

Axiological Futurism

The Systematic Study of the Future of Human Values

John Danaher,

Senior Lecturer

NUI Galway

john.danaher@nuigalway.ie

Abstract: Human values seem to vary across time and space. What implications does this have for the future of human value? Will our human and (perhaps) post-human offspring have very different values from our own? Can we study the future of human values in an insightful and systematic way? This article makes three contributions to the debate about the future of human values. First, it argues that the systematic study of future values is both necessary in and of itself and an important complement to other future-oriented inquiries. Second, it sets out a methodology and a set of methods for undertaking this study. Third, it gives a practical illustration of what this ‘axiological futurism’ might look like by developing a model of the axiological possibility space that humans are likely to navigate over the coming decades.

1. Introduction

Axiological change is a constant feature of human history. When we look back to the moral values of our ancestors, we cannot help but notice that they differed from our own. Our grandparents, for example, are quite likely to have harboured moral beliefs that would count as prejudiced and bigoted by modern standards; and we are quite likely to harbour moral beliefs that they would find abhorrent. As we go further back in time, the changes become even more pronounced (Pleasant 2018; Appiah 2010; Buchanan 2004; Pinker 2011).

Axiological variation is also something we see today when we look around the world and take note of the different cultures and societies that care about and prioritise different values (Flanagan 2017). What consequences does this axiological change and variation have for the future? Should we plan for and anticipate axiological change? Can we study the future axiological possibilities of human civilisation in a systematic and insightful way?

This article tries to answer these questions in three stages. First, it makes the case for a systematic inquiry into the future of human values — termed 'axiological futurism' — and argues that this inquiry is both desirable in its own right and complementary to other futurological inquiries. Second, it outlines a methodology for conducting this inquiry into the future of human values. And third, it presents a sketch of what this inquiry might look like by proposing a tentative model of the 'axiological possibility space' that we are likely to navigate over the coming decades.

In other words, this article explains *why* axiological futurism is needed; *how* we might go about doing it; and *what* it might look like if we did.

2. Making the Case for Axiological Futurism

Broadly construed, axiological futurism is the inquiry into how human values could change in the future. Axiological futurism can be undertaken from a normative or descriptive/predictive perspective. We can inquire into how human values *should* change in the future (the normative inquiry) or we can inquire into how human values *will* (or are *likely to*) change in the future (the descriptive/predictive inquiry).

Axiological futurism is both desirable in and of itself, and complementary to other futurological inquiries. As noted in the introduction, we know that the values people subscribe to have changed across time and space. This means they are likely to change again in the future. This is true even if you think that there is a timeless and unchanging set of values (i.e. an eternal and universal set of moral truths) that is not susceptible to change. Why so? Because even if you accept this you would still have to acknowledge that people have

changed in both their awareness of and attitude towards those timeless and unchanging values over time. Perhaps this is because we are getting closer (or further away) from the eternal moral truth. Either way, our values are always changing and if we want to understand the future we have to factor this change into our accounts.

To illustrate, consider some examples of historical axiological change. One clearcut example is the changing attitude toward the moral status of different groups of people. For a very long time, most societies adopted the view that some adult human beings (e.g. slaves and women) were morally inferior to others (adult, property-owning men). This view may always have been contested to some extent (Pleasants 2018, 571; Aristotle *Politics* 1253b20-23), but it was the received moral wisdom for the majority and was reflected in their daily beliefs and practices. This changed over the course of the 19th and 20th centuries, and although the old moral attitudes linger in certain contexts and places, the shift has been quite dramatic (Appiah 2010). Something similar is true for attitudes toward practices like gruesome torture and wanton animal cruelty (Pinker 2011).

There are other clearcut examples of moral variation if we look across cultures. Owen Flanagan points this out by comparing Buddhist moral communities and Western liberal moral communities (Flanagan 2017). In doing so, he highlights how members of those respective communities have very different attitudes towards the value of the individual self and the emotion of righteous anger. Buddhist communities usually deny or downplay the value of both; Western liberal communities embrace them.

Given the facts of value change, it is prudent to anticipate and plan for future changes. The current moral paradigm is unlikely to remain constant over the long term and it would be

nice if we know how it is likely to change. This is true even if we approach axiological futurism from a normative perspective as opposed to a purely descriptive one. Normatively speaking, there is no guarantee that our current moral paradigm is the correct one and so we might like to see where future moral progress lies and try to get ahead of the curve (Williams 2015). Indeed, this kind future-oriented moral reasoning already features in some normative decision-making. For example, in certain constitutional law cases in the US — which oftentimes engage abstract moral values like justice, fairness and equality (Leiter 2015) — judges have reasoned their way to particular conclusions out of a desire to be on the ‘right side of history’ (McClain 2018). Conversely, even if you are convinced that the current moral paradigm is the correct one, you should still care about the ways in which it might change in the future, if only because you want to protect against those changes.

Axiological futurism is also complementary to most other futurological inquiries. Most futuristic studies are value driven, even if only implicitly. People want to know what the future holds because they value certain things and they want to know how those things will fare over time. If we had no value-attachments, we probably wouldn’t care so much about the future (if nothing matters then it doesn’t matter in the future either). Value attachments are common in the debates about existential risks (Bostrom 2013; Torres 2017). Take, for example, the inquiry into climate change. Much of the debate is driven by an attachment to a certain set of values. Some people worry about climate change because it threatens the capitalistic conveniences and consumer lifestyles they currently value. Others are more concerned about central or basic values, such as the value of ongoing human life, and worry that this is a put at risk by climate change. Something similar is true in the case of AI-risk. People who worry about the impact of superintelligent AI on human civilisation are deeply concerned about the preservation of human flourishing. They worry at that the space of

possible artificial minds is vast, and that the sub-set of those minds that will be ‘friendly’ to human flourishing is quite narrow (Armstrong 2014; Bostrom 2014; Yudkowsky 2011).

These risk-oriented futurological inquiries are either implicitly or explicitly value driven: they are motivated by an attachment to certain human values and by the worry that socio-technical changes will threaten those values. It is interesting then that these futurological inquiries often assume a relatively fixed or static conception of what the future of human values might be (indeed oftentimes a quite anthropocentric and culturally specific set of values). In other words, they assume that there will be great technological change in the future, but not necessarily great value change. Or, even if they do anticipate some value change, it is relatively minimal or narrow in nature. There is, consequently, a danger that these inquiries suffer from an impoverished axiological imagination: they don’t see the full range of possibly valuable futures. There are some exceptions to this (notably Bostrom 2005; and, in part, Baum et al 2019) but even those exceptions would benefit from a more thorough and systematic inquiry into the space of possibly valuable futures (Van De Poel 2018). Such an inquiry might encourage greater optimism about the future — by showing that the space of valuable futures is not quite so narrow or fragile as some suppose — or alternatively encourage realistic pessimism by showing how narrow and fragile it is.

One could object to axiological futurism on the grounds that it is impossible to say anything meaningful or predictive about the axiological future. Any attempt to do so will be hopelessly speculative and will more than likely get things wrong. This is, of course, a criticism that could be thrown at all futurological inquiries. Futurists are notorious for getting the future wrong and looking somewhat foolish as a result. This doesn’t mean the criticism is unwarranted, it just means that it is not unique to axiological futurism. The best response to

this criticism is to argue that the point of axiological futurism is not to precisely predict the future of value. The point of axiological futurism is to map out the broad space of possible axiological trajectories that we could take in the future; to anticipate and imagine the different scenarios; and to help us to plan for those possibilities. The goal is not to give a precise axiological forecast; it is to engage in axiological scenario planning (cf. Baum et al 2019 who do something similar from a non-axiological perspective). I hope to show, rather than tell, how this might be done in the next two sections of this article, drawing lessons from past moral revolutions for guidance (Pleasants 2018; Appiah 2010).

