How should utilitarians think about the future?

Tim Mulgan (Philosophy, University of Auckland, New Zealand; and Philosophy, University of St Andrews, United Kingdom)
t.mulgan@auckland.ac.nz

ABSTRACT

Utilitarians must think collectively about the future, because many contemporary moral issues require collective responses to avoid possible future harms. But current rule utilitarianism does not accommodate the distant future. Drawing on my recent books *Future People* and *Ethics for a Broken World*, I defend a new utilitarianism whose central ethical question is: What moral code should we teach the next generation? This new theory honours utilitarianism's past and provides the flexibility to adapt to the full range of credible futures – from futures broken by climate change to the digital, virtual, and predictable futures produced by various possible technologies.

BIOGRAPHICAL NOTE

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), *Ethics for a Broken World* (Acumen/McGill-Queens University Press 2011), and *Purpose in the Universe: the moral and metaphysical case for ananthropocentric purposivism* (Oxford University Press, 2015).

KEY WORDS: utilitarianism, rule utilitarianism, future people, intergenerational justice, broken world, moral imagination, virtual reality

Drawing on my recent books *Future People* and *Ethics for a Broken World*, I defend a new utilitarianism whose central ethical question is: What moral code should we teach the next generation? I argue that this new theory both honours utilitarianism's past and provides the flexibility to adapt to the full range of credible futures – from futures

broken by climate change to the digital and virtual futures produced by various possible technologies.

1. Ideal moral outlook utilitarianism.

I first outline my approach to utilitarianism (Mulgan 2001, 2006, 2007, 2015d). I begin with a simplified contrast between two strands in the utilitarian tradition. *Act utilitarianism* says that the right act in any situation is whatever produces the best consequences. Act utilitarianism is notoriously demanding, alienating, and counter-intuitive. It is also inadequate for contemporary environmental problems, which can only be understood collectively. (It makes little sense to ask what *I* should do about climate change.) *Rule utilitarians* seek a moderate, liberal alternative to act utilitarianism. They picture morality as a collective enterprise, and evaluate moral codes by their collective impact on human well-being. For rule utilitarians, the fundamental moral questions are: 'What if *we* did that?', and 'How should *we* live?'

I develop a new theory based on Brad Hooker's recent rule utilitarianism (Hooker 2000; Mulgan 2001, 53-103, 2006, 130-160). Any rule utilitarian theory consists of two stages. We first identify the *ideal moral code*, and then we assess acts *indirectly*: the right act is the act that would be performed by someone who had internalised the ideal code. My main concern in this paper is the first stage: identifying the ideal code.

Although 'rule utilitarianism' is now the established name, it is misleading. Talk of 'rules' and 'codes' is distracting. I focus instead on the *Ideal Moral Outlook* (IMO), and therefore on *Ideal Moral Outlook Utilitarianism* (IMOU). This more cumbersome name leaves open whether the utilitarian ideal is a code of rules, a set of dispositions, a package of virtues, a set of priorities, a general moral outlook, or (as seems most likely) some combination of these.

The IMO is the outlook that best promotes human well-being. But a moral outlook on its own cannot promote anything. We must first map candidate outlooks onto specific possible futures, and then rank those possible futures using our preferred utilitarian value function. IMOU ranks competing moral outlooks by asking what would happen if we (the present generation) attempted to teach that outlook to the next generation.

The *ideal* outlook is the best one for us to teach to them. This sets aside the cost of changing existing moral beliefs, but factors-in the cost of (for instance) trying to get a new generation to accept a very demanding ethic.

IMOU includes several vague terms: 'best', 'us', 'teach', 'next generation'. Our next task is to cash these out. Contemporary moral philosophy offers many ways to evaluate states of affairs (e.g., Parfit 1984, 351-454; Broome 2004). Should we consider actual consequences, objective expected value, or subjective expected value? Should we maximise or satisfice? Should we give priority to the worst-off, to people near some significant threshold of well-being or lexical level, to basic needs, to human rights? Should we be risk averse – giving priority to avoiding especially bad results? Should we discount the future or be temporally impartial? How should we aggregate the well-being of individuals within a population: total utilitarianism, average utilitarianism, diminishing marginal view, lexical view, critical level view? Should we give direct weight to factors other than well-being – such as distribution, equality, desert, rights, justice, ideal goods, etc?

In theory, IMOU could incorporate any package of views about value and aggregation. I take the simplest option: asking what maximises expected total well-being. I offer three justifications. First, IMOU's distinctive features emerge most clearly if we eschew eccentric views about value, aggregation, or risk-aversion. Second, maximising expected total utility is the default starting-point in the contemporary literature and has many theoretical advantages over its rivals (Broome 2004). Finally, I argue elsewhere that the most common (and most compelling) objections to 'total utilitarianism' are really objections to the combination of total utilitarianism and a particular (extreme) account of right action: namely, act consequentialism (Mulgan 2006, 55-81). Combined with act consequentialism, total utilitarianism makes extreme demands. But act consequentialism is equally counterintuitive whatever its axiology. Total utilitarianism is much more palatable when incorporated into IMOU.

One issue for all collective consequentialists is scope (Hooker 2000, 169-174; Mulgan 2001, 53-103; Kahn 2012). Does 'we' include all rational agents at all times, or some proper subset of them, or even all *possible* rational agents? Competing desiderata

point in different directions. On the one hand, wider scope increases the distance between collective consequentialism and act consequentialism – thereby enhancing whatever comparative advantages drew us to collective consequentialism in the first place. (For instance, wider scope typically means less onerous demands (Hooker 2000, 169-174).) Also, any particular restricted scope looks arbitrary. Why restrict our attention to only those rational agents? Why draw the line exactly *there*?

On the other hand, universal scope threatens to make collective consequentialism unwieldy, indeterminate, and impractical. How could we ever know what it will be best for *all* rational agents to do? Is there even any fact of the matter to be discovered? Wider scope also exacerbates familiar worries about 'rule worship' (Hooker 2000, 93-99). Why think that the appropriate thing *for me to do* is to follow a rule that would only bring good consequences if it were followed by countless billions of others who will never actually do so?

In this paper, I largely set questions of scope aside. Unless otherwise stated, 'we' includes all present adult members of a large industrial society. Scope is a general difficulty for any collective consequentialism, not a specific problem for IMOU. IMOU can borrow the best rule utilitarian solution, whatever that turns out to be, with one exception. Given its emphasis on teaching the next generation, IMOU insists that 'we' can only include *present adults*: it cannot include future people.

We aim to identify the IMO. Scope only matter if it affects which outlook is best. In many cases, disambiguation is unnecessary because the same outlook comes out best whoever 'we' are. This follows from another feature of rule utilitarianism that IMOU shares: pervasive uncertainty. We cannot hope to specify the IMO in any detail. At most, we can identify some of its general features. But often that is enough. We might, for instance, be confident that the IMO does not support gratuitous torture, even if we don't know exactly what it says about other issues. Similarly, we might be confident that different scopes deliver the same moral outlook. It is unlikely that all human beings would do best to encourage torture while everyone in our society does best by avoiding it!

