Interview with John Danaher on axiological futurism: in pursuit of a better understanding of the relationship between new technologies, risks, and ethics considering value changes[1]

Murilo Mariano Vilaça[2]

Murilo Karasinski[3]

Presentation

Scientific and technological development has broadened the possibilities for human intervention over, so to speak, external (environment) and internal (oneself) nature. The expansion of the human capacity for intervention constitutes a long-term process that has produced what Bailey (2004) calls liberation from biological constraints, but, on the other hand, generates as a side effect of threat to our well-being, to the preservation of the environment and, in the limit, even to our survival. Thus, the debate on the

[2] Ph. D. Researcher at National School of Public Health, Oswaldo Cruz Foundation (ENSP/ Fiocruz), Rio de Janeiro, RJ – Brazil. Research Fellow at FAPERJ (JCNE) and CNPq (APQ – PRÓ-HUMANIDADES). https://orcid.org/0000-0001-9720-5552. E-mail: murilo.vilaca@fiocruz.br.

[3] Ph. D. Professor at Pontifical Catholic University of Paraná (PUCPR), Curitiba, PR – Brazil. https://orcid.org/0000-0002-6099-6968. E-mail: k.murilo@pucpr.br.

relation between benefits and risks of techno-scientific progress is a hallmark of modern societies.

Often, one sees a polarization in the debate. That is, while some focus on the benefits, others highlight the risks. Thoughtful approaches – which are far more complex, given the variety of factors involved in such an analysis – are uncommon.

One of the fundamental factors that are systematically overlooked is value. Briefly put, something is seen as a benefit (or as beneficial) if it promotes a value (or something we value); something is seen as a risk if it puts a value (or something we value) under threat. That is, there is an intrinsic relationship among benefit, risk and value that has not been adequately addressed.

It is precisely this relationship that our interviewee addresses, raising the issue – which is already quite complex in itself – to an even higher level of difficulty. Instead of focusing on the relationship among science, technology and value in the present, he proposes that we investigate the *future of values*. This field he calls *axiological futurism*. According to Danaher (2021, p. 1-2),

> [...] 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. In other words, 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).

John Danaher is Senior Lecturer in the School of Law, NUI Galway, University Road, Galway, Ireland. He has an impressive set of publications in internationally prestigious journals (*Neuroethics*; *Law, Innovation and Technology*; *Science and Engineering Ethics*; *American Journal of Bioethics*; *Cambridge Quarterly of Healthcare Ethics*; *Bioethics*, *Futures*, among others). Many publications relate specifically to analyses of new technologies and their applications.[4]

In the kindly given interview, Danaher, starting with a brief recovery of his trajectory, addressed central points of the concept of axiological futurism, presenting examples of changing values, clarifying his interesting proposal. In addition, he indicated his next steps in the development of the proposal, as well as authors who have influenced it or who have dialogued with it in some way.

---

[4] For detailed information, see https://www.nuigalway.ie/our-research/people/law/johndanaher/.

Besides contributing for axiological futurism to become known in Brazil, we hope[5] that the interview helps in the reflection about the relation among new technologies, risks and ethics considering the value change factor.

Interview

Murilo Karasinski (MK)/Murilo Vilaça (MV): So let's start. To begin our interview, we would like you to comment on your academic trajectory. Why have you researched and written the things that you have? If possible, we would like to propose a common thread: much of your research concerns the variety of important topics grouped under what we might call the "human enhancement debate". So that's our first question, John.

John Danaher (JD): Thanks for the implied compliment in the question. I haven't thought about the trajectory of my academic work in the way that you suggest in the question. I suppose you could say that it does all fit within the framework of human enhancement, and that's what I have focused on, where enhancement is broadly conceived. I suppose traditionally within bioethics, or neuroethics, enhancement is conceived somewhat narrowly as involving the use of pharmacological treatments to enhance mental function in some dimension, or maybe performance enhancing drugs in sport would be counted as a type of human enhancement. I guess more laterally and more recently, people have looked at things like deep brain stimulation or transcranial magnetic stimulation as being enhancing technologies as well. But, of course, if we take a very broad view of enhancement, virtually all technology could count as a type of enhancement, or potentially perhaps as a type of enhancement. In fact, there are people like Nicholas Agar, from New Zealand, who would argue that "external enhancements" – the use of technology to augment human biology through our hands and our eyes and so forth – that's a more effective form of enhancement than these pharmacological and brain stimulation forms of enhancement[6]. And you can even go further and say institutions are a kind of enhancement, right? So you

---

[5] A start of discussion on the subject was debated at the Cambridge Conference on Catastrophic Risk 2022, under the title "Making it More Complex: Axiological Futurism within the reflection on Existential Risks", whose video can be found, in full, at this link: https://www.youtube.com/watch?v=Sw5khDf5phI.