One could also object to axiological futurism on the grounds that it is, in some sense, conceptually impossible. We are all trapped inside a particular moral paradigm.¹ These paradigms shape how we perceive and understand moral value. We cannot get outside these paradigms and imagine other axiological possibilities. The obvious response to this is that we cannot know whether this is true until we have tried. Furthermore, this criticism is more of a concern for the normative version of axiological futurism than the descriptive version. It may be true that our moral perceptions and emotions are so tied to a particular paradigm that we cannot feel any moral attachment to a different one, but we can at least try to describe and understand what it might be like to inhabit a different paradigm (Pleasants 2018). Historians and anthropologists do this all the time — they become tourists to different worldviews, both historical and cross-cultural. As Thomas Kuhn once argued, a contemporary scientific historian might not believe in geocentrism or the existence of phlogistan, but they can at least try to figure out what it might have been like to believe in those theories during the relevant historical era (Kuhn 1962). The axiological futurist can do the same: they can become tourists to new axiological paradigms. It may even be possible for axiological futurists to genuinely

¹ As MM pointed out to me, this could also be a problem with any moral philosophical inquiry that tries to get outside a current moral paradigm. Consequently, since moral inquiry takes place all the time, we might have a reduction of the critique.

feel a moral attachment to new paradigms by taking their attachment to current values to their logical extremes, e.g. by imagining what it might be like to care about all sentient life equally, or to care for robots in the same way that they care for human beings.

One could also take issue with axiological futurism on the grounds that it is nothing new. People have been doing it for decades, albeit without the fancy title. For example, people who argue that we are transitioning into a ‘post-privacy’, ‘transparent’ society as a result of technological change are doing axiological futurism (Brin 1998; Peppet 2015). Similarly, someone like Yuval Noah Harari, in his ‘future history’ books, is doing axiological futurism when he imagines a future ‘religion’ of ‘dataism’ in which data is valued above all else and individualism and humanism are forgotten (Harari 2016). More recently, Ibo Van De Poel (2018) has even called for the design of technologies to be sensitive to the possibility of value change. I do not deny this nor claim that the project envisaged in this article is wholly original. Of course, people have been imagining the future of value for quite some time. What is distinctive about axiological futurism is that it calls for a systematic and explicit examination of the future of value. Rather than focusing on one specific way in which technology might change our values, or on one pet theory of future value, it proposes a fullscale, systematic exploration of future axiological possibility spaces.

Finally, one could object to axiological futurism on the grounds that it will be self-fulfilling or self-defeating. This might be a particular problem if axiological futurism is pursued from a normative perspective. Imagined future axiologies can be enticing or intimidating. For example, those who like the idea of a post-privacy society can use the idea to argue for changes to current social and legal norms; those who hate it can lobby against any such changes. The result is that the imagined axiology either comes into being because

people want it to or never gets off the ground because people don't. But this is, of course, a problem for all futurological inquiries. Since human activity is one of the things that will determine what kind of future we have, there is always the danger that an imagined future compels action in a particular direction (Popper 1957). This seems unavoidable to some extent and yet still not a reason to avoid inquiry into the future. Indeed, one potential benefit of axiological futurism is that it could encourage a less knee-jerk and emotional response to the future. By expanding our axiological horizons we might see less reason to jump to conclusions about how desirable or undesirable the future might be.

3. The Methodology of Axiological Futurism

How could we actually go about doing axiological futurism? What's the methodology? In this section I will sketch an answer to that question. I emphasise, at the outset, that this is just a sketch — something that that other people can refine and improve upon.

It helps if we start with a more precise definition of axiological futurism. I defined it informally in the preceding section. A more formal definition is now required:

Axiological Futurism: The systematic and explicit inquiry into the axiological possibility space for future human (and post-human) civilisations, with a specific focus on how civilisation might shift or change location within that axiological possibility space.

We'll unpack the elements of this definition as we go along. From a methodological perspective two crucial things emerge from it (a) that axiological futurists must provide some theory of the 'axiological possibility space' and (b) that axiological futurists must identify the

methods that can be used to explore that possibility space and the potential trajectories within it.

(a) What is axiological possibility space?

Let's start with the theoretical aspect of the methodology: the idea of *axiological possibility space*. Our goal, as axiological futurists, is to explore this space, to figure out the ways in which it might vary and change in the future, and to identify some possible trajectories that human civilisation might take through this possibility space. To fully understand this idea we need to get into some of the basics of moral theory and axiology. This will help us to determine what the boundaries of the possibility space might be.

Moral theories are usually made up of two main elements: (i) a theory of what is good/bad (i.e. an *axiology*) and (ii) a theory of what is right/wrong (i.e. a *deontology*). Moral theories are then usually directed at two kinds of entities (iii) moral patients/subjects (i.e. those who can be benefitted/harmed by what is good/bad) and (iv) moral agents (i.e. those who can perform actions that are right/wrong). Many times these different elements coincide in a single moral theory. For example, most adult human beings are viewed as moral patients and moral agents and hence are deemed eligible subjects for both an axiology (i.e. there are things that are good for them) and a deontology (i.e. there are rules about what they ought to do). Nevertheless, sometimes the elements can pull apart. For example, most young children are thought to count as moral patients, but not moral agents. They can be benefitted and harmed, but they cannot perform morally right or wrong actions. It may also be possible, under certain moral theories, for things to be good or bad *simpliciter* (in and of themselves) without them being good or bad for some moral subject. Certain theories of environmental ethics, for example, claim that features of the natural world are intrinsically valuable without

being valuable for someone. That said, for the most part, an axiology goes hand in hand with a theory about who has moral patiency and a deontology goes hand in hand with a theory about who counts as a moral agent.

If our goal was to explore the entirety of moral possibility space, then we would have to concern ourselves with all four of these elements (axiology; deontology; patients; and agents). But I am suggesting that we concern ourselves primarily with the axiological elements. Why is this? Because I make an assumption — which could be challenged — that axiology ultimately drives and determines deontology. In other words, I maintain that you need to know what is good/bad (and who can be harmed/benefitted) before you can figure out what to do about it. If you know that animals can be harmed and benefitted, then you know that our actions toward them have a moral dimension. But if you don't know that, or if you don't accept that, you won't think of your actions toward them having a moral dimension, at least not directly (they may have a moral dimension because of their consequences for other moral patients such as your fellow humans or yourself). Your axiological beliefs about animals ultimately shape your deontological beliefs. This doesn't mean they shape them in a straightforward or simple way, but they do constrain how you can think about the morality of the actions you perform towards animals. This is why I think we should focus primarily on axiology. In doing this, we may well generate some conclusions or hypotheses about future deontologies — it would be surprising if we did not given the relationship between axiology and deontology — but this is not the primary object of inquiry.

An 'axiology' will consist of three main things (i) some list or specification of what is good/bad; (ii) some identification or specification of who counts as an object of moral concern (i.e. who the moral patients are) and (iii) some specification of the *relationships*

between the various elements within the axiology (what's most important/least important? what is intrinsically valuable and what is instrumentally valuable?). Axiological possibility space, then, is the full set of possible axiologies, i.e. all the different possible combinations of goods, subjects and relationships between them. Presumably, axiological possibility space is vast — much larger than anyone can really imagine. But equally presumably, many of the 'possible' axiologies within this space are not that plausible or interesting: e.g. a world in which the subjective pleasure we experience while scratching our knees is the only recognised good may be *possible* (in some thin sense of the word 'possible') but is not very plausible and should not concern us greatly.