IMOU is closer to traditional rule utilitarianism than it initially appears.[i] 'Teaching' is not limited to explicit preaching about ethics. It also includes implicit lessons, role-modelling, story-telling, exemplification, and any other present behaviour that impacts on the moral outlook of those who are influenced by us. Furthermore, the 'next generation' is an abstraction. Human beings are not bumble bees, arriving neatly packaged into discreet generations. The next generation is everyone directly influenced by our teaching.[ii]

IMOU is still distinctive in two ways. First, it focuses on the impact of our actions *on other people's moral outlook* (thus setting aside all other consequences); and second, it considers only people we directly influence – those we teach, as opposed to (say) people living in the distant future whose moral outlook may be influenced by our actions in many indirect ways.[iii] My aim in this paper is to justify these distinctive features of IMOU.

IMOU always asks the same question: What can we teach now that will maximise well-being into the future? But this constant question receives different answers across the generations. What best promotes well-being in our generation may be neither what would have been best in the past nor what will be best in the future. IMOU begins with a simple question, and then allows complexity to emerge empirically, because the answer to that simple question depends on facts about human nature and about our ability to teach or learn moral codes. Our central question is timeless in its formulation, but context-specific in its application.

IMOU is distinctive in the recent rule utilitarian literature. I argue that it has four main advantages over its rivals.

1. IMOU is closer to the spirit of the classical utilitarians, especially Jeremy Bentham and J. S. Mill. They too began with simple moral principles and allowed complexity to emerge empirically in response to our evolving knowledge of human nature and the human situation. IMOU also captures a perennially attractive picture of morality as a collective human enterprise passed on from one generation to the next.

2. IMOU's central empirical question is of independent interest, especially to moral educators. Many other rule utilitarians ask questions that could never relate to any possible practical situation. (No-one is ever in a position to choose whether or not everyone everywhere will follow some moral code.) By contrast, our new question is practical. Indeed, if we interpret 'moral teaching' broadly, then that question is inescapable. We *will* teach the next generation *some* moral code. IMOU asks what we *should* teach. Even if we are not rule utilitarians – even if we don't want to use the ideal code to judge individual actions – we surely do want to know which moral code it would be best to teach. Moral philosophers, moral educators, and others who observe that moral outlooks have changed in the past, may wonder how they might change in the future. And this prompts the further question: How *should* moral outlooks change? If we could get the next generation to follow/adopt/internalize any moral outlook, which one should it be? If our job is to influence the next generation's moral outlook, and we are at all sympathetic to utilitarianism, then the IMO obviously matters.
3. I argue elsewhere that IMOU solves some puzzles that plague traditional rule utilitarianism (Mulgan 2006, 130-160). For instance, it neatly by-passes debates about what percentage of the ideal population should be assumed to have internalised the ideal code, and to what extent. (Hooker 2000, 80-85; Ridge 2006; Hooker 2008; Smith 2010.) Rule utilitarians face a familiar dilemma. If we idealise to perfect compliance, then we get a code of rules that has no resources to cope with other people's wrongdoing – because we test our codes by imagining a world without any wrongdoing. But any specific level of partial compliance seems ad hoc. IMOU provides a simple solution. Instead of being stipulated ad hoc in advance, different degrees of internalization reflect the relative difficulty of *teaching* different moral codes in different circumstances.[iv]
4. IMOU accommodates our obligations to distant future people much better than its rivals. This final advantage is the focus of this paper.

2. Why utilitarians must think about the future.

Surprisingly, rule utilitarians appear not to have thought much about the future.[v] Rule utilitarianism is typically static, concentrating on short-term impacts on present people. Despite their obvious importance, future consequences often drop out of the picture very quickly. This temporal myopia doesn't distinguish rule utilitarianism from most other contemporary ethics, but it is at odds with central utilitarian commitments.

Perhaps some non-utilitarians can simply ignore the distant future. If we base justice on reciprocity, actual sentiment, existing relationships, shared projects, or any other connection that is lacking between us and distant future people, then we can safely set the distant future aside. Alternatively, if we adopt a high pure temporal discount rate, then the fate of distant future people fades into irrelevance. (Discount rates are especially relevant in debates about the optimal response to climate change. See, e.g., Broome 2009, 133-155.) But no utilitarian can dismiss the future so easily. A fundamental commitment of *utilitarian* justice is that well-being matters equally – no matter whose it is *or when it occurs*. Our value theory is temporally impartial, and our pure temporal discount rate is zero (Cowen and Parfit 1992).

Rule utilitarians should not ignore the future. But temporal impartiality notoriously threatens extreme demands for present people, because infinitesimal improvements in the lives of innumerable future people will dominate our moral calculus. So far as I am aware, rule utilitarians have not yet explicitly confronted this threat. One possible explanation for this omission is that rule utilitarians are confident they can discount the distant future on instrumental grounds. Even if our goal is to maximise long-term well-being, things go best if we focus on present people.

Rule utilitarians who ignore distant future people could offer three familiar justifications:

> 1. *Ignorance*. We cannot evaluate the long-term consequences of our actions, and so we cannot take distant future people into account.
>
> 2. *Convergence*. What is best for present people is also best for future people, and so we need not consider them explicitly.

3. *Improvement.* Future people will be much better-off, and (thanks to diminishing marginal utility) utilitarians should adopt a general principle of discounting additional benefits to better-off people.

These are all good utilitarian arguments. But they are also all contingent on two key presuppositions about the distant future:

1. *Optimistic Convergence.* If we do what is best for ourselves, then distant future people will be better-off.
2. *Environmental Stability.* Human actions do not affect the natural background conditions over the long-term, and therefore have only minimal impact on distant future people.

Unfortunately, these presuppositions are now highly questionable. In my book *Ethics for a Broken World*, I imagine a future broken by climate change, where a chaotic climate makes life precarious, each generation is worse-off than the last, Rawlsian favourable conditions (Rawls 1971, 178) no longer apply, and our affluent way of life is no longer an option (Mulgan 2011, 2014a, 2015a, 2015c, 2015d, 2016a, 2016b).

This is a *credible* future. No-one can reasonably be confident that it won't happen. When we teach our ideal code to the next generation, we must consider the possibility that they (or their descendents) will inhabit a broken world. But then our two presuppositions no longer hold. Distant future people may be worse-off because of what we have done to their environment.

While rule utilitarianism has hitherto largely ignored the future, IMOU forces us to think directly about the situation of the next generation; and also indirectly about their impact on later generations. IMOU thus counteracts our natural human tendency toward temporal myopia, thereby making us much more likely to give intergenerational issues their due.

A number of recent developments raise the importance of future ethics. (Consider threats such as climate change, resource depletion, over-population, dangerous

technological acceleration, etc.) These developments thus also increase the appeal of IMOU. One central moral task for all utilitarians is to equip future people for the moral challenges they will face. IMOU highlights the fact that teaching a moral outlook to future people is the most important thing we do.