[6] This argument present in the human enhancement debate is addressed by Danaher (2016).

have the rule of law as a way of enhancing human society, make it function more smoothly or something like that… or having effective markets that are incentivize specialization, trade and economic growth and they constitute a form of enhancement. I mean, if you take that broad view of enhancement, you could say that a lot of what I've written falls within that field of inquiry, and that's what I'm interested in. But I would say that when I'm writing individual papers or books or whatever, I don't necessarily think of it as having a unifying theme. I tend to just focus on particular questions or problems or puzzles that interest me. Maybe I'm drawn to particular questions or problems or puzzles because they fall within that area, but from the inside, I haven't thought of my research as having a trajectory, if you want to put it that way.

MK/MV: All right. Thank you, John. We will start our second question here: if we are right, the proposal to create a new field of research – about axiological futurism – is part of the debate about the relationship between risk and value, which is very broad and relatively old. If it is possible, we would like you to comment on what gaps in this debate motivated your proposal.

JD: So this question is specifically about the paper I wrote on axiological futurism (DANAHER, 2021), right? Yeah. So, in the end, you're correct in that the debate about risk and value is a very old one. I suppose a lot of futurist inquiries focus on what are the possible forms of future human civilization, or what are the trajectories of technology and technological development and growth over time. And they often involve, extrapolating from what we might call "hard features" of human society, so most obviously material technologies. We have particular kinds of computers and computer technology nowadays and we think they might develop in the future... well, you know, they'll get faster, they'll get more intelligent, and that will have an impact on human life in various ways. Some of those ways might be positive, some of those ways might be negative. So we focus on the kind of material technology and how that will change over time, and then draw inferences from our predicted trajectory of technological development to the impacts on human society. And we see that debate playing out in just across most debates about responsible innovation – how nuclear technology will develop over time, or how energy harnessing technologies in general will develop over time, the impacts of industrial technologies on the environment… We say

that, well, this is the current trend line, it's going to continue on this upward trajectory for X number of years, then there's going to be a crash, we're going to have less of whatever kind of technology we use over time. What are the implications of that for human society? That's usually how we think about the risks and rewards of technology. We focus on the material technology first, and then focus on the impacts on human society values, institutions, rules, norms, that kind of thing, afterwards. And I suppose a simple critique of that style of thinking about the future is that it assumes that a lot of our values, norms, and rules are static in some sense, that the technology is the thing that changes and the values, norms and rules stay the same and can be used to evaluate the technological change. If you look at, let's say, the European Union, which is where I'm based, most of the debate about innovation says that we have these core values in European civilization – human dignity, freedom, equality. These are our core civilizational values, and we need to make sure that any technology that we are developing is developed in a way that is consistent with those values. So we treat those values as static, fixed, and we have to try and make sure that the technology doesn't step outside the boundary or risk or potential harm or damage those values. So I think that's a useful way of thinking about the future. I know I followed that pattern myself in the past, but I think it is also important to bear in mind that values are also things that change over time, and that the values that our ancestors had or even if you go back one or two generations, the values of my parents and grandparents had, they're similar to my values but they're different as well. I grew up in a country that is, I guess, historically a conservative catholic country, so social morality was largely conservative catholic in nature. That has changed over the past 50 years. My parents grew up in a time where a church-imposed social morality was very common, and they still carry with them elements of that social morality. I don't carry the same kind of cultural baggage, so we have different values. That's a common experience for most people, in that your children will end up having different values than you and you have different values, to some extent, to your parents. And that difference grows over time. If you look further back into the deeper recesses of history, the values of somebody living in the 1800s would be quite different to the values that I have nowadays. The values of somebody living in the Middle Ages would be even more different from the values we have nowadays. And so we have to factor that in as well when we think about the future. There's no reason to think that we've somehow arrived at the final enlightened set of human values and norms and rules.