Still, the vastness of axiological possibility space poses a challenge for the axiological futurist. We need some constraints on the boundaries of axiological possibility space to make the project feasible. Fortunately, we can constrain the axiological possibility space to some degree by taking advantage of the work that has already been done in defining axiologies, and by considering some of the ongoing debates within axiological theory. Doing so, we can quickly gain some sense of the kinds of things that could be included in any possible list of goods/bads. They would include (as 'goods') things like: subjective pleasure, desire satisfaction, knowledge, friendship, beauty, education, health, money, family, food and so on. We also know the kinds of entities that could count as moral subjects. They would include: humans, cognitively complex mammals, all 'persons' (human, animal or artificial), all sentient beings, all living entities, and possibly some non-living entities of great beauty or ecological significance. Finally, we also know the different possible relationships that could exist between these goods/bads and the moral subjects: all could be treated equally, there could be a clear hierarchy of goods and moral subjects, there could be a cyclical or rotating ranking of goods and moral subjects, some goods could be valued intrinsically and some

instrumentally, or there could be multiple different combinations of these relationships. So we know, roughly, the broad constraints on axiological possibility space. It is still vast, and it will be a challenge to explore it, but this what makes axiological futurism an important and intellectually fascinating endeavour.

Of course, it is not enough to identify the boundaries of axiological possibility space. We also have to think about how changes can come about within that possibility space. This is key to the ‘futuristic’ aspect of axiological futurism. Axiological futurism is not conceived as a purely abstract, intellectual exercise in which we map out all the plausible, possible axiologies that could be taken on by human (and post-human) civilisations. We also want to know something about the mechanics of axiological change. How do changes come about? Are some changes inevitable or irreversible? How are things likely to change in the future, particularly in response to technological change? This doesn’t require falling into the trap of precise prediction; but it does mean thinking carefully and systematically about how axiologies can vary over time and space.

This might seem like a daunting task, but we know that the mechanics of axiological change are relatively simple. There are three things that can happen to change an axiology: (i) there can be some expansion or contraction of the circle of moral concern (i.e. the set of beings who count as moral subjects); (ii) there can be some addition to or subtraction from the set of goods/bads; and (iii) there can be some change in how we prioritise or rank goods/bads and/or moral subjects (cf Van De Poel 2018). We see clear evidence for all three kinds of change in human history. The end of slavery and the enfranchisement of women can be interpreted as either an expansion of the circle of moral concern or a change in how we rank moral subjects (towards greater equality). Contrariwise, the rise of fascism and

nationalism can be interpreted as a contraction of the moral circle or a demotion in the moral ranking of certain subjects. Likewise, the loosening grip of religion over the moral lives of industrial society has often brought with it changes to the list of goods/bads (e.g. premarital sex or uncontracepted sex is now no longer seen as 'bad'). Examples could be multiplied.

There is considerable debate in moral philosophy as to whether the changes we see over history are, broadly, progressive, or whether certain kinds of axiological change can be reliably identified as 'progressive' (Moody-Adams 1999; Stokes 2017). One apparent lesson from history is that expansions in the circle of moral concern are usually considered progressive, and viewed positively in the light of history, but there is no guarantee that this trend will continue. For example, some people emphatically reject the idea that expanding the circle of moral concern to include artificial entities would be progressive (Bryson 2018), while others are more open to the idea (Gunkel 2018). There is no need to enter into these debates if axiological futurism is pursued from a descriptive/prescriptive stance since the goal is not to get axiological changes right but rather to understand how and why they happen. There may, however, be a need to get into these debates if axiological futurism is pursued from a normative perspective. In that case, we want to be able say whether or not the direction of axiological change is positive or negative.

This is to talk about what must happen to change an axiology. What actually drives those changes? Ultimately, all moral change is cashed out at the individual and institutional levels: people change in their axiological beliefs, practices and attitudes; and institutions espouse and promote those changes. But how do individuals and institutions change? Broadly speaking, there are two main drivers of change: intellectual drivers and material drivers.

These correspond to the drivers of change that are widely discussed in other sociological and ideological debates (e.g. Marxism vs Hegelianism).

Intellectual drivers of change arise from the application of fresh ideas, theories and reasons to an axiology. Sometimes they arise from within an axiology. Applied moral philosophy is of this kind. Applied ethicists spend much of their time identifying problem cases in moral theory and explaining why moral beliefs, attitudes and practices must change in response to these cases (Campbell 2014). But there are also non-moral intellectual drivers of change. For example, non-moral and non-rational methods of persuasion or example-setting are sometimes key to moral reform (Moody-Adams 1999; Pleasants 2018; Appiah 2010). People change because an attractive person espouses or exemplifies change; or because they are made to fear staying the same; or for some other non-moral reason (cf Fernandez-Armesto 2018 on the influence of ideas on values over time).

Material drivers of change are, obviously, different. They are changes in the material conditions of existence which bubble up into changes in axiological beliefs and practices. New technologies, for example, often make new actions possible (Currier 2015). They can change how we prioritise and rank goods/bads and moral subjects. For example, it is probably no accident that slavery was legally abolished *after* the industrial revolution got going: advances in mechanisation obviated some of the need for slave labour that became important after the shift to agriculture. Ian Morris has developed an extensive historical theory of why this happened, focusing in particular on how changes in technologies of energy capture changed values (Morris 2015). On a smaller scale, the technology ethicist Shannon Vallor has argued that certain technologies bring with them their own set of 'values-in-use', which can have a dramatic impact on our overall axiology (Vallor 2016). For example, she argues that

advances in the global reach of technology, particularly communications technology, means that our circle of moral concern must now be global and not local (Vallor 2016, chapter 2). In the next section I will outline another approach to understanding how technology might drive axiological change. Nevertheless, we should be sceptical as to whether axiological change is ever completely intellectual or material. It is more likely that there is a complex feedback loop between both drivers of change. Intellectual factors might drive the creation of new technologies, which in turn affect how we behave and what we perceive as valuable; or changes in technology might inspire our imaginations to consider new axiological possibilities. We do not need to be doctrinaire materialists or idealists to be axiological futurists. We can be a bit of both.

In sum, the methodological goal of axiological futurism is to inquire into the axiological possibility space for future human and post-human civilisations. The job of the axiological futurist is to sketch different possible axiologies, and anticipate the future trajectories we might take through the axiological possibility space.

Elements of an Axiology	Axiological Change	Drivers of Change
(i) Set of values (goods/bads), i.e. what do people care about and promote	(i) Adding to or subtracting from the list of values	(i) Intellectual drivers of change, i.e. changes to how people think about their axiologies
(ii) Moral subjects, i.e. who or what is worthy of moral consideration	(ii) Expansion or contraction of the circle of moral concern (i.e. the set of moral patients/subjects)	Inconsistencies and contradictions within the axiology
(iii) Relations between values and subjects, i.e. who or what is most important? Do the things and subjects that are valued matter intrinsically or instrumentally?	(iii) Reprioritisations of re-rankings of the values and subjects	Rational reasoning, teasing out the implications or consequences of axiological belief Non-moral, non-rational persuasion
		(ii) Material drivers of change, i.e. changes to the material circumstances of life Axiologies that are internal to particular technologies (e.g. Shannon Vallor <i>Technology and</i>

		<i>the Virtues</i> Axiologies that are driven by or necessitated by external technological change (e.g. Ian Morris <i>Foragers, Farmers and Fossil Fuels</i>)
--	--	---

Table 1 – Understanding Axiological Possibility Space

(b) What methods can we employ to explore the axiological space?

Now that we have a clearer sense of the task of axiological futurism, we can consider the methods that might assist in performing this task. The methods are going to be a grab-bag. Axiological futurism is an inherently speculative and imaginative exercise. We cannot experimentally control, manage and predict the future. Karl Popper’s classic arguments against predictive social science apply well to axiological futurism (Popper 1957): human history is, ultimately, a single unique event, you cannot easily account for the human factor in the evolution of societies (particularly since humans discover, react and respond to predictions made about their futures), and while you may be able to eliminate some possible futures from consideration you can never really narrow it down to one, predictable, future trajectory for human civilisation. Consequently, axiological futurism cannot be a ‘science’ in the strict sense. It is an exercise in *informed speculation*. Anything that helps to inform that speculation is a viable method of doing axiological futurism.