3. Towards an intergenerational utilitarianism.

We need a moral theory to guide us in a world where human actions have an unprecedented long-term impact on the natural environment. Utilitarianism has a distinct advantage in this challenging new territory, because it presupposes neither optimistic convergence nor environmental stability, and easily absorbs new information about the future. (Utilitarianism thus contrasts favourably with Rawlsian liberalism, for instance, which explicitly limits itself to favourable conditions where optimistic convergence holds true.)

When thinking about the future, rule utilitarians must first ask the right question. One mistake would be to try to discover the moral code of distant future people – asking how they will live, and then seeking to emulate them. This makes for great science fiction, but not for credible moral philosophy or practical ethics. We cannot hope to predict the ideal distant future code. And even if we found it, that code would probably be too alien and demanding to be of any practical use to us.

Another tempting mistake is to ask what code would provide the best results if *it* were followed from the present day into the distant future. This presumes that one single moral code will pass down the generations more or less intact. This is extremely implausible. It also contradicts an enduring utilitarian commitment. A central feature of Mill's utilitarianism, in particular, is his faith in empiricism and moral progress (Skorupski 1989, especially 1-47; Donner and Fumerton 2009, especially 59-60). It may seem incongruous to speak of moral progress alongside broken futures, but there is no contradiction here (Mulgan 2015a). Mill is not a naïve optimist about social progress. Instead, he is an empiricist who is confident that future people will know more than us about what is valuable and that future ethical inquiry might move in very surprising directions. At the very least, utilitarians should not build the absence

of moral progress into their theory at the outset by assuming an unchanging moral code! (I return to moral progress in the final section of this paper.)

A third mistake – as we saw in the previous section – is to think we can simply ignore the future altogether.

We need a utilitarian question that counts distant future people equally without asking us to imagine or imitate their moral thinking, and also allows for moral change. My new IMOU fits the bill perfectly. IMOU focuses directly on the next generation and only indirectly on the distant future. It does not ask: What is the code that would maximise well-being if it were followed by all subsequent generations? IMOU asks only what we should teach *to the next generation*. On the other hand, the moral codes of later generations do enter our utilitarian calculations, because they are important consequences of our initial teaching.

A central debate within utilitarianism is when we should idealise, and when we should be realistic. For instance, act utilitarians idealise only *my* actions – and keep everyone else's behaviour fixed. At the other extreme, some rule utilitarians implicitly idealise the behaviour of all moral agents – both present and future. IMOU takes a middle-road. It idealises only the behaviour of our generation. IMOU asks what it would be best for us to teach, but it does not project this imaginary utilitarian thought experiment into the future. IMOU does not ask: 'What code would it be best for the next generation to teach the third generation?' Instead, it asks: 'What will actually happen if we teach this code to the next generation?' This involves also asking what code the next generation will *actually* teach – and what effects their teaching will have on later generations. After all, this is how moral education works. We teach the next generation. We cannot teach distant future people.

4. What is the ideal moral outlook?

We now turn – at last – to the content of the IMO. What *should* we teach the next generation? I begin by dismissing two simple answers. The first is that things go best if everyone always tries to maximise well-being. The IMO is the utility principle and IMOU collapses back into act utilitarianism. All rule utilitarians must deny this

collapse. Otherwise they would not have a distinct theory! They cite human fallibility, partiality, and cognitive limitations to explain why utility is not maximised by teaching human beings to be single-minded utility maximisers (Hooker 2000, 93-99). IMOU borrows this reply.

Another simple answer is that the IMO is identical to our existing common-sense morality. Drawing on arguments made famous by Mill – and before him by William Paley, who argued that God instituted morality to promote human well-being – rule utilitarians have long argued that the ideal code must include the familiar permissions and obligations, and the rights and freedoms, of moral common-sense. Some rule utilitarians embrace this collapse into conventional morality. For instance, Hooker argues that 'the best argument for rule consequentialism is that it does a better job than its rivals of matching and tying together our moral convictions' (Hooker 2000, 101; see also Hooker 1994, 29; Miller 2000). What rule utilitarianism adds is an explanation of *why* commonsense rules are correct.

Instead of following Hooker, I argue that IMOU will depart from commonsense morality in crucial areas – or at least that it offers guidance where commonsense morality is silent or confused. In particular, as I argue in the next section, the ideal moral outlook for a broken future diverges sharply from our current moral common-sense. I regard this as an objection to common-sense morality, not a problem for IMOU. Our considered moral judgements have evolved to fit our affluent world. We have no reason to expect them to be (even) prima facie reliable when we contemplate a broken future.

However, unlike the most radical *act* utilitarianism, IMOU also limits the potential alien-ness of the IMO, because that outlook must be taught by current human beings to a new generation of humans. Human nature is not infinitely plastic. Any plausible IMO will include familiar moral dispositions such as honesty, generosity, promise-keeping, courage, murder-aversion, and so on. People who internalise the IMO will not walk callously past children drowning in ponds, take pleasure in the sufferings of others, or reject the basic goods of human life.

Can we say more about the distinctive content of the IMO? I believe we can. Relative to our current morality, the IMO will include both a greater emphasis on benevolence towards (temporally) distant strangers; and a greater appreciation of the long-term consequences of one's actions – including greater awareness of the impact of *collective* actions to which one contributes. (Dale Jamieson proposes a new 'green' virtue of 'mindfulness' to capture what is needed here (Jamieson 2014, 187).)

All utilitarians should find these two elements uncontroversial. Things will obviously go better if people think more clearly about the future. I concentrate on a more controversial and more interesting claim: that the IMO will prompt agents to think more deeply about the nature of value and morality. My third distinctive feature is a new emphasis on moral imaginativeness – especially the imagining of new moral norms suited to various possible futures. In our culture, this task is largely confined to speculative fiction. It has not been prominent in moral philosophy. In future-oriented utilitarian ethics, by contrast, imaginative moral experiments in living (to adapt a phrase from J. S. Mill) will be essential elements in everyone's moral repertoire.

We need moral imaginativeness because we are multiply uncertain about the future. Our ignorance has three overlapping dimensions: empirical, metaphysical, and evaluative. We don't know what will happen, we don't know what the world is ultimately like, and we don't know what really matters.

In the rest of this paper, I illustrate the need for moral imaginativeness by examining three contrasting pairs of possible futures that demand different moral sensitivities. (I explore moral imaginativeness more fully in Mulgan forthcoming a)

5. Broken futures.

The broken world illustrates why utilitarians must think about the future. The need to think about the future clearly influences IMOU. The relevance of details *about* that future is less obvious. Do future people need a *different* moral outlook in different possible futures, or will they simply interpret the same moral outlook in different ways?

The future is uncertain. The IMO must therefore work well across a wide range of possible futures. I will now argue that this flexible outlook differs significantly from the moral outlook it would be best to teach if we knew what particular future our descendents will face. The IMO will also therefore differ from the ideal codes of traditional rule utilitarians who (implicitly) presuppose that the future resembles the present.

