Ethics for Possible Futures

TIM MULGAN (AUCKLAND/ST. ANDREWS)
PROCEEDINGS OF THE ARISTOTELIAN SOCIETY
135TH SESSION

ISSUE NO. 1
VOLUME CXIV
2013 - 2014

ETHICS FOR POSSIBLE FUTURES

TIM MULGAN
UNIVERSITY OF AUCKLAND
UNIVERSITY OF ST. ANDREWS

MONDAY, 4 NOVEMBER 2013
17.30 - 19.15

THE WOBNUR SUITE
SENATE HOUSE
UNIVERSITY OF LONDON
MALET STREET
LONDON WC1E 7HU
UNITED KINGDOM

This event is catered, free of charge, &
open to the general public

CONTACT
mail@aristoteliansociety.org.uk
www.aristoteliansociety.org.uk

© 2013 THE ARISTOTELIAN SOCIETY
BIOGRAPHY

Tim Mulgan is Professor of Philosophy at the University of Auckland, and Professor of Moral and Political Philosophy at the University of St Andrews. He is the author of The Demands of Consequentialism (Oxford University Press 2001), Future People (Oxford University Press 2006), Understanding Utilitarianism (Acumen 2007), and Ethics for a Broken World (Acumen/McGill-Queens University Press 2011). He is currently completing a manuscript for Oxford University Press entitled Purpose in the Universe: the moral and metaphysical case for ananthropocentric purposivism.

EDITORIAL NOTE

The following paper is a draft version that can only be cited or quoted with the author’s permission. The final paper will be published in Proceedings of the Aristotelian Society, Issue No. 1, Volume CXIV (2014). Please visit the Society’s website for subscription information: www.aristoteliansociety.org.uk.
I explore the moral implications of four possible futures: a broken future where our affluent way of life is no longer available; a virtual future where human beings spend their entire lives in Nozick’s experience machine; a digital future where humans have been replaced by unconscious digital beings; and a theological future where the existence of God has been proved. These futures affect our current ethical thinking in surprising ways. They raise the importance of intergenerational ethics, alter the balance between competing moral theories, and push morality in a more objective direction.

IN this paper, I examine the impact on moral and political philosophy of four credible futures. These futures are designed to question several commonplace presuppositions of contemporary philosophy. Imagining specific futures gives our obligations to future people a new urgency, and also influences our current ethical thinking in several surprising ways.

I begin with a future I have discussed in detail elsewhere: the broken world (Section I). Section II shows how a virtual future modelled on Nozick’s experience machine pushes morality in a more objective direction. Sections III and IV briefly explore two more outlandish futures where intelligent machines discover strange new facts about value or God.

I. THE BROKEN WORLD

In my book, *Ethics for a Broken World*, I imagine a future where resources are insufficient to meet everyone’s basic needs, where a chaotic climate makes life precarious, where each generation is worse-off than the last, and where our affluent way of life is no longer an option. (Mulgan 2011. See also Mulgan 2012, 2014a and 2014b.) In a philosophy class in that broken world, students and teachers look back in disbelief at a lost age of affluence. They struggle to make sense of the opulent worldview of late-affluent philosophers such as Nozick and Rawls, and the behaviour of affluent citizens like us.

The broken world lacks two ubiquitous but often unacknowledged presuppositions of recent moral and political thought: that future people will be better-off than present people; and that the interests of different generations largely coincide. To get a stark contrast with our own
affluent world, I also stipulate that the broken world lacks favourable conditions.

A society enjoys favourable conditions if it has reached a level of sophistication and prosperity such that its members can establish liberal democratic institutions that meet all basic needs without sacrificing any basic liberties (Rawls 1971, p. 178). Modern liberal democracies in North America, Western Europe, and Australasia clearly enjoy favourable conditions. Indeed, Rawls argues that virtually all modern societies enjoy them (Rawls 1999, p. 108). In the broken world, favourable conditions are gone. No broken society can meet all basic needs, and therefore none could possibly establish Rawlsian liberal institutions that both meet basic needs and protect basic liberties.

I picture this future scarcity, not as a one-off catastrophe, but as an ongoing fact of life. (A parallel might be the regular seasonal fluctuations in food supply experienced by traditional Inuit communities – Rawls’s own example of a society who might lack favourable conditions.) In a broken world, thanks to the scarcity of material resources (especially water) and the unpredictable climate, societies periodically face population bottlenecks where not everyone can survive.

