inuit as rich in resources, and the ships and crews were well-prepared for an Arctic sojourn. But every member of the expedition perished. The well-equipped and well-prepared crew were unable to learn the skills needed for survival. A little more than 50 years later, the Norwegian explorer Roald Amundsen spent two winters on King William Island. He survived, with the help of the Netsilik Inuit, who taught him many of the skills he needed. Presumably the British could have acquired the relevant skills, but despite the preparation, and indeed experience of the sailors (Sir John Franklin, who headed the expedition, was on his fourth Arctic trip), they were unable to work out for themselves how to survive in the harsh environment (Boyd, Richerson and Heinrich 2011).

Similar stories, both of the death of explorers and survival only with the help of the indigenous population, can be multiplied using Australian examples. In 1861, members of the ill-fated Burke and Wills expedition to cross the Australian continent were given cakes made from the seeds of the Nardoo plant by local Aboriginals. Unwilling to rely on assistance from people they saw as inferiors, they spurned further assistance and gathered Nardoo for themselves. They harvested the seeds and, as they had seen the Aborigines do, they ground them into a powder which they mixed with water. But they missed a step: the Aborigines roasted the seed cases prior to grinding them. The prepared Nardoo satisfied the explorers' hunger but robbed them of Vitamin B and probably hastened their deaths (Burcham 2008). Again, the well-prepared expedition failed catastrophically because the explorers were unable to deduce how to survive in the harsh environment.

These stories manifest the limitations of human ingenuity. Human beings are remarkably clever animals, but we have not been able to spread to a huge range of environments in virtue of our intelligence alone. In fact, our intelligence is not up to the job: no one could ever come up with all the innovations needed to survive in the Arctic or the central Australian desert. If it is not our intelligence that explains our success, however, what is it? Following Richerson and Boyd (2005), I suggest it is our culture.

The Inuit can survive in the Arctic winter because they have a cumulative culture. Survival depends on numerous acquired technologies and learned behaviours. To survive, you must know when to hunt, what to hunt, how to hunt, and how to make and use the weapons needed. You need to make and wear specialized clothing, specialized footwear, and even specialized goggles. These technologies and behaviours are the product of generations of innovation, each of which improves on the available tools incrementally. The innovations are gradual and cumulative, allowing for the population to adapt to the environment and colonize places that, prior to the innovations, were inaccessible to them. They are also distributed across the population: not only are they beyond any one person's inventing; they may even be beyond any one person's learning. Culture allows for a distribution of labour, so that no one person needs the full range of skills required for survival and flourishing.

Intelligence certainly plays a role in developing and applying the expertise needed for survival, but the role is smaller than might be thought. The innovations are aided by intelligence (though innovations may often be serendipitous). Retaining successful innovations and rejecting unsuccessful ones requires some intelligence (often very little: it may take little intelligence to see that one bow is more effective than another). In fact, even experts often lack a detailed understanding of the tools and other technologies they use. Food taboos provide a rich source of examples. These taboos are often functional, but the functional explanations may be unavailable to the people who accept them (Henrich and Henrich 2010).

Instead of reasoning our way to the tools and techniques we need for flourishing, we acquire them from those around us, especially as children. Because the acquisition of local tools and techniques is so crucial to human flourishing, we have acquired psychological adaptations which dispose us to such acquisition. We are over-imitators, compared to other primates. Nagell, Olguin and Tomasello (1993) compared the performance of human children and chimps on the acquisition of a skill that was demonstrated to them. Experimenters used a rake, *tine side down*, to draw sweets that were otherwise out of reach toward themselves. Using a rake that way is highly inefficient: many sweets slip through the gaps in the tines. When they were given the opportunity to perform the task, chimps flipped the rake so that the flat side acted as a far more efficient tool, with few sweets escaping. But children did not: they tended to imitate the actions demonstrated. Similarly, infants shown by experimenters how to turn on a switch by butting it with their heads will themselves turn it on by butting, rather than using their hands (Meltzoff 1988). Chimps acted more intelligently than human children, in a central respect: they exercised their causal knowledge to hit upon a more efficient technique. We are disposed to imitate *rather than* to innovate intelligently, which suggests that acquisition of local ways of proceeding outperforms the application of intelligence[6].