Elsewhere, I explore several places where current ethical thinking must be reinterpreted for a broken future. While some specific impacts of the broken world are predictable, others are more surprising. Consider versions of 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 (Jackson 1999; Mulgan 2015b); or the many strands of contemporary moral philosophy built on intuitions that are very closely tied to our affluent present (Singer 1972; Thomson 1976; Mulgan 2015c); or theories of rights and distributive justice that implicitly presume a world where the central elements of a worthwhile life can be guaranteed to everyone (Mulgan 2011, 18-68; 2016a). These familiar ethical ideas must all be re-imagined to fit a broken world.

Due to the scarcity of material resources (especially water) and the unpredictable climate, broken world societies periodically face population bottlenecks where not everyone can survive. They must therefore institute *survival lotteries* – bureaucratic procedures that determine who lives and who dies. And no broken world society will endure unless most citizens regard its actual survival lottery as (at least reasonably) just. A central concern of broken world ethics is thus to design a *just* survival lottery.

'Survival lottery' is a term of art. It may not involve any *actual* lottery. For instance, a *libertarian survival lottery* might simply consist of a collective decision to allow the 'natural' distribution of survival-chances to remain uncorrected. However, broken world *liberals*, *egalitarians*, or *contractualists*, who all seek a fair redistribution of the burdens imposed by scarce resources and chaotic climate, probably *do* need literal lotteries. (For instance, a Rawlsian survival lottery must fairly distribute the benefits and burdens of both social cooperation and the natural lottery.)

Broken world dwellers must be prepared to countenance trade-offs between lives. They must also be willing to sacrifice present basic needs to preserve or enhance their society's capacity to meet future basic needs. In a world of declining resources, a sustainable survival lottery cannot always privilege the present over the future.

Any adequate broken world moral outlook must therefore include a willingness to *contemplate* survival lotteries, to ask which possible survival lotteries are *more* just, and to endorse an existing survival lottery if (but only if) it is reasonably just. Familiar Millian utilitarian arguments about the desirability of broad participation in the design of political institutions are *especially* compelling in a broken world (Mulgan 2011, 133-147). The broken world IMO will definitely not favour unthinking acceptance of the status quo!

By contrast, a central tenet of contemporary liberalism is its *unwillingness* to even contemplate violations of basic rights. A liberal society guarantees a private realm where the individual's rights, needs, and choices cannot be traded-off against the common or aggregate good. Liberal rule utilitarians, following Mill, have long argued that the IMO is liberal in this respect. Things go best if people are secure in the knowledge that their rights act as trumps in the public sphere.

Of course, liberal rule utilitarianism is controversial. Sceptics will object that we cannot expect the IMO to guarantee a liberal private sphere. ('So much the worse for liberalism', say non-liberal rule utilitarians; 'so much the worse for rule utilitarianism', reply non-utilitarian liberals.) However, I believe that, at the very least, liberal rule utilitarianism is one very promising, and hitherto under-explored, approach to our obligations to future people. I therefore propose to grant the liberal rule utilitarian assertion that, *if* we were confident that favourable conditions would persist, then the best moral outlook for us to teach to the next generation *would* include a (more or less absolute) reluctance to trade liberal rights against aggregate well-being. Indeed, that reluctance is *especially* necessary for utilitarians, who endorse a very strong disposition to promote the good. Without strong liberal safeguards, our IMO is too likely to lead to the erosion of rights, freedoms, and other safeguards that, once lost, are very difficult to resurrect.

The best utilitarian moral outlook for a world of enduring favourable conditions is thus particularly *ill-suited* to a broken world. A central feature of any liberal departure from act utilitarianism is an emphasis on individual *rights*. But when *nothing* (not even bare survival) can be guaranteed to everyone, rights must either be abandoned or radically reinvented (Mulgan 2011, 56-68, 113-122, 185-197; 2014c; 2015c). This is *why* the broken world is so ethically unsettling. It is also why we must take broken futures seriously: if our descendants will inherit a world where they must think the unthinkinable, then we do them no favours if we also bequeath a moral outlook that prevents them from thinking it. Indeed, our reluctance to countenance trade-offs may itself *contribute* to a broken future. The longer we refuse to countenance intergenerational survival lotteries, the harsher such lotteries must be once they are (inevitably) introduced.

The best moral outlook for a world of enduring favourable conditions is thus ideal, neither in a broken world, nor even in an affluent present where favourable conditions are under threat. In the former, it offers no advice; while in the latter, it preserves an unsustainable status quo too long, thus spreading the inevitable cost inequitably across the generations.

6. Post-scarcity futures.

If future people face a broken future, then we must radically rethink our ethical teaching. If we *knew* the future was broken, we would at least know how to proceed. But this future is not inevitable. Indeed, some possible futures are quite the opposite. Imagine instead a *post-scarcity world* where some new technology has removed all conflicts over resources. Perhaps nanotechnology has produced 'Cornucopia machines' capable of re-assembling air molecules to create any desired object (Stross 2005). Or perhaps future people have uploaded themselves into a digital realm, where available resources are effectively infinite relative to their digitised wants and needs. (I explore this particular post-scarcity scenario in section 8.)

Affluent liberal society promises to meet all basic needs, but not to satisfy all desires. Once basic needs and basic liberties are guaranteed, the primary focus of the affluent theory of justice is the distribution of resources that remain scarce *relative to desires*.

In the broken world, the affluent promise is broken: not all basic needs can be met. In the post-scarcity world, the affluent limitation is removed: desires no longer conflict, because all can be met simultaneously.

In *Ethics for a Broken World*, some of my imaginary broken world philosophy students wonder how affluent philosophers found anything to talk about (Mulgan 2011, 135). For those students, justice is all about the distribution of (scarce) chances to survive. If survival were guaranteed to all, they ask, wouldn't ethics become trivial? We may be similarly dismissive of post-scarcity 'ethical problems'. Without conflicting desires, what is left to worry about? If future people inhabit a post-scarcity world, then life will be good whatever their moral outlook. So we can safely ignore this possible future, and concentrate on other scenarios where morality does matter.

Unfortunately, we cannot so easily set post-scarcity futures aside. In a post-scarcity world, it is *possible* to simultaneously satisfy all desires. It does not follow that this happens automatically or permanently. Future people need a moral outlook that enables them to realize the post-scarcity promise.

This desideratum may still seem trivial. In an affluent world, liberal egalitarian politics is non-trivial because people's desires conflict – and the rich, powerful, or talented may prefer illiberal or inegalitarian alternatives. But what could possibly compete with the fulfilment of the post-scarcity dream? Why would anyone *not* prefer a world where everyone's every desire was satisfied?