On the other hand, my broken world is not apocalyptic. Functioning human societies do exist there. Some people even have time to sit around wondering about justice. But the content of those wonderings, and the society in which they take place, are very different from our own.

The broken world is a credible future. No-one can reasonably be confident that it won’t happen. It involves no outlandish claims, scientific impossibilities, or implausible expectations about human behaviour. Climate change – or some other disaster – might produce a broken future. This is not to say, of course, that the broken future will happen. Many other futures are also credible. Some are much better, others are much worse. Our epistemic situation does not allow us to make confident predictions either way. But the broken world is one very real possibility.

In this paper, I take the credibility of my broken future as given, and explore its implications for moral philosophy.

Contemporary ethics presupposes that future people will be better-off than present people; that the interests of different generations largely coincide; and that favourable conditions will persist indefinitely. The removal of these three presuppositions has a significant impact on moral philosophy in the broken world. Elsewhere, I have explored a number of places where current ethical thinking must change in a broken future. One affluent ethical concept that is especially difficult to translate is our
notion of rights. Will rights be a luxury that future people cannot afford?

I begin with a familiar utilitarian story about rights. Since Bentham, utilitarians have been suspicious that the rights of the rich will trump the needs of the poor. Rights are not much use if you cannot stay alive. Many modern utilitarians respond, not by rejecting rights, but by expanding them – adding a right to subsistence, or a right to have your basic needs met. Utilitarians insist that, without this extension, rights cannot be regarded as trumps.

This modern utilitarian account of rights faces obvious difficulties in the broken world. If we cannot meet all basic needs, then we cannot hope to honour a universal right to subsistence. And basic needs will inevitably conflict with other individual rights. Social survival in a broken world may require restrictions on personal liberty on a scale that people have only previously accepted in times of war, or other temporary crisis. Private land and individual labour might be requisitioned to grow food; the use of fossil fuels for private purposes might be severely curtailed; and individual lifestyle choices – especially reproductive decisions – might be very tightly regulated and constrained.

Rights will look very different in a broken world. Indeed, if we insist that a right is something that must be guaranteed to every individual, then the very idea of rights seems to disappear. If we cannot guarantee everyone’s survival, how can we hope to honour all their rights? One might conclude that future people will simply abandon the discourse of rights. A more interesting possibility is that they will find new ways to think about rights. A theme of my book is that every broken world society will need to institute a survival lottery – some bureaucratic procedure to determine who lives and who dies. To implement such a lottery in our affluent world would be a monstrous violation of rights. However, in the chaotic climate of the broken world, survival lotteries might be necessary to protect rights. Participating in the lottery, and in the preceding deliberation, may be the best way for future people to fine-tune their views on the balance between freedom and survival. (Would you rather have a high probability of bare survival, or a longer shot at a more affluent life?) Perhaps, for them, this is what a right will be: an equal input to collective deliberation, and then an equal chance to live or die.

Broken world philosophy will revolve around the design of a just survival lottery. Theories of freedom, autonomy, rights, responsibilities, supererogation, human flourishing, authority, punishment, and much else, will all need to earn their keep within some over-arching vision of a just society governed by a fair lottery. To design their lotteries, future philosophers may draw inspiration from debates in our affluent philosophy about the role of lotteries in allocating our scarce resources –
such as medical technologies, political offices, university places, or other limited opportunities.

If we encountered an isolated population, perhaps on some distant planet, living without favourable conditions and operating survival lotteries, that would be unsettling enough. But because the broken world may be our future, it also has a significant impact on our current ethical thinking.

A credible broken future teaches us four main ethical lessons. First, it undermines our tendency to ignore our obligations to distant future people. Philosophers have traditionally marginalised intergenerational issues, because they were confident that they could set the future aside. If we create a stable liberal democratic society in our own generation, then our descendants will inevitably be better-off than us, and therefore their interests do not conflict with ours. In Rawls’s liberal society, for instance, the only intergenerational question is the ‘just savings problem’: how much better-off we should leave our descendants? (Rawls 1971, pp. 251-259.) The prospect of a broken future undermines this optimistic presentism. We no longer take it for granted that we will leave our descendants better-off, or even that we can. And many of our most urgent moral dilemmas involve intergenerational conflict. We now realize that future people might be worse-off because we have looked after ourselves.