Suppose the hypothesis I have sketched here, according to which our success as a species is due to cumulative culture rather than sheer intellect, is true. Then it follows that in acquiring equal or greater intelligence, AIs will not have acquired the capacities that explains our success. It does not follow that they will not be able to outcompete us, however. Outside the arena of sport, in which rules constrain how we compete, success in competition does not depend on any one set of capacities (human beings may outcompete birds without learning to fly). Even if we don't flourish in virtue of our intelligence, machines may dominate us in virtue of *their* greater intelligence.

That's certainly true, but we ought to be wary of assuming that their greater intelligence will give them greater intellectual powers. The impression that it will may be the product of our Martian intuitions: we tend to think that it is our intelligence that explains our success, so we are disposed to think that a more intelligent machine will be more successful than us. It may be that greater intelligence doesn't translate into a capacity to exercise power over environment or over us.

That said, there are in fact grounds for thinking that the high intelligence of AIs will translate into intellectual mastery. Cumulative culture explains our success because there are features of our minds, on the one hand, and our environments, on the other, that have made it more successful than the Martian model. AIs may be thought to lack the first kind of feature and be able to control the second kind, thereby ensuring that their intelligence outcompetes our culture.

First our minds. Human cognition is characterized by a number of trade-offs. We utilize heuristics and dispositions that are adaptive, in ancestral environments, but which misfire in novel environments (Gilovich, Griffin and Kahneman 2002). We tend to be over-impressed by the easily observable and blind to statistical regularities for similar reasons. We test hypotheses in biased ways – we are motivated reasoners (Lord, Ross and Lepper 1979), who look for evidence in favour of hypotheses and overlook evidence against those to which we are committed (Nickerson 1998). Relying on such dispositions may have been adaptive prior to the development of cumulative culture due to resource limitations and restrictions on processing speed, but cumulative culture allows us to circumvent them.

For instance, while motivated reasoning is a severe restriction on the intellectual capacities of individual cognizers, distributed cognition may overcome this restriction: when cognitive labour is distributed, the conflicting preferences of individual reasoners ensures that the space of options is thoroughly explored. I am well disposed toward my own hypothesis, *h*, and motivated to look for evidence in its favour. That fact makes it hard for me to come by proper justification for *h*. But you are not well disposed toward *h*; rather, you favour *h**. You are therefore motivated to look for evidence against *h* (and I am motivated to look for evidence against *h**). Neither of us are very good at assessing our own theory, but we are good at assessing one another's (Ditto and Lopez 1992). By distributing the cognitive labour, the scientific community is able effectively to pursue truth, even though none of us is particularly good at tracking it (Mercier and Sperber 2011).

If cumulative culture is so important because it allows us to overcome our limitations, then AIs may be successful on the basis of intelligence alone if their intelligence is not limited in these ways. Psychologists talk about strategies for debiasing cognition. For instance, one can guard against the confirmation bias by reminding yourself of the need to conduct symmetrical memory searches; there is some evidence that doing so reduces its effects on cognition (Lilienfeld, Ammirati and Landfield 2009). But obviously a bias toward one sided searches can be designed out of AIs from the get go. They will not need debiasing, because they are not biased. They will look as much for disconfirming evidence as confirming. Similarly, they may not need to rely on heuristics because their much faster processors and much more extensive resources ensure that they will not need to trade speed against accuracy (or, perhaps more likely, that it is only when they face much more demanding problems than those we ever contemplate that they will need to make such trade-offs).

Whether it is possible to design better cognizers by designing out our limitations depends on the nature of the trade-offs being made. If we are trading accuracy for speed because accuracy is too costly, then faster and cheaper processors will make the trade-off unnecessary. But if we accept lower accuracy in one domain in return for higher accuracy in another, then the trade-off may be indispensable. Many heuristics are rational not only because we have resource limitations but because the search space is unbounded. We need ways of tagging information as relevant, or stopping rules, because there is an indefinitely large number of things we might take into account. Consider a simple problem, like where to eat dinner tonight. The range of considerations that might rationally be taken to bear on the question is open ended: the price of the food, its quality, the type of cuisine, its atmosphere, its distance from where we are, its distance from our next destination (and why take that to be fixed?), its ambience, the weather, the day of the week, the phase of the moon, and so on. Deciding where to eat might require, first, settling what type of cuisine we want, and innumerable considerations bear on *that*. And so on. We encounter a combinatorial explosion of considerations. The isotropy of cognition – the way in which anything might be relevant to anything else ("in principle, our botany constrains our astronomy, if only we could think of ways to make them connect", Fodor (1983, 105) writes) – ensures that greater processing resources don't obviate the need for heuristics. Evolution is not an optimizing process, and no doubt some of our limitations can be avoided with better design, but some may be inevitable (and perhaps AIs will be subject to new ones).