There are several reasons why post-scarcity conflict is still possible. First, powerful individuals whose desires are all *already* satisfied might oppose a new system designed to satisfy everyone else's desires as well. Why take the risk?[vi] Second, people may reasonably disagree about how post-scarcity life should be organised. And we cannot dismiss such disagreements as trivial matters of taste. On any account of human well-being other than crude actual-present-desire-maximisation, some possible post-scarcity scenarios are (very much) better than others. It is good (other things equal) if all desires are satisfied. But it also matters what those satisfied desires are *for*. What will people do with their cornucopia machines? Will they all descend into a drug-induced stupor, or retreat into mindless virtual realities? (What do present

people do with the potentially infinite resources of the world-wide-web?) Will anyone have the incentive or the drive to invent or explore *new* patterns to programme into cornucopia machines, or new ways to spend their (now effectively limitless) leisure time? Cautionary tales of wishes granted by duplicitous literal-minded genies, Asimov-literal robots, and other post-scarcity fictions teach us that a world where everyone gets what they want could be shallow, unstable, or otherwise very grim. (In Stross 2005, for instance, the arrival of cornucopia machines escalates existing social tensions into an all-out civil war.) Finally, unless we imagine beings with radically non-human desires, our post-scarcity world cannot literally involve everyone getting everything they want. People's strongest desires often ineliminably involve other people. And those desires inevitably conflict. (Consider Hobbesian desires for power or pre-eminence; the never-ending consumerist urge to keep up with the Joneses; or the desire for a reciprocal and exclusive romantic relationship.) Once again, it matters how these post-scarcity conflicts are resolved.

I hope to explore the ethics of post-scarcity more fully elsewhere. My present argument only needs two modest claims. First, we cannot set post-scarcity aside on the grounds of triviality. Things could go very badly in such a world if future people lack the appropriate moral outlook. Second, the best utilitarian moral outlook for *this* particular possible future may not work very well in other possible futures. (Conversely, neither a broken world IMO nor an affluent one is well-suited to the peculiar demands of post-scarcity life.)

We do not know whether future people will inhabit a broken world, an affluent world, or a post-scarcity one. How should we decide what moral outlook to teach? Should we maximise expected utility, or give priority to avoiding future disaster – even if the resulting moral outlook might deprive future people of the greater benefits available in affluent or post-scarcity worlds? Even if we are risk-averse, can we be sure that the worst-off possible future people live in *broken* futures? Or might a badly-run post-scarcity world be even worse?

This simple situation involving only three possible futures (broken, affluent, post-scarcity) is challenging enough. Sadly, our real epistemic situation is much more complex. Uncertainty about the brokenness of the future is only one aspect of our

ignorance. In the next two sections, I illustrate another source of ignorance by exploring the complexities surrounding two different post-scarcity scenarios.

7. Evaluative uncertainty and virtual futures.

Utilitarians seek to maximise well-being. But the nature of human well-being is a site of perennial philosophical controversy. We are unsure what makes life worth living. Contemporary debate contrasts three positions: *hedonism* says well-being is pleasure and the absence of pain; *preference theory* says well-being is getting what you want; and the *objective list theory* offers a list of things that are good in themselves irrespective of the agent's attitude to them, such as knowledge, achievement, friendship, and so on (Parfit 1984, 3-4, 493-502; Fletcher, 2013; Crisp 2015).

There are two ways that rule utilitarians might address this pervasive disagreement. The *exclusivist* picks her favourite theory of well-being and ask what maximises *that*. The *agnostic* seeks the ideal code that maximises whatever makes life worth living. She trusts that, *in practice*, preference, pleasure, and objective value all coincide and the same ideal code maximises well-being *whatever well-being turns out to be*.

Unfortunately, exclusivism and agnosticism both come unstuck when we consider possible futures where nuances of well-being really matter. Suppose some not-too-distant future generation must choose between broken and post-scarcity futures. But there is a catch. The latter must be a *virtual future* where people abandon the real world altogether and spend their entire lives plugged into an experience machine that perfectly simulates any possible human experience. This virtual reality is all anyone has ever known, and they find it perfectly satisfactory.

How should we think about the choice between broken and virtual futures? How do we want future people to think about that choice? *On its own terms*, the virtual world is more abundant than any affluent reality (past or present), and certainly better than any broken world. But should those terms be accepted? IMOU asks: How should we teach the next generation to think about such choices?

My virtual future is modelled on Robert Nozick's famous experience machine (Nozick 1974, 42-45). Like all good thought experiments, it works by prizing apart things that typically go together. When pleasure is entirely cut adrift from achievement, which matters more? Nozick's discussion is tantalizingly brief, and his dialectical purpose is unclear. However, one popular interpretation presents Nozick's thought experiment as a reductio ad absurdum of hedonism (Feldman 2011). Elsewhere, following Nozick, I argue that the objective list theory has a decisive advantage over both hedonism and preference theory in relation to virtual futures (Mulgan 2014a, 2014b). My present argument only assumes that the three theories *disagree* about these futures. This claim is much less controversial. Life in the virtual future is phenomenologically indistinguishable from the 'real thing', and everyone is content with their lot. Hedonists and preference theorists will thus accept the virtual future on its own terms, agreeing that it offers perfectly satisfactory substitutes for real goods. (Or, rather, it offers the only *real* good – pleasure or preference-satisfaction.) The choice between broken and virtual futures is simply a choice between brokenness and post-scarcity. And that choice is a no-brainer. By contrast, many objectivists find virtual futures very deficient. 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.

Our uncertainty here is *normative*. Is the virtual a fully adequate substitute for the real? Can what is lost in the virtual transition be replaced without remainder? If we teach the next generation to be thoroughgoing hedonists, they will embrace the virtual future without regret, thus maximising both pleasure and preference satisfaction. This is unproblematic if hedonism or preference theory is correct. But what if both theories are wrong?[vii]

My virtual world is not purely imaginary. It is one *credible future*. (And even if perfect virtual reality remains forever elusive, milder trade-offs between the virtual and the real are already here.) To reliably promote well-being, we need a moral outlook that works across all credible futures. But once virtual futures enter the picture, it makes a huge difference *what* we seek to maximise. It really matters whether pleasure and preference exhaust what is good for people. We can no longer

remain agnostic, assuming that the same moral outlook will maximise pleasure *and* preference-satisfaction *and* objective goods. But it would be reckless, given our philosophical uncertainty, to select one account of well-being and simply seek to maximise *that*. If we select the wrong account, then our 'ideal' moral outlook could be very sub-optimal indeed.

Liberals (including liberal rule utilitarians) argue that the best moral outlook for an affluent present is neutral between competing stories about the good life. They thus hope to avoid the dangers of presupposing the wrong theory of well-being. But a *neutral* outlook could prove disastrous or unworkable in the face of a virtual future. How can we remain neutral when so much is at stake?

It is hard to find the best moral outlook for the *choice* between broken and virtual futures. But now suppose, instead, that a virtual future is inevitable. (Perhaps the world becomes so bleak that there is no other choice. Or perhaps *we* cannot now hope to influence that future decision.) How can we best prepare future people for the choices they will face within their virtual world? It is not obvious that the best moral outlook for the choice *whether* to enter a virtual future is also best for life *within* that future. For instance, suppose we are objectivists about well-being. Faced with the prior choice, objectivism stresses the importance of what could be irretrievably lost in an experience machine. But once the virtual future is inevitable, and especially once it has already arrived, objectivism aims instead to preserve genuine values in a purely virtual environment – and to find more nuanced ways to distinguish one virtual (substitute) good from another. Once again, what we should emphasize in our ethical teaching depends on what challenges we expect future people to face.