The way in which I sometimes think about it is that… and this might be a bit grandiose… but we can talk about Copernican shifts in our thinking. The pre-Copernican thought was roughly that the Earth was at the center of the universe. Post Copernicus, post Galileo, that view changed. We shifted perspective. The Earth is no longer at the center, the Sun is at the center. And obviously since then we've even gone through further shifts in perspective in that the Sun is just one star amongst many, and we live in one galaxy amongst many, right? So we've constantly decentered ourselves, cosmologically speaking, through a series of Copernican shifts. People often think about the impact of Darwin's theory of evolution is in similar terms. The pre-Darwinian view was that humans sat atop this kind of ladder of creation, that we were the pinnacle of biological life on Earth, that we were somehow special. But the post-Darwinian view is that actually there's nothing special about us, we're just one branch amongst many of the possible evolutionary branches. So that's a kind of the Copernican shift in how we think about our position relative to other biological creatures. I think of the core idea in that axiological futurism paper as involving or requiring or encouraging people to take a similar Copernican shift when it comes to this space of values or norms that we have, that there's a tendency to think that our current value system is somehow perfected and that there's no possible way of getting outside of it, but actually our value system is just one among many. There's been historical variations in values. There are also cross-cultural and geographical variations in values, and so we need to decenter ourselves when it comes to thinking of values, too. These are things that also change over time.

MK: Luciano Floridi has a book called *The 4th Revolution* (FLORIDI, 2014), and he has some ideas that are interesting regarding this theme of Galileo, Darwin, Freud and things like that… One thing that came to my mind, it wasn't in our previous questions, John, just… in the sense that values change over time in our places, do you see any difference between changing values or values changing in Ireland or Brazil, for example, in the next 100 years or something like that? Or does this kind of thing doesn't belong to your study, this thing regarding changing values in different places all over the world?

JD: Yeah, so axiological futurism, broadly conceived, would have to factor in cross-cultural differences in moral trajectories of values. I'm not sufficiently well-informed about Brazil to make any comments about how its value system or structures might change in comparison to Ireland, where I live. I just don't have enough knowledge of it, of the history, social context, or political context to say anything about it. But if you take the evolutionary analogy, you know, that species that develop in different geographical locations, in different continents, can take different trajectories. They can originate in the same point but they follow different pathways, and they can evolve apart over time. Something similar can happen with societies for a variety of reasons. I tend to focus predominantly on the role that technology plays in changing values and norms. That's the thing that I am most interested in. But absolutely there are lots of other things that change values and norms over time. So cultural history, the kind of institutional history of a country, makes a bit of a difference. That's actually something I've been writing a paper[7] about recently with Jeroen Hopster, from the University of Twente. One of the examples we have is a comparison between different Constitutional systems, and I picked the Irish Constitution and the U.S. Constitution just as an illustration because they are two jurisdictions that I knew reasonably well. The Irish Constitution is very easy to amend or change. You can change any provision in the Irish Constitution by a simple majority of the voting population, which means it has kind of a flexibility built into it. And it has, in fact, been amended, I think, over a dozen times in the past 20 years. And the Irish Constitution is not unusual in that respect. You can find similar amendable Constitutions in many European countries. Switzerland is a famous example of a country that changes its Constitution using a regular a participatory system of changing Constitutional rules over time. The U.S. Constitution is very hard to change. I mean, it was designed that way, but it's very difficult to change. Without going into the technical details, there are two ways to amend the US Constitution: either you get a super majority of both houses of Congress to change the Constitution, which is not going to happen at the moment given the kind of polarization within the political climate in the US, or a super majority of all the states can change the Constitution, and that hasn't happened in the past 50 years. In fact, the last change that they had was a very minor change that was actually an amendment that was proposed over 200 years ago and was about the payment of Congress people, so they all had an interest in changing

---

[7] The paper was published: DANAHER, J.; HOPSTER, J. The normative significance of future moral revolutions. **Futures**, v. 144, p. 1-15, 2022.

it. So there really haven't been substantive changes in the US Constitution in quite a long time. That kind of institutional framework doesn't have the same flexibility built into it. So if you think about it, that means that countries with a more flexible Constitutional system are open to more of the value changes over time versus Constitutional systems that have less flexibility built into them. So that has nothing to do with technology per se. It has to do with the institutional structure of the two countries, and that means that they can take very different paths over time.

MK: All right. Thank you, John. I am going to ask you the third question. I was discussing with Murilo that this is a tricky one. I don't know if it's clear enough but let's go.