Our cognitive limitations may reflect trade-offs in other ways: it may be that our defects are side-effects of dispositions that work very well when cognition is distributed and culturally embedded. Distributed cognition works well not by *overcoming* the confirmation bias but by *harnessing* it: the disposition to defend one's antecedent view and to attack rivals may be more productive for identifying significant truths than a disposition to reason 'dispassionately'. Designing our cognitive limitations out of AIs might, in some cases at least, be designing out *features*, not bugs.

There is however another reason to think that IAs might outcompete us without requiring cumulative culture: cumulative culture is an adaptation to features of the environment in which we evolved, and the environment in which we compete with AIs may lack those features, ensuring that our capacity for cumulative culture is not advantageous. Richerson and Boyd (2005: 118) note that selection "favours a heavy reliance on imitation whenever […] environments are neither too variable nor too stable". If environments are very stable across time, encoding solutions to recurrent problems into genes becomes feasible; selection in favour of such genes will lead to less reliance on cumulative culture. If environments are not stable enough, acquiring behaviours and technologies from others is not adaptive, because what worked for *them*, in the past, is not likely to work for the imitator. In very rapidly changing environments, innovation beats imitation.

It is quite possible that the environment to which we will have to adapt in the relatively near future will be more rapidly changing than the Holocene has been. Equally, however, it is likely that our new methods of disseminating and adopting cultural innovations – through the mass media and, especially, the internet – make it possible for imitation to keep pace with this turbulence. Just as the invention of literacy allowed the accumulation of cultural knowledge well beyond what would be possible were we required to rely on encoding in human memory, so the rapid exchange of information and the fine grained distribution of cognition that computer networks allow for may allow us to flourish in a much more rapidly changing environment. While the world may soon – may already – change too rapidly for our imitative dispositions to keep pace with unaided, they are not unaided. These dispositions may have enabled the development of a scaffold that will allow them to cope with this challenging environment, and to provide us with the capacities to generate the next generation of external scaffoldings for ever more rapid cultural innovation.

There is however one scenario in which AIs might achieve greater intellectual powers than us: they might (somehow) very greatly retard the pace of change. If they did that, then to the extent that cumulative culture is adaptive in environments that change (though not too rapidly), they might be able to render it redundant as a means of expanding our own powers. We could then find ourselves in the following situation: the AIs would possess greater domain-general intelligence than us *and* all or most of the resources that our cumulative culture makes available (insofar as the AI can glean the techniques and know how that make up our cumulative culture from our anthropological and sociological and political journals and books, our newspapers and our novels – all conveniently digitized for easy access, of course). Since further development of cumulative culture would not produce a significant increase in intellectual powers, given that the environment is static, the AIs' greater intelligence combined with a more or less equal facility with the resources of cumulative culture so far would grant them greater intellectual powers than us[7].

But this scenario is highly unlikely. Given that the world is changing more rapidly than ever, in ways that seem to make cumulative culture more powerful and more significant than ever, we should not expect stasis. In fact, the development of AIs would seem to increase the pace of change, rather than decelerate it. Perhaps sufficiently powerful AIs could impose stasis on the world in a way that gives them a decisive advantage (though it is not obvious that they could in fact do this: recall Lampedusa's adage, that if we want things to stay the same, we're going to have to change), but imposing stasis requires them to *already* possess the intellectual powers that, we are supposing, the existence of stasis might give them. In a rapidly changing world, the AIs won't be able to do without the further development of cumulative culture: access to its contents at any one time will not give them the resources to cope with the ways in which the environment changes.