8. Digital futures.

Another possible instantiation of the post-scarcity future is a *digital future* where flesh-and-blood humans have been replaced by digital beings – intelligent machines and/or digital copies of human brains (Mulgan 2014a, 2016b, forthcoming b).[viii] A future generation already inhabiting a virtual environment might opt to 'upload' to a fully digital virtual world. Or future people might face an earlier choice between broken and digital futures. (Perhaps only digital beings can survive some catastrophe

that will destroy both real world creatures and non-uploaded virtual life. Or perhaps we have sufficient resources to upload, store, and 'run' a billions minds, but not to preserve a comparable number of brains-in-vats.) Future people must also choose between many *different* digital futures. (Should they opt for destructive uploading or digital copying or the development of non-human-based artificial digital intelligences? And which *form* of uploading or copying or AI is best?)

How do we want future people to think about the choice between virtual and digital futures, or between broken and digital futures, or between different digital futures? And what is the best moral outlook for life within a digital future? As in the virtual future, these questions can easily come apart. Perhaps the sensitivities needed to choose whether to upload (or how to go digital) differ from those needed to take advantage of all the opportunities available *within* any given digital future.

One especially disturbing prospect is an *unconscious* digital future, where both intelligent machines and digital humans lack any phenomenological experience, inner life, or 'qualia'. Unconscious digital futures are credible. Consciousness might be simply a matter of patterns of information processing – something machines could easily share. But it might instead be an emergent feature specific to our biology. Experts – whether scientific, religious, or philosophical – disagree. (Contrast, e.g., Hofstadter 2007 and Searle 1997.)

Our ignorance here is both metaphysical and normative. What counts as a *successful* digital transition depends on what is most *important* about flesh-and-blood human life. And *that* is a normative question. Is the unconscious digital future desirable? If it is bad, how bad is it? Would an unconscious digital future be a catastrophe on a par with human extinction, or can what is lost be replaced (with or without remainder)? If we are risk-averse, is the unconscious digital future the worst possibility – something to be avoided at all costs? Or are there worse possible fates?

With virtual futures, objectivists complain that hedonists cannot see what is missing. In our new future, the risk is reversed. Unconscious intelligent machines might have access to (some) genuine objective values despite lacking any phenomenological experience. But if hedonism is true, then any unconscious future is a valueless void. If

we presuppose either the wrong metaphysics or the wrong story about value, then we risk the annihilation of value itself.[ix]

Future people may face a new ethical dilemma. Should they recognise the personhood of digital beings? (Mulgan 2014a) There are very significant risks *on both sides*. If we falsely assume that our digital descendents are *un*conscious, then we risk losing vast improvements in human well-being, or mistreating real moral persons. But, if we falsely assume that digital beings *are* conscious when they are not, then we risk the total annihilation of human value. (After all, unconscious intelligent machines will probably regard consciousness – 'whatever that is!' – as unimportant. If they take over, they may remove consciousness without thinking twice.) This is a new ethical predicament, because no credible future raises analogous doubts about other expansions of ethical concern. We don't have similar worries that animals will turn out not to be sentient, for instance.

Virtual and digital futures both require sensitivity to questions about value. But they require different kinds of sensitivity and involve different risks. The best moral outlook for the former is unlikely to be ideal for the latter.

9. Predictable and unpredictable futures.

My final contrastive pair of possible futures concerns the predictability of human behaviour. Imagine a *predictable future* where (perhaps thanks to computer models exploiting big data), future actions can be predicted as reliably as any past action could ever be discovered. Any third-party, and perhaps even the person herself, can be as confident at t2 that A will do X at t3 as she could ever be that A did Y at t1.

This is a prima facie credible future. Even if prediction is in its infancy, it seems premature to rule it out on empirical grounds. Predictable futures raise many fascinating philosophical questions. (Can you thwart predictions about your own future behaviour? Would predictions of uncommitted crimes be admissible in court? Is the public use of predictions consistent with liberal neutrality, when some moral or religious traditions deny the possibility of prediction? And so on.) I hope to address

these questions fully elsewhere. My present concern is whether prediction threatens moral behaviour.

In a world where future actions are always predicted, libertarianism about free will may come to seem incredible. In such a future, moral outlooks that *presuppose* libertarianism may thus fail to motivate people – especially if those outlooks make onerous demands. To avoid moral collapse in predictable futures, we must teach a moral outlook that can survive without libertarianism. (This may be a very difficult task if, as libertarians argue, human beings are naturally disposed to link moral responsibility and libertarian freedom.)

Compatibilists, who believe that freedom is compatible with prediction, will simply reply that the IMO should presuppose compatibilism. As most utilitarians are already compatibilists, this seems the obvious solution. However, there are other possible scenarios where the *rejection* of libertarianism is even more disastrous. Suppose libertarian freedom *is* essential to a worthwhile human life, but also that digital beings (whose behaviour is governed by finite computer programmes) can never be genuinely unpredictable (Boden 2014). It would follow that digital beings cannot enjoy either libertarian freedom or moral responsibility – and therefore that they cannot flourish in any way that is truly valuable. Faced with the option of a digital future, dogmatic compatibilists may not realise how much is at stake.[x] Without resolving perennial philosophical debates about the nature and value of freedom, we cannot decide what moral outlook we should teach.

Our final possible future is the opposite. Consider an *unpredictable future* where technological developments give all agents (whether digital or not) the ability to re-programme their desires, motivations, and decision-procedures, in ways that make it impossible to predict their future behaviour. Time-honoured (albeit fallible) methods of prediction – based on character, past behaviour, statistical generalisations etc. – are all useless. Perhaps people *deliberately* introduce random elements into their thought processes to subvert enhanced prediction technology! Or perhaps extreme libertarians, discovering that human beings have never *previously* enjoyed genuine freedom because our desire-caused behaviour was always in principle predictable, strive to create the first truly free beings. If compatibilism is true, this is a quixotic risk. But if incompatibilism is true, it is essential for the emergence of genuine value.

We cannot rule this future out, especially given our ignorance about the metaphysical nature of freedom; the connections between freedom, responsibility, and human flourishing; and the space of possible future agents.[xi] We must ask what moral outlook would best suit it. For IMOU, the question is whether predictable futures, normal futures, and unpredictable futures all require *the same* moral outlook. As ever, this seems unlikely.

10. An imaginative moral outlook.

My goals in this paper have been to introduce ideal moral outlook utilitarianism, to emphasise the need for utilitarians to think more clearly about the future, and to highlight the difficulties posed by different possible futures. I have argued that we need a flexible moral outlook that promotes well-being across a wide range of possible futures, possible stories about value, and possible metaphysical pictures. I close by sketching how IMOU might deliver the necessary flexibility. (I develop these ideas at greater length in Mulgan forthcoming a.)