Imagining a broken future also raises the motivational significance of our intergenerational obligations. Our duties to contemporaries naturally engage our moral sentiments, because we must justify ourselves to the actual people whose real-life interests are affected by our actions. By contrast, distant future people are very remote from our everyday concerns. Asking how our actions might impact on actual future people can help redress this imbalance, by giving our obligations to future people the same felt urgency as our obligations to one another.

A second lesson is that the broken future alters the comparative plausibility of competing moral and political theories, simply because some theories cope better than others with obligations to future people in general. I will discuss two examples. In the next section, I argue that objective accounts of human well-being deal with the future better than subjective ones. In this section, drawing on my own recent work, I argue that utilitarianism accommodates the future more easily than contractualism, which is its main rival in the contemporary intergenerational literature (Mulgan 2006, chapter 1 and 2; Mulgan 2014c).

Utilitarianism bases all our obligations on the fact that our actions impact on the well-being of sentient beings. Obligations to future people are theoretically on a par with obligations to present people. While
utilitarians endlessly debate the precise details of our intergenerational obligations, they have no difficulty making sense of them. By contrast, contractualists have great difficulty accommodating any obligations to future people at all.

Contractualist accounts of intergenerational obligation face two barriers. The first is Parfit’s non-identity problem (Parfit 1984, Part 4). Contractualists model morality or justice on a bargain or agreement among rational individuals. But how can we begin to imagine contracts, bargains, or cooperative schemes involving future people whose existence and identity depend upon what we decide? Contractualists as diverse as Immanuel Kant, John Rawls, David Gauthier, and T. M. Scanlon all face serious difficulties here. The second barrier is the lack of reciprocal interaction between present people and distant future people. We can do a great deal to (or for) posterity, but, as the saying goes, what has posterity ever done for us? We cannot bargain, negotiate, or cooperate with those who will live long after us. A contract with distant future people seems incoherent.

Of course, many contractualists do try to accommodate intergenerational justice. They cite the motivations of present people, exploit contracts between overlapping generations, appoint trustees or ombudsmen for the future, or construct imaginary intergenerational bargaining situations where the parties know neither when nor whether they exist (Gosserries and Meyer 2009). But these intergenerational contracts all seem troublingly ad hoc. It is hard to escape the conclusion that, for the consistent social contract theorist, intergenerational justice is (at best) an afterthought – an optional extension of a theory of justice designed for contemporaries. Contractualists cannot accommodate the future as easily as utilitarians do.

If conflicts between generations were rare, or if we could be confident that future people would be better off, then this comparative weakness of contractualism might not matter. (After all, no theory is perfect, and utilitarianism certainly has problems of its own.) But, if we face a broken future, then our need for a credible account of our obligations to future people is must greater. This doesn’t prove that utilitarianism is superior all-things-considered, but it does significantly enhance its comparative appeal.

Our third lesson is that some moral theories handle a broken future better than others. Working through the ethical implications of the broken future, I have been struck by the number of different ways that philosophers help themselves to optimistic assumptions about the future. Consider four disparate examples: traditions in naturalistic meta-ethics that identify moral facts with the end-points of processes of empirical moral inquiry that may turn out to be inextricably linked to an unsustainable way of life (Mulgan 2015); the many strands of
contemporary moral philosophy built on *intuitions* about simple cases – intuitions that are very closely tied to our affluent present (Mulgan 2014a); libertarians who presuppose that initial acquirers can leave ‘enough and as good’ *for all future people* (Mulgan 2011, pp. 18-68); or Rawlsian liberals who insist that ‘justice’ only applies while favourable conditions persist (Mulgan 2011, pp. 160-196). The recognition of a credible broken future thus counts against naturalist meta-ethics, intuition-based ethics, libertarianism, and (Rawlsian) liberalism. As philosophy is a comparative business, the broken future thus supports non-naturalism, theoretical ethics, and alternative political philosophies such as utilitarianism.