Cultural Computers

I have been arguing that we ought to distrust the intuition that highly intelligent AIs will have greater intellectual powers than us, on the grounds that this intuition is the product of the fact that the Martian model is itself intuitive, but the Martian model is false. Though we certainly cannot be sure that enough intelligence won't translate into greater intellectual powers than we possess, we ought to be wary of thinking it will. But if our success is due not to our intelligence, but to our cumulative culture, can't intelligent AIs can outcompete us by enculturating themselves? They thereby will reap the rewards of culture on top of whatever rewards their intelligence will bring them. If that's right, then even if intellectual mastery is due to cumulative culture rather than intelligence (alone), once AIs develop cumulative culture of their own they will be in a position to dominate us.

There is no in principle obstacle that I can see to AIs developing cumulative culture. Culture is no more mysterious than intelligence (perhaps it is less mysterious): if intelligence can be instantiated by machines (and there is no compelling reason to think that it cannot be), they can instantiate culture too. So despite my doubts about the reach of intelligence, I do not doubt that AIs can achieve intellectual mastery equal to our own.

But if AIs can achieve intellectual mastery equal to our own, then it seems that they can achieve greater intellectual mastery. The reason for thinking this is the case is closely parallel to the reason for thinking that the singularity may occur relatively soon: the rapid increase in speed and efficiency of processing will lead to IAs that have better cumulative culture than ours. From then on, as they take over the process of designing the next generation of AIs, we ought to expect a rapid increase in capacities. Alongside, or instead of, an intelligence ratchet, we can expect a cultural ratchet.

There are some reasons to be cautious about postulating a cultural ratchet. An important reason is that culture is unlike intelligence in a central way. Though there is a great deal of dispute about whether intelligence can be reliably measured, and concerning whether it is single or multiple, almost everyone accepts that we can make at least rough comparisons of the intelligence of different individuals at least along some dimensions. Very few people would deny that in very important respects adult human beings are very much more intelligent than mice, for example, and few people would deny that, again in very important respects (though perhaps not in every respect), Richard Feynman's intelligence was greater than Donald Trump's. But it is far from obvious that these kinds of comparison are sensible with regard to cultures. Today, few people are happy to claim that one culture is better than another.

The claim that there might be a cultural ratchet does not depend on our being able to rank cultures, however. It depends, rather, on the less controversial (if not entirely uncontroversial) claim that some cultures enable greater problem solving capacities, across many significant domains, than others. This claim ought to be much less controversial than the claim that some cultures are, overall or in important dimensions, better than others. Surely it is true that we have greater problem solving capacities today in multiple fields than we had 50 years ago. The claim that we have been growing rapidly in this kind of problem solving capacity at least since the scientific revolution should be relatively uncontroversial[8]. So the incomparability of cultures along other dimensions is not an obstacle to the cultural ratchet.

It might be doubted that AIs can design in better cumulative culture. It should certainly be conceded that we have little idea how the processes that result in cumulative culture could be enhanced, and probably fair to say we have less idea how this can be done than how intelligence can be enhanced. Nevertheless, these seem to be limitations that can be overcome, via the application of intelligence and cultural experimentation. Just as it is unlikely that the psychological dispositions that realize our intelligence are optimally designed, so it is unlikely that the suite of dispositions that enable cumulative culture are optimal. The transmission biases which lead us to embrace certain innovations and reject others might be fine-tuned, for example. Alternatively, sufficiently intelligent computers might be able to construct evolutionary models to allow for experimentation with cultural variants, thereby enabling a massively speeded up process of testing such variants and embracing the more successful. I therefore doubt that there are insuperable obstacles to AIs becoming successful cultural machines.

If AIs become successful cultural machines, however, they can reasonably be expected to have greater intellectual mastery than we – as we are right now – can have. They might be able to design better processes of cultural accumulation. Even if they can't, their greater intelligence will give them an advantage over us, given equality of cumulative culture. While such a process could result in AIs possessing the power to eliminate or marginalize us, I am optimistic that they will pose neither an existential threat to us nor a threat to our significance. The reason is this: the mechanisms of collective deliberation which (partly) underlie the power of cumulative culture work more powerfully given a diversity of opinions and assumptions. Consider the confirmation bias (Nickerson 1998). This pathology of individual level thought is transformed into an epistemic virtue in collective deliberation, given sufficient diversity of views: our disposition to defend our views against rivals ensures that all views are scrutinized (Mercier and Sperber 2011). Without genuine diversity, collective deliberation is prone to groupthink and group polarization (Isenberg 1986), whereby groups become more extreme in their antecedent views. With sufficient diversity, these problems are avoided and the hypothesis space is explored more deeply. IAs might try to garner the benefits of cultural diversity by emulation or simulation, but it may be more efficient to use existing cultural diversity as its own best model. That fact entails that they have reason to allow us to continue to exist, in sufficiently good conditions to be exploring questions in which they are interested, and entails that our intellectual culture will remain relevant to the frontiers of knowledge[9].