What follows are some suggested methods of inquiry. Many of these methods are already being employed by researchers in philosophy, psychology and social sciences. They are just not being employed specifically in the service of axiological futurism. What I suggest

here is that they can be repurposed and reconceived for that end. This is a tentative list. The hope is that it will be added to in the future.²

Logical Space Methods: The first task of the axiological futurist is to map out the contours of axiological possibility space. One obvious way to do this is to map out the logical space of variation for a given value or set of values. The resulting logical space will help us to identify the different ways in which a value might be specified and how it might relate to other values. This is something that is already done by moral philosophers with respect to individual values and pairs of values (e.g. Roemer 1998; List and Valentini 2016). For example, a lot of work has been done on the logical contours of values such as ‘equality’ and ‘freedom’. Philosophers have identified dimensions or parameters along which different conceptions of these values can vary. A theory of equality for example might vary along two dimensions: equality of opportunity and equality of outcome. Given these two parameters, a researcher can construct a simple 2 x 2 logical space for the value of equality, classifying different possible axiologies depending on whether they score high or low on those two dimensions. More complex variations are also possible.

Constructing logical spaces is usually just a matter of carefully reading the theoretical literature and spotting the patterns and variations among the different theories associated with different values. Christian List and Laura Valentini (2016) adopt this approach when trying to understand the value of political freedom. They note that theories of political freedom tend to be concerned with interferences with individual behaviour, but then vary depending on the kinds of interferences with which they are concerned. Some theories are concerned only with *actual interferences* with individual behaviour (‘freedom as non-interference’ theories); some

² For a complementary list of methods, focused specifically on AI futurism, see Shahar 2019

theories are concerned with possible interferences with individual behaviour ('freedom as non-domination' theories). Similarly, some theories are only concerned with immoral (unjustified) interferences, whereas some theories are concerned with all possible interferences, be they moral or immoral. This suggests to List and Valentini that theories of freedom vary along *modal* and *moral* dimensions and they then use this to construct a 2 x 2 logical space of freedom.

Both of these examples involve logical spaces for individual values. It is also possible to use this method to construct logical spaces to represent the different ways of valuing moral subjects and, crucially, for mapping the possible relationships between different values. For example, you could imagine a simple axiology in which there are three main values: equality, freedom and well-being. Each of these values represents a dimension of variance for a possible society. We can then define a three-dimensional axiological space within which possible societies can be classified and organised. Some societies may value all three highly and try to maximise all three; some will value well-being over freedom and freedom over equality (and so on). It may also be the case that certain axiologies that seem to be possible within this space are not in fact possible. For example, it may not be possible to maximise freedom and equality (or equality and well-being) at the same time. Some tradeoffs and compromises may be (logically/physically/technologically) necessary (e.g. Kleinberg, J., Mullainathan, S. and Raghavan, M. 2016 on the impossibility of reconciling different conceptions of fairness). Historical, cross-cultural and psychological inquiries could be an important guide in this respect as they will give us a sense of what has been and is currently possible for human societies when it comes to different combinations of values. For example, moral foundations theory, which is a psychological theory suggesting that there are five (maybe six) basic dimensions of value in human moral psychology, might set an important

limit on what is possible when it comes to human axiology (Graham et al 2013; Haidt 2012; Flanagan 2017). Whatever the case may be, thinking about the relationships between values in terms of logical spaces allows for a more systematic and thoughtful inquiry into axiological possibility space.

There are, however, limits to what individual humans can do when it comes to constructing and exploring logical spaces. Humans can handle two to three dimensions of variance with relative ease. Beyond that, it gets much trickier to intuitively conceive of a logical space of possibility. Formal and computer-assisted methods of mapping logical spaces may consequently be necessary to make the vast space of axiological possibility more tractable and manageable.

Causal Relationship Methods: Mapping the logical space is a first step. Ultimately, what we want to know are the causal relationships between potential drivers of change in axiology and actual changes in axiology. Figuring out those causal relationships is crucial if we are to make the space of possible future axiologies comprehensible. Working out these causal relationships will be tricky. Again, we cannot run civilisation-wide experiments on possible future axiologies, particularly if axiologies are partly determined by forms of knowledge and technology that are yet to be discovered and invented. To make headway on this, we have to rely on historical studies of axiological change, cross-cultural studies of axiological variance and psychological and small-scale experimental studies of change and variance. Each kind of study gives us a different insight into the possible causal relationships at play.

Historical studies, particularly if they are grand in scope and scale, are helpful because they give us a sense of how axiologies have changed in response to (and in conjunction with) other social and technological changes. William Lecky's *The History of European Morals* (1959) is a classic example of this style of inquiry, being one of the first studies to consider how material factors drove changes in European axiologies. But intellectual studies of axiological change should not be neglected. They show how evolving conceptions and ideologies can drive changes in axiological belief systems. JB Schneewind's study on *The Invention of Autonomy* (1998) is a good example of this style of inquiry. It provides a detailed map of the intellectual debates that gave rise to the modern liberal axiology, with its focus on the autonomous individual as the ultimate locus of value. Similarly, Kwame Anthony Appiah's study of three historical moral revolutions makes the case for thinking that changing conceptions of honor played a key role in moral change (Appiah 2010). Deeper historical studies are useful too. Michael Tomasello's examination of the 'natural history' of morality, for example, gives a sense of the evolutionary steps that had to be taken for human moral systems to arise (Tomasello 2016). Likewise, Ian Morris's aforementioned study of how changes in the technology of energy capture drove changes in axiologies of violence, equality and fairness gives a sense of the major socio-technical forces that might be at play in axiological change (Morris 2015). This is just a small sample of the historical inquiries that can assist axiological futurism. Examples could be multiplied and all are somewhat useful in helping us to tease out the potential causal mechanisms behind axiological change and thereby extrapolate from the past to the future.

Cross-cultural studies of axiological variance are helpful because they give us a chance to learn from 'natural experiments' in axiological possibility space. Different axiological communities arise in different geographical locations, and in different socio-technical

contexts. Comparing and contrasting the axiological variance across communities can give us a sense of both (a) the causal factors that might be responsible for this variance and (b) how broad the axiological possibility space really is (and thereby help us to overcome the parochialism and short-sightedness that often comes with being locked in one axiological worldview). Owen Flanagan's book *The Geography of Morals* (2017) is a good manifesto for this kind of inquiry, demonstrating how anthropological and psychological research can support this cross cultural analysis, and providing some detailed normative evaluations of the axiological variations between Western and Buddhist societies. There are, of course, many other ethnographic and anthropological studies that can assist in studying axiological variance. Cross-cultural comparison can be particularly fruitful from a futuristic point of view if some communities are further along in their socio-technical development than others. As William Gibson once famously observed, the future is already here, it is just unevenly distributed. Axiological futurists can take advantage of that uneven distribution to further their aims. Indeed, there are some examples of this kind of inquiry already taking place. For example, Jennifer Robertson's *Robo Sapiens Japonicus* (2017) gives insights into the axiological beliefs and practices of Japanese society with respect to robots. This is interesting because Japan is a society in which robots are generally more accepted and more widely used than in most Western societies. Similarly, Virginia Eubanks's *Automating Inequality* (2017) takes explicit inspiration from Gibson and argues that poorer communities are subject to much greater algorithmic surveillance and governance than wealthier communities, and so give us some insight into how axiological beliefs and practices might change if and when algorithmic governance technologies become more widely distributed.