The fourth impact of the broken future is that it raises troubling practical questions about how we should live now. Can we reasonably justify a refusal to adopt the ethical outlook of the broken world *for ourselves*? If future people will be worse-off, partly as a result of our actions, should we reduce our aspirations, and bring our notion of what is necessary for a worthwhile human life into line with theirs? Can we insist for ourselves on goods and opportunities that will not, as result of that very insistence, be available to future people?

The removal of favourable conditions raises an even more disturbing question. Suppose we conclude that, while we can guarantee our own basic needs, our descendants will need to run a survival lottery. Can we still insist on guaranteed survival for ourselves, or should we move in their direction – operating a survival lottery across the generations? (And what would that lottery look like?) The survival lottery strikes us as morally unthinkable. But if we leave future people in a place where they must think the unthinkable, then perhaps we should think it too. Perhaps the design of a just survival lottery should be *our* central philosophical concern as well. (I develop a survival lottery inspired by J. S. Mill’s liberal utilitarianism in Mulgan 2014d.)

II. THE VIRTUAL WORLD

Imagine a *virtual future* where people have abandoned the real world altogether and spend their entire lives plugged into an experience machine that perfectly simulates any possible human experience (Nozick 1974, pp. 42-45). Perhaps this virtual reality is the best option in a broken world. The natural environment is so polluted, and so resource-poor, that people have been forced to dream away their lives with no direct contact to any reality outside the machine. This is not a sceptical scenario. Future people are fully aware that their reality is merely virtual. But this is all anyone has ever known, and they find it perfectly satisfactory. No-one misses bird song, clean air, blue skies, or any of the other wonders their rapacious ancestors have destroyed.
When Nozick first presented it in 1974, the experience machine was science fiction. In 2013, the virtual world is one credible future. Something like this could well happen. Even if we discount the hype surrounding all new technologies, no-one can be confident that genuine virtual reality will not emerge. Like the broken world, the virtual world may not be our immediate future. And it may never happen. (The future might be so broken that the necessary technology never develops.) But it is one medium-term possibility.

My specific virtual future would still be worth exploring even if it were not credible, because the issues it raises are relevant to a wide range of very plausible futures. Even if future people never face a choice between instituting a survival lottery and retreating forever into a world of perfect illusion, they will confront more mundane choices where technological or economic advances conflict with long-cherished connections to the natural world. (Should we put our energies into conservation or into developing ever-more elaborate video games?) And while perfect virtual reality may remain forever elusive, less-than-perfect experience machines could still be very appealing for future people living in (more or less) broken worlds.

As with the broken future, I shall take the credibility of the virtual future as given, and explore its implications. We must first address a new question. No-one advocates a broken future, and it is obviously something to avoid. By contrast, the virtual world has many advocates, and some people are striving to make it a reality. Our first question, then, is whether the virtual world is desirable. Should this future worry us? Should we try to avoid it? If individuals are the best judges of their own interests, then the virtual future is unobjectionable. All that matters is that people are content with their lot. But many of us do find the virtual future very worrying indeed. My aim in this section is to explore that worry, and to see where it leads.

My virtual future is modelled on Nozick’s experience machine. Nozick himself argues forcefully that it is a mistake to choose the experience machine. Experience is not the only thing that matters. We want to do things, not merely to have the illusion of doing them. And we need a connection to some reality that is deeper than the imagination of a video game designer.

Nozick’s thought experiment is so powerful because most people share his reaction. Most of us agree that something vital is lost if one spends one’s entire life in a virtual world, however perfectly it replicates the real thing. My virtual future is considerably worse than the experience machine. In Nozick’s original tale, each individual decides for herself whether to enter the machine, and then selects her own experiences. In my new tale, one generation imposes a particular virtual future on another. Future people prefer their world, but they have not
chosen it. Anyone who thinks it is a mistake to enter the experience machine should find the imposition of a virtual future especially troubling.

For the present discussion, I stipulate that the imposition of a broken future is (at least) morally problematic. Something of great value is lost in the transition to a broken future, and the fact that its inhabitants do not mind only makes things worse. If we impose a virtual future, we harm its inhabitants. Unless we have a very good excuse, our actions are wrong. These moral judgements are not uncontroversial. But they are widely shared, and it is therefore worth asking where they lead. Every moral inquiry has some controversial premises, and this will be one of mine.