While there is evidence that even ignorant dissent improves collective deliberation (Surowiecki 2004), our cumulative culture will be more epistemically valuable to AIs if we do not fall too far behind them in our intellectual mastery. We might keep up through enhancement, but even in its absence (or to compensate for its limitations), we can take advantage of the fact that the cultural ratchet has positive externalities. The cultural innovations made by AIs become available to us (even if AIs are unfriendly to us, this might be the case: there is plenty of evidence of cultural innovations spreading across rival groups in the historical and archaeological record. In fact, that is one way in which group conflict is sometimes resolved: one group embraces the cultural innovations of another and thereby comes to be absorbed into it). Cultural innovations very significantly involve artefacts and niche construction: altering the physical environment to better support cognition and behaviour. If we allow AIs to increase their intellectual mastery by such innovations, we may thereby allow them to increase our own. Perhaps, indeed, the distinction between 'them' and 'us' might eventually become meaningless, as they join us in the project of building better beings. If that's right, then Kurzweil's vision of the singularity – on which it involves our becoming vastly more intellectually powerful in concert with the machines – is more likely than the visions of the pessimists. Contra Kurzweil, however, there may be little need for us to increase our intelligence for this to occur.

Conclusion

I have argued that we owe our intellectual capacities very significantly to our cumulative culture. Culturally embedded cognition allows us to distribute cognition across groups, allowing problems to be broken down into parts, with each solved separately and for our cognitive limitations to be transformed into virtues. Our disposition to imitate allows for the gradual accumulation of successful solutions to recurrent or persistent problems, and allows for an incremental improvement in these solutions. The fact that our impressive problem solving capacities (and our equally impressive problem creating capacities) are due to these features of ourselves, and not to our intelligence, should give us pause when we contemplate the possibility that AIs might soon have intellectual mastery over us. The Martian intuition is powerful, but it is mistaken: sheer intelligence is not what accounts for the reach and extent of our capacities. For this reason, we ought to be wary of thinking that super-intelligent machines will have a longer, more extensive, reach than we do, in virtue of their intelligence.

I have conceded, of course, that AIs can themselves distribute cognition, and accumulate innovations. They can take advantage of the cultural ratchet just as we have. But, I have suggested, that may not pose a threat to us, or place our fate in the AIs hands. Cultural diversity is epistemically beneficial, and that fact gives the AIs reason not merely to tolerate our existence but to ensure that we continue to flourish (such that we can continue to engage in inquiry that will generate knowledge that they care about). We may keep up with them in part at least by taking advantage of positive externalities generated by the cultural ratchet. The innovations it makes available are generally available for adoption. The cultural ratchet may indeed provide an opportunity for AIs to increase their problem-solving capacities beyond our current levels, but in so doing it may allow us to increase our own capacities to the same extent.

Nothing I have said here constitutes anything like a decisive argument for the claim that the singularity won't occur, or that it will occur in a benign form. It may be true that we owe our success to culture and not unaided intelligence, yet still be the case that the superintelligence of AIs allows them to possess unparalleled problem solving capacities without the need for cultural scaffolding. Alternatively, it might be true that AIs take advantage of the cultural ratchet in a way that generates more negative externalities for us than positive, such that they outpace us culturally. Or they may be able to simulate cultural diversity in a way that is more conducive to intellectual mastery than the preservation of actual diversity. Detailed modelling of these processes is required to cast light on the probability that that will occur. What I take myself to have achieved is nevertheless significant. I have suggested that the anxiety that superintelligence will translate into intellectual mastery – over us, most significantly – is driven by the Martian intuition. Since the Martian intuition is false, we should hesitate from thinking that superintelligence is the threat that thinkers like Bostrom believe. There are other live options. Even if superintelligence emerges soon, the future may not be bleak.