Finally, psychological and other experimental studies are helpful because they can provide insight into the causal mechanisms that might underlie larger scale axiological

change. Studies in moral psychology on the foundations of moral belief and practice (Graham et al 2013; Haidt 2012; Greene 2013) and the mechanisms of moral change are obviously of great relevance. They help the axiological futurist identify possible upper limits on the manipulability of axiological systems in response to intellectual and material drivers of change. Studies that focus in particular on how moral beliefs and practices might change in response to new technologies are also of particular relevance to the axiological futurist. The ‘Moral Machine’ experiment, run by researchers based in MIT could be an example of the genre, although not conceived by its authors in these terms. This experiment was a largescale examination of how axiological (and deontological) belief systems might respond to autonomous driving technology (Awad et al 2018), specifically how people would reason about dilemmas involving sacrificing different groups of people. The experiment helped to reveal different biases in the ranking of moral subjects across different cultures and in doing so gives us some sense of the contours of axiological possibility space with respect to one particular use of technology.

As with the mapping of axiological possibility space, some assistance from formal computer-assisted modelling could be a useful complement to these experimental approaches. For example, computer models of repeated games (like the Stag Hunt or the Prisoners’ Dilemma) can give some insight into the causal factors that might be responsible for changing social norms over time (Alexander 2008; Bicchieri 2016; Skyrms 1997).

There will always be limits to the informativeness of these experimental approaches. They will usually involve small groups of experimental subjects or simplified model environments. Most studies will only model a handful of axiological changes and causal factors. Even the MIT Moral Machine experiment — which was impressive for the fact that it had millions of experimental participants — was limited insofar as it only focused on one

type of technological change and one set of moral beliefs (specifically moral responses to so-called ‘trolley dilemmas’). Limitations of this sort are inevitable and they necessitate caution when it comes to extrapolating from these experiments to society-wide axiological changes. Still, this shouldn’t negate the great importance of these studies to the axiological futurist.

Collective Intelligence Methods: There is one final class of methods that is worth discussing. As we have seen so far, axiological futurism is an exercise in informed speculation in which the theorist tries to (a) map the contours of axiological possibility space, (b) determine the causal relationships between drivers of change and resulting changes of location within that axiological space, and (c) use this to speculate about the possible future trajectories that human and post-human civilisations will take through axiological possibility space. The axiological futurist draws upon different disciplines and methods to assist in these three tasks, including anthropology, history, psychology and moral philosophy.

As noted earlier, there are already some people conducting inquiries that could be classed as a type of ‘axiological futurism’. What is notable, to date, is that most of these inquiries are the product of individual authors who do not conceive of their projects in the terms outlined in this article. This kind of solo-authored inquiry has advantages — it’s relatively easy to do and if pursued to the hilt, and sketched in full imaginative depth, it can be quite visceral and effective (e.g. Hanson 2016). I mentioned some academic examples earlier on but it is true in the case of fiction too. For example, many science fiction novels and stories have a strong axiological futurist aspect to them. They often depict dystopian futures in which humanity has taken a wrong turn in axiological possibility space. They warn us against doing the same. Good examples of this would include George Orwell’s *1984* and Dave Eggers *The Circle* both of which sketch out possible futures in which there is near

perfect government and corporate surveillance. Both novels give a sense of what the resultant social axiologies might be (a fixation on transparency and conformity), and neither paints a flattering picture. Utopian science fiction can do the same thing, albeit with the opposite purpose in mind.

The problem with solo-authored work of this sort is that it is often narrow-minded and biased. Individual authors have their own axes to grind. They focus on one or two technological or social changes and consider one or two consequences for our axiologies. They don't think about multiple changes in parallel nor the possibility of multiple different future axiologies. For example, they might focus solely on changes to surveillance technology, and imagine what might happen if that technology gets really good, while at the same time ignoring similar changes to other technologies such as genetic engineering, cyborgification, robotics, space exploration and so on. Better work will try to consider multiple streams of change, but individuals are always limited in what they can do.

One way of overcoming these individual limitations is to use methods that allow groups of people to collaborate effectively on the axiological futurist project. We can call such methods 'collective intelligence' methods (Mulgan 2017; Malone 2018; Hogan, Johnston and Broome 2015; Hogan, Hall and Harney 2017). It may seem a little odd to single these out as a distinctive subset of methods. After all, the hope is not that groups will pursue different methods but, rather, pursue the methods outlined above more efficiently and effectively. It may also seem a little redundant: surely all research projects are ultimately pursued by groups, even if only indirectly? No person is an island unto themselves: even quintessential lone geniuses like Einstein and Darwin had collaborators to help with their projects.

This is all true, but it is worth addressing collective intelligence methods separately for three main reasons. First, the idea of ‘collective intelligence’ features centrally in the proposed map of the future axiological possibility space that is outlined below in section 3. Foreshadowing it here will make it easier to understand what is discussed there. Second, as should be clear, axiological futurism necessarily draws upon multiple disciplines. If we had some formal method for usefully collecting and harnessing multiple insights from these respective disciplines we could greatly enhance the scope and depth of axiological futurism. Third, people have already begun to argue that we should think about collective intelligence methods as their own distinct method (Malone 2018, Mulgan 2017, Hogan, Hall and Harney 2017). It is not enough to simply get a group of people with diverse backgrounds and different areas of expertise together in a room and hope that they will produce insights that are greater than the sum of their parts. Good group work is hard to do (Straus, Parker and Bruce 2009). Groups often fail to produce better insights than individuals. They often develop their own biases and collective group think; particular individuals can dominate discussions and deliberations, thereby substituting their own agenda for that of the group; people can also get ‘blocked’ or be overly timid in groups, resulting in them producing fewer, not more, insights than they might achieve on their own steam. So although group work has the potential to overcome the limitations of individualism, it can only do this if it is done in a systematic and thoughtful way.

Fortunately, people have already started to do this and have developed formal methods for enabling groups to work together effectively. I confess to having a vested interest in this idea. In previous work, along with my colleagues, I used formal collective intelligence methods to get an interdisciplinary group to think about how technological transformation might change the future of social governance, and to consider the research questions that

need to be answered as a result of this (omitted). We did this by organising group work into three main phases of activity: (i) an idea generation phase, in which we encouraged individuals within the group to generate as many different ideas as possible in response to a particular research question; (ii) a deliberation and discussion phase, in which members of the group added to and critically evaluated one another's ideas and (iii) a convergence/consolidation phase, in which we got the group to coordinate on producing a particular output (in our case a draft agenda of research questions keyed to relevant methods for answering those question). Breaking group work down into these phases might sound like common sense but it is striking how infrequently it is done. Furthermore, when done explicitly and thoughtfully it is possible to plan specific group work activities — such as idea-writing and structured dialogues — that make the maximum use of each phase. Doing this helped our group to produce an output that would have been impossible if we had worked on our own. There is reason to hope that similar collective intelligence methods could be a boon to the axiological futurist project.

(c) Interim Conclusion

This concludes the discussion of how we might go about doing axiological futurism. I have summarised the key ideas from this section in the table below.