MK/MV: The third one is: the disruptive possibilities arising from the extreme techno-scientific development that characterizes the epoch called the Anthropocene demand a "research of the future", which involves anticipation and preparatory research. This proposal is perhaps one of the most ambitious in this sense, not least because it moves toward a post-human future. In our view, it has made research into the risks of human enhancement technologies much more complex. How does this "looking into the future" fit into the broader landscape of moral reflection? More specifically, we would like to hear from you about how the idea of axiological possibility space, which would be characterized by the mapping of what changes diachronically (values), such as deontology, which is traditionally understood in the Kantian sense, that is, as something that would be resistant to change, adaptation, accommodation. I don't know if it's clear enough. Well, that's our third question. We would like to hear from you if it's understandable.

JD: Yeah, I feel like I could take this into two parts, in the sense that there's a query here about the fact that if you include values and morals in the discussion about what changes over time, and the things that future-oriented inquiries are interested – that makes future oriented inquiries much more complex. So I think that's certainly true because, again, you're not just thinking about one thing changing like, again, just to use the initial example, like we were talking about how material technology changes over time,

how computer technology changes over time, and so you can maybe make meaningful predictions about how that will change over time if you get to some new law of technological development like Moore's law or something like this, right? That technology will speed up every 18 months to 2 years, or more, and that makes prediction about the future relatively easy. But if you start to include changes in soft features of human society, like institutions and norms and values, that's a much more complex thing, and there are maybe fewer constraints about the possible forms that human civilization could take. This is one of things I say in the paper. I say that the goal of axiological futurism is to try to map the space of possible axiological futures for human civilization and how we might move about within that space. But one of the points that I make is that axiological possibility space is vast, just like there's an incredible number of possible value systems, it's not clear to me that you can… You couldn't possibly describe them all, ok? There have been some attempts to do this. I don't know if you're familiar with Oliver Scott Curry's work. He is an anthropologist, partly affiliated with the University of Oxford, and he developed this idea of the theory of morality as a cooperation, which is an evolutionary theory of morality. Mark Alfano is a philosopher as well and has done some work with him. So they have a paper on moral combinatorics, which uses basic combinatorial analysis to try and put some shape on the possible types of moral systems that you could find in human civilization. But their theory works because they assume that there are 7 types of normative systems that are all based on different cooperative games. They say there are 7 cooperative games, and there's basically a limited number of strategies that you can adopt in each of those games that would be stable over evolutionary time and each of those strategies constitutes a possible normative system. So they start to actually put numbers on the possible different moral societies we could have. And I remember in the paper, they ended up with a figure of just over 2,000 possible systems. But then they also want to know: "well, actually, there's a possibility of even more variation than we're suggesting because you can combine the different games in different ways as well, so suddenly it sort of explodes into a vast space of possibilities". So that is a challenge, I mean, trying to reduce the space of possibility that we are inquiring into to make it something more manageable is a challenge. So the only way to really do it is to start introducing constraints in some way that can limit the space of possibility. Take an example that I had a moment ago of the US Constitutional order with its relative inflexibility versus another Constitutional order, you can use that as a constraint and say: "well, that means that certain kinds of changes

aren't possible because at the moment we have that institutional framework that doesn't change". You have to introduce some things like that to make it a more tractable field of inquiry, right? And that means that any practical form of axiological futurism will have significant limitations. Prediction is hard, especially when you're trying to predict the future, to quote Niels Bohr, but the hardness of it isn't the reason to shy away from the inquiry either. I think we can still have the inquiry. We can still try and sketch new possible scenarios, and anticipate possible scenarios, and that can be a valuable and worthwhile endeavor, even if we can't possibly sketch every possible future that arises. So that's one point I would make about that.