Acknowledgements

I am grateful to Anders Sandberg for comments on an earlier version of this paper, as well as to an anonymous reviewer for this journal for a genuinely helpful and challenging set of comments.

References

- Agar, N. 2016. Don't Worry About Superintelligence. *Journal of Evolution & Technology* 26: 73-82.
- Boyd, R., Richerson, P.J. and Henrich J. 2011. The Cultural Niche: Why Social Learning is Essential for Human Adaptation. *Proceedings of the National Academy of Sciences* 108: 10918-10925. http://dx.doi.org/10.1073/pnas.1100290108
- Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Burcham, P.C. 2008. Toxicology Down Under: Past Achievements, Present Realities, and Future Prospects. *Chemical Research in Toxicology* 21: 967–970. http://dx.doi.org/10.1021/tx8001252
- Chalmers, D.J. 2010. The Singularity: A Philosophical Analysis. *Journal of Consciousness Studies* 17: 9 - 10.
- Dickens, W.T., and J.R. Flynn. 2002. The IQ Paradox Is Still Resolved. *Psychological Review* 109: 764-771. http://dx.doi.org/10.1037/0033-295X.109.4.764
- Ditto, P.H., and D.F. Lopez. 1992. Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology* 63: 568-584. http://dx.doi.org/10.1037/0022-3514.63.4.568
- Fodor, J. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.
- Gilovich, T., D. Griffin, and D. Kahneman. 2002. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.
- Goertzel, B. 2015. Superintelligence: Fears, Promises and Potentials. *Journal of Evolution & Technology* 24: 55-87. http://dx.doi.org/10.1017/CBO9780511808098
- Goertzel, B. 2016. Infusing Advanced AGIs with Human-Like Value Systems: Two Theses. *Journal of Evolution & Technology* 26: 50-72.
- Henrich, J. and N. Henrich. 2010. The evolution of cultural adaptations: Fijian food taboos protect against dangerous marine toxins. *Proceedings of the Royal Society B: Biological Sciences* 277: 3715-3724. http://dx.doi.org/10.1098/rspb.2010.1191
- Isenberg, D.J. 1986. Group Polarization: A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology* 50: 1141–1151. http://dx.doi.org/10.1037/0022-3514.50.6.1141
- Kurzweil, R. 2005. *The Singularity Is Near: When Humans Transcend Biology*. London: Penguin.
- Levinovitz, A. 2014. The Mystery of Go, the Ancient Game That Computers Still Can't Win. *Wired*, 15 December 2014. http://www.wired.com/2014/05/the-world-of-computer-go/
- Levy, N. 2014. *Consciousness and Moral Responsibility*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198704638.001.0001
- Levy, N. Forthcoming. Embodied Savoir-Faire: Knowledge-How Requires Motor Representations. *Synthese*. 10.1007/s11229-015-0956-1
- Lilienfeld, S.O., T. Ammirati and K. Landfield. 2009. Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science* 4: 390-398.
- Lord, C.G., L. Ross and M.R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psycholo*gy 37: 2098–2109. http://dx.doi.org/10.1037/0022-3514.37.11.2098
- Meltzoff, A.N. 1988. Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology* 24: 470–476. http://dx.doi.org/10.1037/0012-1649.24.4.470
- Nagell, K., R.S. Olguin and M. Tomasello. 1993. Processes of social learning in the tool use of chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Journal of Comparative Psychology* 107: 174 –186. http://dx.doi.org/10.1037/0735-

7036.107.2.174

- Nickerson, R.S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2: 175-220. http://dx.doi.org/10.1037/1089-2680.2.2.175
- Pinker, S. 2010. The Cognitive Niche: Coevolution of Intelligence, Sociality and Language. *Proceedings of the National Academy of Sciences* 107: 8993-8999. http://dx.doi.org/10.1073/pnas.0914630107
- Richerson, P.J., and R. Boyd. 2005. *Not by genes alone*. Chicago: University of Chicago Press.
- Silver, D., et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529: 484–489. http://dx.doi.org/10.1038/nature16961
- Mercier, H., and D. Sperber. 2011. Who do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* 34: 57-111. http://dx.doi.org/10.1017/S0140525X10000968
- Stanley, J. 2011. *Know How*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199695362.001.0001
- Stross, C. 2005. *Toast*. Cosmos Books.
- Surowiecki, J. 2004. *The Wisdom of Crowds*. New York: Doubleday.
- Wallach, W., and C. Allen. 2008. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.
- Weir, A. 2014. *The Martian*. London: Penguin.
- Yudkowsky, E. 2015. *Rationality: From AI to Zombies.* Machine Intelligence Research Institute.