Logical Space Methods	Causal Relationship Methods	Collective Intelligence Methods	Pitfalls/Things to Avoid
<p>Methods dedicated to working out the contours of axiological possibility space.</p> <p>Establish dimensions of variance for particular values and map the resulting logical space.</p> <p>Establish dimensions of variance for multiple values and</p>	<p>Methods dedicated to working out the causal drivers of change (intellectual and material) within axiological possibility space</p> <p>Use historical studies that focus on value change over time</p> <p>Develop cross-</p>	<p>Methods dedicated to getting interdisciplinary groups to collaborate effectively on mapping the possibility space and working out the causal drivers of change</p> <p>Identify and assemble group members</p> <p>Adopt ‘divergent thinking’ methods to enable the group to</p>	<p>Narrow-framing, i.e. focusing on only one technological driver of change or one value/set of values</p> <p>Cultural and individual bias, i.e. being too wedded to one set of values (particularly problematic for descriptive axiological futurism)</p>

<p>map their relations to one another</p> <p>Use cross-cultural and historical analysis to figure out the variation in particular values</p> <p>Use psychological studies on moral psychology (e.g. moral foundations theory) to determine upper limits on the flexibility of moral standards.</p>	<p>cultural studies of natural experiments in value change</p> <p>Use psychological and other experimental studies to examine relationships between causal variables and value change</p> <p>Use computer modelling/game theoretical studies of shifting value equilibria</p>	<p>generate diverse insights or thoughts (e.g. responding to a trigger statement)</p> <p>Adopt processing and collaborative methods to enable the group to comment on and develop one another's ideas (e.g. idea-writing, group deliberation/dialogue)</p> <p>Adopt a 'convergent thinking' method to get the group to coordinate on a shared output.</p>	<p>Group think (when collaborating with others)</p> <p>Mono-disciplinarity</p>
--	---	---	--

Table 2 – Methods for Exploring Axiological Possibility Space

4. The Three Intelligences: A Model of the Axiological Possibility Space

Now that we see *why* axiological futurism is valuable and *how* we might go about doing it, let's turn to *what* it might look like if we did. In this section, I present a map of (a portion of) the axiological possibility space that humans are likely to navigate in the coming decades. Included within this map will be a model of the causal relationships between technological change and axiological change. The implicit claim made by this model is that if we promote and encourage certain technological changes, then we will also promote and encourage certain axiological changes, and vice versa (assuming there is a feedback loop between the intellectual and material drivers of change). The map and the model are the result of my own *informed* speculation, with all the caveats that entails.

The model is inspired by Ian Morris's aforementioned theory of value change (Morris 2015). According to this theory changes in the technology of energy capture affect societal value systems. In foraging societies, the technology of energy capture is extremely basic:

foragers rely on human muscle and brain power to extract energy from an environment that is largely beyond their control. Humans form small bands that move about from place to place. Some people within these bands (usually women) specialise in foraging and others (usually men) specialise in hunting. As a result foraging societies tend to be quite egalitarian. They have a limited and precarious capacity to extract food and other resources from their environments and so they have to share when the going is good. They are also tolerant of using some violence to solve social disputes and to compete with rival groups for territory and resources. They display some gender inequality in social roles, but they tend to be less restrictive of female sexuality than farming societies. Consequently, they can be said to value inter-group loyalty, (relative) social equality, and bravery in combat. These are the foundations of their value systems. Farming societies are quite different. They capture significantly more energy than foraging societies by controlling their environments, by intervening in the evolutionary development of plants and animals, and by fencing off land and dividing it up into estates that can be handed down over the generations. Prior to mechanisation, farming societies relied heavily on manual labour (often slavery) to be effective. This led to the moralisation and justification of social stratification and wealth inequality, but less overall violence. Farming societies couldn't survive if people constantly used violence to settle disputes. There was a focus on orderly dispute resolution, though the institutions of governance could be quite violent. There was much greater gender inequality in farming societies because (a) women were required to take on specific roles in the home, and (b) the desire to transfer property through family lines placed a special value on female sexual purity. This affected their foundational values around gender and wealth equality. Finally, fossil fuel societies capture enormous amounts of energy through the combustion and exploitation of fossil fuels (and later electricity, nuclear power, and renewable energy sources). This enabled greater social complexity, urbanisation, mechanisation, electrification

and digitisation. It became possible to sustain very large populations in relatively small spaces, and to facilitate more specialisation and mobility in society. As a result, fossil fuel societies tend to be more egalitarian than farming societies, particularly when it comes to political and gender equality, though less so when it comes to wealth inequality. They also tend to be very intolerant of violence, particularly within a defined group/state.

The model I develop here takes two key ideas from Morris's theory. The first is the notion of an '*ideal type*' of social order. Human society is complex. We frequently use simplifying labels to make sense of it all. We assign people to general identity groups (Irish, English, Catholic, Muslim, Black, White etc) even though we know that the experiences of any two individuals plucked from those identity groups are likely to differ. We also classify societies under general labels (Capitalist, Democratic, Monarchical, Socialist etc) even though we know that they have their individual quirks and variations. Max Weber argued that we need to make use of such 'ideal types' in social theory in order to bring order to the complexity (Weber 1949), while being fully cognisant of the fact that the ideal types do not necessarily correspond to social reality. Morris makes use of ideal types in his analysis of the differences between foraging, farming and fossil fuel societies. He knows that there is no actual historical society that corresponds to his model of a foraging society. But that's not the point of the model. The point is to abstract from the value systems we observe in actual foraging societies and use them to construct a hypothetical, idealised model of a foraging society's value system. It's like a Platonic form — a smoothed out, non-material 'idea' of something we observe in the real world — but without the Platonic assumption that the form is more real than what we find in the world.

This brings me to the second idea. The key motivation for the model I will now develop is that one of the main determinants of our foundational values is not the technology of energy capture that we rely upon but, rather, the form of intelligence that is prioritised and mobilised in society. I here define ‘intelligence’ as the capacity to solve problems across different environments (Malone 2018; Mulgan 2017). Intelligence is a basic resource and capacity of human beings and human civilisations (Tainter 1988; Turchin 2007; Henrich 2015). It’s what we rely upon for our survival and it’s what makes other forms of technological change possible. For example, the technology of energy capture that features in Morris’s model is, I would argue, itself dependent on intelligence.

I submit that there are three basic forms that intelligence can take: (i) individual, *i.e.* the problem-solving capacity of individual human beings, (ii) collective, *i.e.* the problem-solving capacity of groups of humans working and coordinating together, and (iii) artificial, *i.e.* the problem-solving capacity of machines. For each kind of intelligence there is a corresponding ideal type of axiology, *i.e.* a system of values that protects, encourages and reinforces that particular mode of intelligence. Since these are ideal types, not actual realities, it makes most sense to think about the axiologies we see in the real world as the product of tradeoffs or compromises between these different modes of intelligence. Much of human history has clearly involved a tradeoff between individual and collective intelligence. It’s only more recently that ‘artificial’ forms of intelligence have been added to the mix. What was once a tug-of-war between the individual and the collective has now become a three-way ‘contest’ between the individual, the collective and the artificial. My contention is that the axiological possibility space that we navigate over the coming decades will be defined by these three ideal types of axiology associated with individual, collective and artificial intelligence.

That's the model in a nutshell. It might seem a little abstract and opaque at this point. Let's clarify by translating it into a picture. In Figure 1, I've drawn a triangle. Each vertex of the triangle is occupied by one of the ideal types of society: (i) the society that prioritises individual intelligence, (ii) the society that prioritises collective intelligence, and (iii) the society that prioritises artificial intelligence. The claim being made is that societies can be classified according to their location within this triangle. For example, a society located midway along the line joining the individual intelligence society to the collective intelligence society would prioritise technologies that enhance both individual and collective forms of intelligence, and would have an axiology that mixed the values associated with both. A society located at the midpoint of the triangle as a whole, would include elements of all three of the ideal types. And so on.