Suppose a virtual future is both credible and undesirable. How does this affect our current ethical thinking? We begin with an old lesson from the broken world. The virtual future reinforces the importance of our obligations to future people, because it represents another way that future people might be worse off than present people. We cannot blithely assume that new technology will enable future people to escape the broken world. Some technological ‘solutions’ are themselves undesirable.

Our first new ethical lesson is that the virtual future supports an objective list theory of human well-being (Parfit 1984, Appendix I). This theory offers a list of things that are good in themselves irrespective of the agent’s attitude to them, such as knowledge, achievement, friendship, individuality, self-development, and so on.

Contemporary debate about well-being contrasts three positions: hedonism (well-being is pleasure and the absence of pain); preference-theory (well-being is getting what you want); and objective list. Objectivists argue that neither hedonism nor preference-theory is satisfactory. Some pleasures are good, some bad, others are neutral. Some preferences improve your life, while others do not. Consider a child who wants to play in the sand rather than go to school. Clearly, we make his life go better if we send him to school. The challenge is to explain why. Education doesn’t simply help people to satisfy their existing preferences. It also teaches them what to desire, and which pleasures to seek. It is important to satisfy people’s desires only because what they value is independently worthwhile. The objects are not valuable because they are desired — they are desired because they are valuable.

A theme of my recent work is that only the objective list theory captures the full range of our obligations to future people (Mulan 2012, 2014b). Hedonism and preference theory are inadequate. The more
importance we attach to those obligations, the more serious this comparative advantage becomes.

Nozick’s experience machine is often read as a decisive refutation of hedonism. Life in the machine is phenomenologically indistinguishable from the ‘real thing’. If it is a mistake to enter the machine, then there must be more to human flourishing than the quality of one’s experiences. Our negative reaction to the virtual world supports this critique, as the hedonist must also find that world unobjectionable.

The impact of the experience machine on preference-based accounts of well-being is harder to judge, for reasons that will emerge shortly. This is where our new tale comes into its own. The preference-theorist cannot capture our unease about the imposition of a virtual future, because its inhabitants are content with their lot. If we only look at individual preferences, then we cannot see what is wrong with avoiding our obligations to future people simply by manipulating their psychology – or their environment – so that they never want the good things we destroy.

By contrast, the objective list theory easily capture both Nozick’s reaction to the experience machine and our reaction to the virtual future. If a connection to the natural world is intrinsically valuable, then human lives go better (and perhaps can only go well) when they instantiate that value. Some things matter, and it matters that people are connected to real values, not virtual ones. Even Peter Singer, the most prominent contemporary defender of preference utilitarianism, has recently acknowledged, on the basis of very similar examples, that we need a more objective account of well-being to make sense of our obligations to distant future people (Singer 2011, p. 244).

Unlike Nozick’s original, and countless other fantastical tales, our virtual world is a credible future. This realisation greatly strengthens the objectivist critique of hedonism and preference-theory. In debates over well-being, as in many other philosophical topics, every theory stumbles over some ingenious imaginary case. It is therefore easy for partisans of hedonism or preference-theory to set aside Nozick’s experience machine. No theory is perfect, after all.

We cannot demand that a theory of well-being perfectly fit all our intuitions about imaginary cases. But we can reasonably insist that moral philosophy provide useful guidance about important actual decisions. An acceptable theory of well-being must help us to think clearly about our obligations regarding credible futures, especially when our present choices might harm future people. Like the limitations of social contract theory, the inadequacies of hedonism and preference-theory can no longer be ignored.
Singer’s own conversion is instructive. As a practical ethicist, Singer focuses on first-order moral issues, such as abortion, our treatment of animals, or our obligations to the distant poor. His shift away from preference utilitarianism is driven by the failure of his own attempts to apply it to the newly urgent practical questions posed by climate change. The practical ethicist can side-step the experience machine, but not the virtual future.

The virtual future also teaches us two other lessons about objectivity. Our second lesson concerns liberal neutrality. Preferences play two distinct roles in contemporary ethics. First, they ground a substantive account of what human well-being is. (A good life is one where you get what you want.) Second, they provide the standard liberal response to pervasive disagreement about well-being. (A just society allows people to decide for themselves.) We have seen that the virtual future undermines the former. It also undermines the latter. To see how, we return to Nozick.