Endnotes

[1] Some researchers think that the route to genuine artificial intelligence lies through building (relatively dumb) learning or evolving machines, which will then develop into machines as or more intelligent than us. If that's right, then we will be obsolete as designers even earlier than if we build the first machines intelligent enough to themselves build or develop into the next generation of AIs.

[2] There is a more technical definition of a singularity, involving extremely rapid development of intelligence to the limits imposed by physics. In common with most writers, however, I will use the term in a looser sense, to refer to the relatively rapid development of intelligence much greater than ours (see Chalmers 2010 for discussion).

[3] The recent success of a computing system at beating an expert Go player (Silver at al. 2016) might be an indication that progress is now once again rapid. Go is harder for a computer to master than chess, and it had recently been predicted that it would take another 10 years for a computer to play the game at an expert level (Levinovitz 2014).

[4] Or rather, if we understand 'singularity' as consisting in the event of AIs crossing some threshold such that in virtue of their capacity to outcompete us intellectually our fate will be in their hands, their coming to be more intelligent than us will not entail the singularity. Obviously, AIs need not be more intelligent than us for our fate to be in their hands: some minimal degree of intelligence combined with sentience, and access to the right effectors, might be even more dangerous to our autonomy or existence than superintelligence. Imagine our fate in the hands of a toddler.

[5] The line "I'm going to have to science the shit out of this" occurs in the film version of *The Martian*, but not the original book. It is, however, an accurate encapsulation of the flavour of the book as well as the film. Interestingly, while the film and the line resonated strongly with many scientists (Neil deGrasse Tyson tweeted that that was his favourite line in the film), there is good reason to think that the success of science is explained by its structure and its division of cognitive labour – by the way in which instantiates features of the cultural model, the rival of the Martian model – rather than by the way it instantiates the Martian model.

[6] One relatively obvious reason to think that intelligence isn't sufficient for intellectual mastery: our intellectual mastery has increased significantly since the advent of behavioural modernity, roughly 50,000 years ago, but our intelligence has probably not increased anywhere near as significantly since the emergence of homo sapiens 200,000 years ago.

[7] This scenario presupposes that it is possible to access the content of cumulative culture intellectually: that is to say, that its contents can all be reduced, without significant loss, to knowledge-that. If cumulative culture is in fact partially embodied – and therefore in a representational format that cannot be reduced to knowledge-that – then accessing books and newspapers, and so on, would not be sufficient for accessing all its significant content. There is a large debate over

the extent to which knowledge-how can in fact be reduced to knowledge-that; see Stanley (2011) for a defence of the claim that it can, and Levy (forthcoming) for a defence of the claim that it cannot. Of course, the claim that knowledge-how cannot be reduced to knowledge-that does not entail that AIs cannot access it; merely that they will themselves need to be appropriately embodied to access it.

[8] It might be objected that our growing intellectual mastery is due to our rising intelligence, not to our culture. IQs have been rising gradually since they were first measured (the so-called Flynn effect), necessitating continual recalibration of the scale. Even if our rising IQs is partially explanatory of our increased problem solving capacity, however, it is very likely that these increases are themselves explained by changes in cultural complexity (Dickens and Flynn 2002). Our problem solving capacities are the product of the co-evolution of culture and brains.

[9] Here's another scenario in which the AIs have greater intellectual mastery than we do. They have more intelligence than us, and in addition they keep up with all our cultural innovations. We possess no advantage over them in terms of culture, but they possess an advantage over us in terms of intelligence. While this seems possible, this scenario would neither represent an existential threat to us nor a threat our significance: the AIs would require us, not as pets or farmed animals, but as flourishing cultural beings so that our cumulative culture could provide them with a resource.