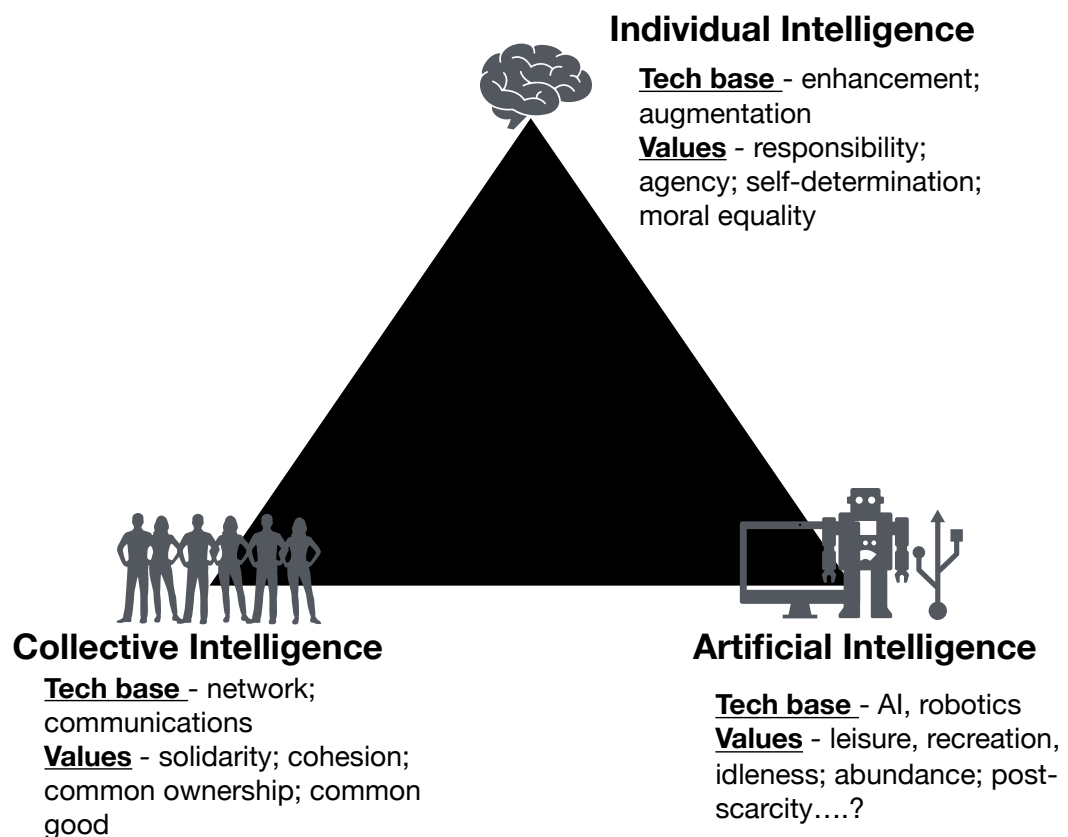


Figure 1 – The Intelligence Model of Axiological Space

The value of this picture depends on what we understand by its contents. What follows is a brief sketch of each ideal type:

Individual Intelligence Society: Individual intelligence is the intelligence associated with individual human beings, i.e. their capacity to use mental models and tools to solve problems and achieve goals in the world around them. In its idealised form, individual intelligence is distinct from collective and artificial intelligence. In other words, the idealised form of individual intelligence is self-reliant and self-determining. It is promoted by any and all technologies that promote individual problem-solving capacity and self-reliance. This includes most ‘tools’ and could also include technologies of individual enhancement (e.g. cognitive enhancers, cyborgification, and genetic engineering). The associated ideal type of axiology will consequently place an emphasis on intelligent individuals as the most important moral subjects and will try to protect their interests, identify their responsibilities, and reward them for their intelligence. It will ensure that the individual is protected from interference (i.e. that they are free and autonomous); that he/she can benefit from the fruits of their labour; that their capacities are developed to their full potential; and that they are responsible for their own fate. In essence, it will be a strongly liberal axiological order.

Collective Intelligence Society: Collective intelligence is the intelligence associated with groups of human beings, and arises from their ability to coordinate and cooperate in order to solve problems and achieve goals. Examples might include a group of hunters coordinating an attack on a deer or bison, or a group of scientists working in lab trying to develop a medicinal drug. Collective intelligence thrives on technologies that enable group communication and coordination, e.g. networking and information communication

technologies. The idealised form of collective intelligence sees the individual as just a cog in a collective mind. The associated ideal type of axiology is one that emphasises the group as the most important moral subject, and values things like group cohesion, collective welfare, common ownership, and possibly equality of power and wealth (though equality is, arguably, more of an individualistic value, and so cohesion might be the overriding value). In essence, it will be a strongly communistic/socialistic and possibly nationalistic axiological order.

I pause here to repeat the message from earlier: these are ideal types of axiological order. There never was a primordial liberal state of nature in which individual intelligence flourished. On the contrary, it is more likely that humans have always been social creatures and that the celebration of individual intelligence came much later on in human development (Schneewind 1998; Siedentop 2012). Nevertheless, I also suspect that there has always been a compromise and back-and-forth between the two poles.

Artificial Intelligence Society: Artificial intelligence is the kind of intelligence associated with computer-programmed machines. It is inherently technological. It mixes and copies elements from individual and collective intelligence (since humans created it and their data often fuels it), but it is also based on some of its own tricks. It functions in forms and at speeds that are distinct from human intelligence. It is used initially as a tool (or set of tools) for human benefit: a way of lightening or sharing our cognitive burden. It can, however, function autonomously and without human input. It is even possible that, one day, AIs will pursue goals and purposes that are not conducive to our well-being (Bostrom 2014). The idealised form of AI is one that is independent from human intelligence, i.e. does not depend on human intelligence to assist in its problem solving abilities. The associated ideal type of axiology is, consequently, one in which human intelligence is devalued; in which machines

do all the important cognitive work; and in which we are treated as (at best) moral patients (beneficiaries of their successes). Think about the future of automated leisure and idleness that is depicted in a movie like *Wall:E* or, perhaps, in Ian M Banks's *Culture* novels. Instead of focusing on individual self-reliance and group cohesion, the artificially intelligent axiology will be one that prioritises human pleasure, recreation, game-playing, idleness, and machine-mediated abundance (of material resources and phenomenological experiences) (reference omitted).

The sketch of this last ideal type of axiology is, admittedly, deeply anthropocentric: it assumes that humans will still be the primary moral subjects and beneficiaries of the artificially intelligent social order. You could challenge this and argue that a truly artificially intelligent order would be one in which machines are treated as the primary moral subjects (Gunkel 2018). That's a possibility that should be entertained. For now, I stick with the idea of humans being the primary moral subjects because I think that is more technically and politically feasible, at least in the short to medium term. I also think that this idea gels well with the model I've developed. It paints an interesting picture of the arc of human history: Human society once thrived on a combination of individual and collective intelligence. Using this combination of intelligences we built a modern, industrially complex society. Eventually the combination of these intelligences allowed us to create a technology that rendered our intelligence obsolete and managed our social order on our behalf. This adds a new element to the axiological possibility space that we will navigate over the coming decades (figure 1).

There are problems with this model. It's overly simplistic; it assumes that there is only one determinant of fundamental values; and it seems to ignore or overlook moral issues that

currently animate our political and social lives (e.g. identity politics). Still, by focusing on the abstract property of intelligence as the major driver of axiological change, the model provides a starting point from which a more complex sketch of the axiological possibility space can be developed. I want to close by suggesting some ways in which this model could be (and, if it has any merit, should be) developed:

- Other potential dimensions of variance and/or ideal types of social order should be offered and evaluated.
- A more detailed sketch of the foundational values associated with the different ideal types should be provided.
- The links between the identified foundational values and different social governance systems should be mapped in more detail.
- An understanding of how other technological developments might fit into this ‘triangular’ model is needed.
- A normative defence of the different extremes, as well as the importance of balancing between the extremes, is needed so that we have some sense of what is at stake as we navigate through the possibility space. This would be essential if we are to pursue axiological futurism from a normative stance.

In short, we need to make full use of the methods outlined in the previous section in order to explore the possibility space as best we can. In this respect, (somewhat ironically) collective intelligence methods could be particularly valuable. Perhaps there could be a series of mock ‘constitutional conventions’ for the future, in which such groups actually draft and debate the different possible ideal type axiologies?