While most commentators agree with Nozick, his reaction to the experience machine is not universally shared. My own experience (having taught this example for nearly twenty years) is that people divide quite sharply. Some are enthusiastic to enter the machine; others are vehemently opposed; and many people, while interested in principle, are wary of putting that much trust in any human technology.

In real-life, we would accommodate this disagreement by allowing each person to choose for herself. (This is certainly what Nozick would do!) If individual experience machines were available, most liberal societies (and all libertarian philosophers) would defer to individual preferences. (The alternative, after all, is for one person to impose her preferences on others.) The appeal of this liberal neutrality goes beyond those who endorse a preference-based substantive account of well-being. Even most proponents of the objective list theory, who believe it is a prudential mistake to plug-in, would still defer to the individual’s right to make her own mistakes. When each individual chooses for herself, we can combine Nozick’s libertarian deference to individual preferences with his claim that it is a mistake to enter the experience machine. Whatever their own personal substantive views about well-being, liberals and libertarians can thus remain officially neutral about the substantive issue.

Thinking about credible futures calls into question our habitual liberal deference to individual preferences. We saw an illustration of this in Section I. One reason the survival lottery is so disturbing is precisely because it takes the most important individual decisions, and subjects them to collective deliberation. The virtual future teaches a similar lesson. Liberal neutrality becomes problematic when you choose the experience machine, not for yourself, but for your descendants. And
neutrality collapses entirely when we collectively impose a virtual world on all future people. When present actions set the parameters for the choices of future people, we cannot responsibly defer to their (adaptive) preferences. We must ask, not what they will want, but what is worth wanting – not what will be desired, but what is desirable.

The virtual future raises the stakes. Our favoured account of well-being is no longer merely a guide to individual prudential decisions within a neutral liberal framework. It must ground collective decisions about the human future. If the objective list theory is to play this role, then it must be based on robust moral facts about objective values. This is the third lesson of the virtual future: it pushes us in the direction of a more objective story about morality itself. This raises two further difficulties. First, we need a meta-ethical story that allows for moral objectivity. Second, if morality involves matters of objective fact, then even our most cherished moral beliefs may be mistaken. Our final two credible futures are designed to bring these two difficulties to the surface.

III. THE UNCONSCIOUS DIGITAL WORLD

Imagine a world where flesh-and-blood humans have been replaced by digital beings. Some are intelligent machines, while others are digital copies of human brains. (Perhaps some well-intentioned machine has kindly upgraded us to a more durable platform.) But this digital future is unconscious. Both intelligent machines and digital humans lack any phenomenological experience or inner life. In J. J. Valberg’s apt phrase, unlike each of us, no digital being finds herself at the centre of an ‘arena of presence’ (Valberg 2007).

Digital futures are credible. No-one can be confident that superintelligent machines will never emerge. And digital beings might not be conscious. The question of machine consciousness is a site of reasonable philosophical disagreement. Consciousness might be nothing but patterns of information processing (e.g. Hofstadter 2007), but it might instead be an emergent feature specific to our biology (e.g. Searle 1997). Consciousness and intelligence may always go together. But, for all anyone knows, they may sometimes come apart. It is thus worth asking what would follow if they did.

Conscious digital entities are a stock device in science fiction, as is the specific trope of copying a person into a computer. A presumption of digital consciousness dominates popular culture. My unconscious digital future is counter-intuitive. But that is a reason to explore it, not to ignore it.

The unconscious digital future reinforces several earlier lessons. It provides new reasons to be suspicious both of technological solutions to
the broken world, and of future preferences. (If intelligent machines pass
the Turing Test, then our descendants may not realise that digital life is
unconscious until it is too late.)

This digital future has two new lessons for current moral thinking.
First, it calls into question one popular moral reading of science fiction.
We often identify moral progress with a broadening of ethical concern to
embrace people of all religions, races, and genders, and even other
sentient animals or the environment. Science fiction invites us to further
expand our ethical circle to include aliens and digital beings.

The presumption of digital personhood is then a moral imperative,
rather than a metaphysical hypothesis. You must always treat intelligent
machines as persons, and therefore as conscious beings. Otherwise, you
risk treating conscious digital people as if they were merely unconscious
things. Even if conscious and unconscious digital futures are both
credible, we should always assume the former.