Ideally, we want to identify the IMO without first resolving our uncertainty about well-being or metaphysics. My solution relies on one specific claim about moral progress. I assume that we *could* produce a next generation whose moral sensitivity and moral imaginativeness were significantly greater than our own, and whose judgements about value and well-being were much more finely-nuanced than ours. If we teach a moral outlook that emphasises moral imaginativeness, then we can reasonably expect to produce a next generation of (comparative) moral experts.

This hypothetical claim about moral progress should be uncontroversial. Collectively, we surely could enhance moral imaginativeness. After all, we know that experts of all kinds are created by education. And if we *cannot* influence future moral outlooks for the better, then this would be a fatal blow, not merely for rule utilitarianism or IMOU, but for all systematic future ethics.

If we *can* count on the next generation's superior moral judgement, then we can use that judgement to side-step our own uncertainty about value. The trick is to delegate

to them the difficult business of deciding what the ideal code should be maximising in the first place! Instead of teaching the next generation to do X or avoid Y (on the grounds that these rules maximise, say, pleasure), we should encourage them to first develop their moral imaginativeness and then to pursue whatever *they* judge to be most valuable. Because of their superior moral and epistemic position, we can be reasonably confident that a moral outlook that emphasises imaginativeness and judgement will more reliably promote well-being than any similar code that doesn't emphasize those things, even if we can't be sure what well-being is.

Many interesting questions remain unanswered. What is moral imaginativeness? How does it differ from non-moral exercises of imagination? How can we encourage it? What might future people need to imagine? Do different moral outlooks feature different levels of moral imaginativeness? Do different possible futures call for different degrees or kinds of moral imaginativeness? But at least we are now asking the right questions. And IMOU offers us one final piece of advice. If we want to embody the ideal moral outlook in our own lives, then we should concentrate on the places where it differs from our current ethics. So we should start by making imaginative moral experiments of our own and trying to imagine the ethics of the future.[xii]

References.

Agar, Nicholas. 2010. *Humanity's End: Why we should reject radical enhancement.* Cambridge, MA: MIT Press.

Agar, Nicholas. 2014. "On the prudential irrationality of mind uploading." In *Intelligence Unbound: the future of uploaded and machine minds*, edited by Russell Blackford and Damien Broderick, 146-160. Oxford: Wiley-Blackwell.

Bayley, Barrington. 2001. *Soul of the Robot*.

Blackford, Russell, and Broderick, Damien. 2014. *Intelligence Unbound: the future of uploaded and machine minds*. Oxford: Wiley-Blackwell.

Boden, Margaret. 2014. "Creativity and AI: A contradiction in terms?" In *The Philosophy of Creativity: new essays*, edited by E. S. Paul and S. B. Kaufman, 224-244. Oxford: Oxford University Press.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Broome, John. 2009. *Climate Matters*. New York: WW Norton.

Cowen, Tyler, and Derek Parfit. 1992. "Against the social discount rate." In *Justice Between Age Groups and Generations*, edited by P. Laslett and J. Fishkin, 144-161. New Haven: Yale University Press.

Crisp, Roger. 2015. "Well-Being," *The Stanford Encyclopaedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/well-being/>.

Donner, Wendy, and Richard Fumerton. 2009. *Mill*. Oxford: Wiley-Blackwell.

Egan, Greg. 2008a. *Diaspora*. Gollancz.

Egan, Greg. 2008b. *Permutation City*. Gollancz.

Fischer, John, Robert Kane, Derk Pereboom, and Manuel Vargas. 2007. *Four Views on Free Will*. Oxford: Blackwell.

Feldman, Fred. 2011. "What we learn from the experience machine." In *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, edited by R. M. Bader and J. Meadowcroft, 59-86. Cambridge University Press

Fletcher, Guy. 2013. "A Fresh Start for the Objective-List Theory of Well-Being." *Utilitas* 25: 206-220.

Goodwin, Barbara. 1992. *Justice by Lottery*. Chicago: University of Chicago Press.

Gosseries, Axel, and Lucas Meyer, eds. 2009. *Intergenerational Justice*. Oxford: Oxford University Press.

Griffin, James. 1986. *Well-Being*. Oxford: Oxford University Press.

Hauskeller, Michael. 2013. *Better Humans? Understanding the enhancement project*. Durham: Acumen.

Heyd, David. 1992. *Genethics: Moral issues in the Creation of People*. Berkeley: University of California Press.

Hofstadter, Douglas. 2007. *I am a strange loop.* New York: Basic Books.

Hooker, Brad. 1994. "Rule-Consequentialism, Incoherence, Fairness." *Proceedings of the Aristotelian Society* 95: 19-35.

Hooker, Brad. 2000. *Ideal Code, Real World*. Oxford: Oxford University Press, 2000.

Hooker, Brad. 2008. "Variable 'versus' Fixed-Rate Rule-Utilitarianism." *Philosophical Quarterly* 58: 344-352.

Jackson, Frank. 1999. *From metaphysics to ethics*. Oxford: Oxford University Press.

Jamieson, Dale. 2014. *Reason in a Dark Time*. New York: Oxford University Press.

Kahn, Leonard. 2012. "Rule Consequentialism and Scope." *Ethical Theory and Moral Practice* 15: 631-646.

Kaczmarek, Patrick. 2016. "How Much is Rule-Consequentialism Really Willing to Give Up to Save the Future of Humanity?" *Utilitas.* Online Early. doi:10.1017/S0953820816000352

Lockhart, Ted. 2000. *Moral Uncertainty and Its Consequences*. Oxford: Oxford University Press.

MacLeod, Ken. 1996. *The Star Fraction.* London: Orbit.

Miller, Dale. 2000. "Hooker's Use and Abuse of Reflective Equilibrium". In *Morality, Rules and Consequences*, edited by B. Hooker, E. Mason, and D. E. Miller, 156-178. Edinburgh: Edinburgh University Press.

Mulgan, Tim. 2001. *The Demands of Consequentialism*. Oxford: Oxford University Press.

Mulgan, Tim. 2006. *Future People*. Oxford: Oxford University Press.

Mulgan, Tim. 2007. *Understanding Utilitarianism*. Durham: Acumen.

Mulgan, Tim. 2011. *Ethics for a broken world: reimagining philosophy after catastrophe*. Durham: Acumen.

Mulgan, Tim. 2012. "Contractualism for a broken world." Paper presented to workshop on contractualism, Universite de Rennes, May 2012.

Mulgan, Tim. 2014a. "Ethics for Possible Futures." *Proceedings of the Aristotelian Society* 114: 57-73.

Mulgan, Tim. 2014b. "What is Good for the Distant Future? The Challenge of Climate Change for Utilitarianism." In *God, The Good, and Utilitarianism: Perspectives on Peter Singer*, edited by John Perry, 141-159. Cambridge: Cambridge University Press.

Mulgan, Tim. 2014c. "Replies to Critics." *Philosophy and Public Issues 4*, 58-92.