5. Conclusion

In conclusion, axiological futurism is the systematic and explicit inquiry into the axiological possibility space for future human (and post-human) civilisations. Axiological futurism is necessary because, given the history of axiological change and variation, it is very unlikely that our current axiological systems will remain static and unchanging in the future. Axiological futurism is also important because it is complementary to other futurological inquiries. While it might initially seem that axiological futurism cannot be a systematic inquiry, this is not the case. Axiological futurism is an exercise in informed speculation. The job of the axiological futurist is to map the axiological possibility space and consider how civilisations might shift and change location within that possibility space in the future. The goal is not precise prediction but, rather, scenario planning. In doing this, the axiological futurist can call upon a number of disciplines for assistance, including philosophy, history, anthropology, and psychology. I have tried to show how this might be done by presenting a model of the future axiological possibility space that focuses on the role of intelligence in shaping our foundational values. I hope that others join the cause and develop axiological futurism into a distinctive branch of research.

Acknowledgments: The author would like to thank MM, SB and MH for feedback on earlier drafts of this paper.

References

Appiah, Kwame Anthony (2010). *The Honor Code: How Moral Revolutions Happen*. New York: W W Norton & Co.

Alexander, JM (2008). *The Structural Evolution of Morality*. Cambridge: Cambridge University Press.

Armstrong, Stuart (2014). *Smarter than us: The Rise of Machine Intelligence*. Machine Intelligence Research Institute.

Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon & Iyad Rahwan (2018). The Moral Machine Experiment. *Nature* 563: 59–64

Bicchieri, Cristina (2016) Norms in the Wild: How to Diagnose, Measure, and Change Social Norms. Oxford, UK: OUP.

Baum, Seth, Armstrong, S, Ekenstedt, T. Olle Häggström, Hanson, R., Kuhlemann, K. Maas, M.M., Miller, J., Salmela, M, Sandberg, A., Sotala, K., Torres, P., Turchin, A. and Yampolskiy, R. (2019). Long-Term Trajectories of Human Civilization. *Foresight* 21(1): 53-83, DOI 10.1108/FS-04-2018-0037.

Bostrom, Nick (2005). Transhumanist Values. *Review of Contemporary Philosophy*, Vol. 4, May 2005.

Bostrom, Nick. (2013). Existential Risk Prevention as Global Priority. *Global Policy* 4(1): 15-30

Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: OUP.

Brin, David (1998). *The Transparent Society*. New York: Basic Books.

Bryson, Joanna (2018). Patience is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology* 20(1): 15-26

Campbell, Richmond (2014). Reflective Equilibrium and Consistency Reasoning. *Australasian Journal of Philosophy* 92(3): 433-453.

Currier, Richard (2015). *Unbound: How Eight Technologies Made us Human and Brought Our World to the Brink*. New York: Arcade Publishing.

[-----]

Eubanks, Virginia (2017). *Automating Inequality*. New York: St Martin's Press.

Fernández-Armesto, Felipe (2019). *Out of Our Minds: What we Think and How We Came to Think It*. London: OneWorld

Flanagan, Owen (2017). *The Geography of Morality*. Oxford: OUP.

Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, & Peter H. Ditto. (2013) Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology* 47: 55-130

Greene, Joshua (2013). *Moral Tribes*. London: Penguin.

Gunkel, David (2018). *Robot Rights*. Cambridge, MA: MIT Press.

Haidt, Jonathan (2012) *The Righteous Mind*. London: Penguin.

Hanson, R. (2016). *The Age of Em*. Oxford: OUP.

Harari, Yuval Noah, (2016) *Homo Deus*. London: Harvill Secker.

Henrich, Joseph (2015) *The Secret of Our Success*. Princeton, NJ: Princeton University Press 2015.

Hogan, MJ, Johnston, H, Broome, B (2015) Consulting with citizens in the design of wellbeing measures and policies: lessons from a systems science application. *Social Indicators Research* 123: 857–887

Hogan, M. J., Hall, T., & Harney, O.M. (2017). Collective Intelligence Design and a New Politics of System Change. *Civitas Educationis*, 6(1), 51 – 78.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores. *ArXiv:1609.05807 [Cs, Stat]*, September 19, 2016. <http://arxiv.org/abs/1609.05807>

Kuhn, Thomas (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Kumar, Victor and Campbell, Richard (2016). Honor and Moral Revolution. *Ethical Theory and Moral Practice* 19: 147-159.

Lecky, William (1955 - originally published 1869), *The History of European Morals*. New York: George Braziller.

Leiter, Brian (2015). Constitutional Law, Moral Judgment and the Supreme Court as SuperLegislature. *Hastings Law Journal* 66:1601-1617

List, Christian and Valentini, L. (2016). Freedom as Independence. *Ethics* 126 (4):1043–1074

Malone, Thomas (2018). *Superminds: The Surprising Power of People and Computers Thinking Together*. London: Oneworld.

McClain, Linda (2018). Prejudice, Moral Progress, and Being 'On the Right Side of History': Reflections on *Loving v. Virginia* at Fifty. *Fordham Law Review* 86:2701.

Moody-Adams, Michele M. (1999). The Idea of Moral Progress. *Metaphilosophy* 30(3): 168-185

Morris, Ian (2015). *Foragers, Farmers and Fossil Fuels*. Princeton NJ: Princeton University Press

Mulgan, Tim (2017). *Big Mind: How Collective Intelligence Can Change Our World*. Princeton, NJ: Princeton University Press.

Peppet, Scott (2015). Unraveling Privacy: The Personal Prospectus and the Threat of a Full-Disclosure Future. *Northwestern University Law Review* 105: 1153.

Pinker, Steven (2011). *The Better Angels of Our Nature*. London: Penguin.

Pleasants, Nigel (2018). The Structure of Moral Revolutions. *Social Theory and Practice* 44(4): 567-592

Popper, Karl (1957) *The Poverty of Historicism*. London: Routledge.

Robertson, Jennifer (2017) *Robo Sapiens Japonicus*. Berkeley, CA: University of California Press.

Roemer, John (1998). *Theories of Distributive Justice*. Cambridge, MA: Harvard University Press.

Schneewind, JB (1998) *The Invention of Autonomy*. Cambridge, UK: Cambridge University Press.

- Shahar, Avin (2019). Exploring Artificial Intelligence Futures. *Journal of AI Humanities* 2: 169-194
- Siedentop, Larry (2011). *Inventing the Individual*. London: Penguin
- Skyrms, Brian (1997) *Evolution of the Social Contract*. Cambridge, UK: Cambridge University Press.
- Stokes, Patrick (2017). Towards a New Epistemology of Moral Progress. *European Journal of Philosophy* 25(4): 1824-1843
- Straus S, Parker A, Bruce J, et al. (2009) Group matters: A review of the effects of group interaction processes and outcomes in analytic teams. *RAND Working Paper*. Available at: http://www.rand.org/content/dam/rand/pubs/working_papers/2009/RAND_WR580.pdf
- Tainter, Joseph (1988). *The Collapse of Complex Societies*. Cambridge, UK: Cambridge University Press.
- Tomasello, Michael (2016). *A Natural History of Morality*. Cambridge, MA: Harvard University Press.
- Torres, Phil (2017). *Morality, Foresight and Human Flourishing: an Introduction to Existential Risk*. Durham, NC: Pitchstone Publishing.

Turchin, Peter (2007). *War and Peace and War: The Rise and Fall of Empires*. New York: Plume

Vallor, Shannon (2016). *Technology and the Virtues*. Oxford: OUP.

Van De Poel, Ibo (2018). Design for Value Change. *Ethics and Information Technology* DOI:10.1007/s10676-018-9461-9

Weber, Max (1949). *Methodology of the Social Sciences* - translated by Edward Shils and Henry Finch. Glencoe, IL: Free Press.

Williams, Evan (2015). The Possibility of an Ongoing Moral Catastrophe. *Ethical Theory and Moral Practice* 18(5): 971-982

Yudkowsky, Eliezer (2011). Complex Value Systems are Required to Realize Valuable Futures. *Machine Intelligence Research Institute*, available at <https://intelligence.org/files/ComplexValues.pdf>