This presumption of digital personhood turns out to be another
moral principle that works for contemporaries, but does not translate
easily to future people. The presumption makes sense when we
encounter already existing digital beings on some distant planet. But
once we ask whether to create digital beings, and especially whether
humans should transform themselves into digital beings, then there are
very significant risks on both sides. If we falsely assume that our digital
descendants are unconscious, then we risk losing vast improvements in
human well-being. But, if we falsely assume that digital beings are
conscious, then we risk the total annihilation of human value. This is a
new ethical predicament, because no credible future raises analogous
doubts about other expansions of ethical concern. (We need not worry,
for instance, that animals will turn out not to be sentient.)

Our second lesson is every more troubling. Suppose that future
(unconscious) intelligent machines regard consciousness as unimportant.
In their view, what really matters about either a machine or a human
being is not her experiences (whatever those are) but her patterns of
thought. If pattern alone matters, then even an unconscious digital future
is much better than any future containing embodied humans. Is this just
a case of competing preferences? Or is it a matter of moral fact? If so,
who is correct? We naturally think that the unconscious digital future is
very undesirable. But what if we are wrong?

IV. A THEOLOGICAL FUTURE

Imagine a future where some clever philosopher, or some even cleverer
machine, has proved the existence of God; or where future science posits
a cosmic purpose to explain the nature and existence of the universe.
(This is not just a future where everyone believes in God, but one where there is a God whose existence has been proved.)

This is a credible future. Our universe is religiously ambiguous. The available evidence is open to radically different interpretations. For all anyone knows, there might be a God. But this ambiguity might itself reflect our cognitive limitations. Perhaps the truth will be obvious to smarter beings. (After all, many people believe that God’s existence can be proved. Perhaps they will turn out to be right.)

It is interesting to ask whether liberal ideals could survive into a future that lacks reasonable religious pluralism. I leave that task for another day, and end with some brief lessons for current ethical thinking.

Taking God seriously raises the credibility of several of our earlier tales. If there is a spiritual person beyond the physical universe, then the possibility that consciousness is not entirely reducible to physical patterns is more credible. Perhaps God only gifts consciousness to God’s creatures, not ours. And, whatever we think of the prospects of atheist moral realism, a morally perfect personal creator certainly makes objective morality more plausible. If God has a purpose for the cosmos, then it is much easier to see how we can make moral mistakes, and why a connection to reality is so important.

We might expect God to prevent a broken world. My theological future would then undermine my broken one. A more worrying possibility is that future non-human philosophers or scientists will discover a distinctly non-human God. These future metaphysical discoveries might then reveal new realms of objective value. Indeed, this is what we should expect. If machines succeed in proving the existence of God where humans have failed, then they probably are more central to God’s plans. Perhaps our theological future is also an unconscious digital one. Digital philosophers prove that pattern is what matters, and that the valuable patterns are distinctly un-human – more intricate, more complex, or just very different. All efforts are now devoted to truly valuable patterns, and human patterns are left to dissolve. From our human perspective, a theological future could thus be very broken indeed. (I explore the possibility of a God who is indifferent to human beings, and the implications for morality, in Mulgan 2015.)

Is it easy to imagine a future where people think differently about value or God. It is much harder to take seriously the possibility that those future people might be right. But, if we are to imagine the most challenging and unsettling possible futures, then this is what we must do.
REFERENCES


PRESIDENT: David Papineau (KCL)

PRESIDENT-ELECT: Adrian Moore (Oxford)

HONORARY DIRECTOR: Lucy O’Brien (UCL)

EDITOR: Matthew Soteriou (Warwick)

LINES OF THOUGHT SERIES EDITOR: Scott Sturgeon (Oxford)

EXECUTIVE COMMITTEE: Sarah Broadie (St. Andrews) / Ben Colburn (Glasgow) / Alison Hills (Oxford) / Rosanna Keefe (Sheffield) / Samir Okasha (Bristol) / Ian Rumfitt (Birkbeck) / Robert Stern (Sheffield)

EXECUTIVE ADMINISTRATOR: Mark Cortes Favis

ASSISTANT EDITOR: David Harris

WWW.ARISTOTELIANSOCIETY.ORG.UK