Mulgan, Tim. 2015a. "Mill and the broken world." *Revue Internationale de Philosophie* 69: 205-224.

Mulgan, Tim. 2015b. *Purpose in the Universe: The moral and metaphysical case for Ananthropocentric Purposivism*. Oxford: Oxford University Press.

Mulgan, Tim. 2015c. “Theory and intuition in a broken world.” In *Intuition, theory, and anti-theory*, edited by Sophie-Grace Chappell, 141-166. Oxford: Oxford University Press.
Mulgan, Tim. 2015d. “Utilitarianism for a Broken World.” *Utilitas* 27: 92-114.
Mulgan, Tim. 2016a. “Answering to future people.” *Journal of Applied Philosophy*. Advance Online Publishing. DOI: 10.1111/japp.12222.
Mulgan, Tim. 2016b. “Theorising about Justice for a Broken World.” In Theorizing Justice: crucial insights and future directions, edited by Krushil Watene & Jay Drydyk, 15-32. London: Rowman and Littlefield.
Mulgan, Tim. Forthcoming a. “Moral Imaginativeness, Moral Creativity, and Possible Futures.” In *Creativity and Philosophy*, edited by Berys Gaut and Matthew Kieran, Oxford: Oxford University Press.
Mulgan, Tim. Forthcoming b. “Moral Philosophy, Superintelligence, and the Singularity.” draft manuscript.
Naam, Ramez. 2012. *Nexus*. Axon.
Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Blackwells.
Parfit, Derek. 1984. *Reason and Persons*. Oxford: Oxford University Press.
Portmore, Douglas. 2009. “Rule-Consequentialism and Irrelevant Others.” *Utilitas* 21: 368-376.
Ridge, Michael. 2006. “Introducing Variable-Rate Rule-Utilitarianism.” *Philosophical Quarterly* 56: 242-253.
Searle, John. 1997. *The mystery of consciousness.* London: Granta.
Sepielli, Andrew. 2014. “What to do when you don’t know what to do when you don’t know what to do …” *Nous* 48: 521-544.
Singer, Peter. 1972. “Famine, Affluence and Morality.” *Philosophy and Public Affairs* 1: 229-243.
Singer, Peter. 2011: *Practical Ethics*. 3rd edn. Cambridge: Cambridge University Press.
Skorupski, John. 1989. *John Stuart Mill*. Oxford: Routledge.
Smith, Holly. 2010. “Measuring the Consequences of Rules.” *Utilitas* 22: 413-433.
Stross, Charles. 2005. *Singularity Sky*. Orbit.
Thomson, Judith. 1976. “Killing, Letting Die, and the Trolley Problem.” *The Monist* 59: 204-217.
Walton, Jo. 2015. *The Just City*. London: Corsair.

[i] The argument in this paragraph was inspired by a discussion following a paper I presented to the Philosophy Seminar at the Australian National University in September 2015, and especially to questions from Kim Sterelny and Steve Stich.

[ii] More restrictive definitions of 'teach' or 'next generation' are appropriate for some limited scope variants of IMOU. If 'we' is restricted to some class of moral educators, such as people professionally engaged in moral education, then perhaps our 'teaching' is what we officially do, and 'the next generation' is our current cohort of students. (Similar restrictions might apply if 'we' are people developing public education policies, etc.) However, these restricted IMOUs are not our primary interest in this paper.

[iii] Problematic boundary cases may still emerge. (What about people in the distant future whose moral outlook is formed directly by viewing Michael Sandel's lectures on justice on some future incarnation of YouTube, for instance?) But I propose to set these aside here.

[iv] IMOU has richer resources to deal with partial compliance than any existing rule utilitarianism, because it does not idealize to full compliance *among the next generation* at all. We idealize *our teaching*, not *their response* to that teaching – and certainly not their subsequent behaviour. The IMO is chosen because of how it would *actually* be implemented by real human beings. This prevents it from becoming too idealistic. Anyone who has internalized the IMO will therefore be well-equipped to respond to partial compliance, especially partial compliance with that outlook's more demanding elements.

[v] I have found no sustained discussion of rule utilitarianism and the future in the literature. A search on The Philosophers' Index in March 2016 for works published in English between 1990 and 2015 that include both 'rule' and 'future' in either title or abstract yielded 128 results. Only two of these dealt with rule utilitarianism – and those were my *Future People* book and a review of that book. (One exception, which I discovered only after I had finished this paper, is Kaczmarek 2016.)

[vi] Post-scarcity technology, if concentrated in a few hands, might enable the few to easily dominate the many. If I own the *only* cornucopia machine (or I alone have the information that makes the machine produce the right *things*), then I might reasonably hope to impose my minority preferences indefinitely.

[vii] Another potential resource for collective utilitarians is the extensive literature on moral uncertainty (e.g., Lockhart 2000; Sepielli 2014). However, that literature is largely orthogonal to the concerns of this paper, because it focuses on uncertainty regarding competing accounts of right action, rather than uncertainty about the correct theory of well-being or value.

[viii] My thinking about digital futures is especially indebted to Agar 2010 and Agar 2014. For general philosophical discussion, see Blackford and Broderick 2014; Bostrom 2014; and Hauskeller 2013, 115-132. Some philosophically rich fictional presentations are Bayley 2001; Egan 2008a, 2008b; MacLeod 1996; Naam 2012 and Walton 2015.

[ix] Not all objectivists will find value in the unconscious digital future. Some will agree with hedonists that it is a valueless void. If we endorse an *experience requirement* (Griffin 1986, 13; Fletcher 2013, 210), then other list items (achievement, knowledge, friendship, preference satisfaction) only add value to a person's life if she *experiences* them. In a world without experience, there can be no valuable lives. There can thus be value in the unconscious digital future only if *some* items on our list are *not* subject to an experience requirement. Of course, if pleasure has *any* value, then worlds with flesh-and-blood humans will still retain an advantage. But unconscious digital futures are now worth *something*. And unconscious digital beings can then outweigh conscious competitors – *if* they can collectively accumulate *enough* valuable achievements to compensate for their lack of awareness. Given the potentially enormous achievements opened up by the digital transition, this is not out of the question.

[x] Predictability and freedom might also be closely linked to consciousness. Suppose consciousness must include awareness of one's own freedom. Predictable beings are thus never truly conscious. If digital beings are predictable, then every digital future is both predictable and unconscious.

[xi] For a taste of the current debates about free will and moral responsibility, in particular, see e.g. Fischer et al. 2007.

[xii] The first version of this paper was written for presentation as a Paduano seminar at the Stern Business School at New York University in April 2015. I am very grateful to Rex Mixon and Bruce Buchanan for the invitation and for their very generous hospitality. Subsequent versions were presented at the Australasian Association for

Applied and Professional Ethics conference in Auckland in July 2015, the Philosophy Department seminar at the Australian National University in September 2015, and a workshop on collective agency at St Andrews in November 2015. I am very grateful to Tim Dare, Seth Lazar, and Samuel Mansell for these later invitations, and to audiences at all four events for very helpful comments.