The second part of the question, my way of thinking about it, is that maybe there's a tension between the assumption underlying axiological futurism, namely, which is that values are things that are flexible and can change versus certain kinds of moral theory, which suggest that maybe values are unchanging and fixed over time. Let's say, something like the Kantian Constructivism which says that there is one moral law that you can arrive at through rational reflection on the nature of human agency and what is means to be a human agent. There are a couple of things that I would say about that. Number one, and this has been pointed out by other people who have written about technology and ethics in particular, I'd say Shannon Vallor has a pretty good book called *Technology and the Virtues*, which involves a critique of the application of Kantianism or Utilitarianism to a changing civilization, which is that like those kinds of abstract normative principle are not action guiding in their basic form. For example, the basic Kantian principle "Act so that the maxim of your will is at the same time a universalizable maxim" – that is, of course, very empty in its abstract form. It's very hard to know what that means applied to a particular social context, so you have to start adding details to it to make sense of it. And once you start adding details to it and apply it to a specific context, you are in a sense changing or modifying this universal law in order for it to actually have some practical real world guidance. Same is true with, let's say, Utilitarianism. So the core Utilitarian principle is something like: "try and maximize the wellbeing of the most number of people". OK. Maybe that is an unchanging moral law, but in a particular social context, what does it mean to maximize wellbeing? What are the tools available to us to do that? That is something that is going to change as a result of technology. So what it means to follow that rule is something that is going to have to

change and adapt, so we will develop these more specific moral rules that apply to particular social, institutional, cultural, and technological worlds. So even if you think that morality is unchanging and fixed, the particular application of it in a given civilization isn't something that is fixed in the way that some moral theorists might think. It's something that is subject to change over time. More generally, I would say that axiological futurism does not make any strong and metaethical commitments, or it doesn't have any strong metaethical commitments underlying it. It is something I get into a little bit more detail in the paper that I have drafted now with Jeroen Hopster, but we argue that axiological futurism is consistent with a realist metaethics, and is also consistent with a constructivist or relativist metaethics. The only kind of metaethics that is not really consistent with, it would be something like a pure subjective egotistical relativism like "the moral rule is just whatever I believe it to be". But it is consistent with most other metaethics because even if the moral rules or laws exist somewhere out there in a Platonic space, our understanding of it clearly changes over time, and our beliefs about what it entails clearly change over time. You would have to be historically and cross-culturally ignorant to believe otherwise, or to think otherwise. So I think axiological futurism fits quite comfortably within the broader landscape of moral reflection, and is consistent with most existing moral theories.

MK/MV: So our fourth question, last but one is that in the paper, you state that the mechanics of axiological change are relatively simple, that one lesson from history is that expansions in the circle of moral concern are usually considered progressive, but that there is no guarantee that this trend will continue. But you gather and condense a number of historical examples and arguments in favor of the effective possibility of investigating the axiological possibility space for future human and post-human civilizations, even in the face of a certain complexity and relative uncertainty about the future. In view of this, we would like to hear from you about your methodological modeling – of the three intelligences – which seems to be the "heart" of your proposal, if we understood it well.

JD: Yeah, so the first part of that, I kind of engaged with some of those questions in the previous answer about the complexity of moral change, but I think… I can't remember the exact wording that I used in the paper. I think

what I would say is that human moral systems, to me, consist of relatively simple component parts. You have values, which are things that we desire or wish to promote and protect. They are states of the world that are desirable in some sense, like pleasure is desirable in some sense; being well-educated and knowledgeable is desirable; having friends is desirable; maybe, you know, beautiful natural environments are desirable in some way, etc. So they are things that… states of affairs that we want to realize, that we desire, that we pursue – these are values. And the other kind of component of a moral system is some rule for behavior. Things that we ought to do, things that we shouldn't do, things that we are permitted to do, etc. So those component parts are relatively simple. And there is a limited number of ways in which those component parts can change. When it comes to values, you can either add or subtract values, or you can shift the priority of values. You might say all values are equal, or that pleasure is better than education, which is better than having lots of friends, something like that, so you can rank them in different ways, and you can re-rank them. When it comes to rules, essentially you either say something like that law is permissible, is no longer permissible, or is now obligatory. And we see those kinds of shifts over time. For example, consider changes in sexual morality over time. Sometimes a sexual relation or sexual act was deemed impermissible, forbidden, now they're deemed permissible and acceptable, and so forth. And it can happen vice versa as well. So there is a limited number of ways in which things can change. So it's simple in that sense. But it's obviously complex when it comes to the complete range of possible values and possible moral rules.

The other way in which the mechanics of it are complex is with respect to what is the actual driver of change. What causes us to change our set of values that we think are important to us, or what causes us to re-rank values? If you take an example like, let's say, again in the realm of sexual morality, it's probably true that in many countries… I mean, I haven't looked at every single country but certainly in most Western European countries, in America… I think this is broadly true in South American countries too, but I'm not as familiar with that. We've come from a world in which casual sexual relationships outside of marriage were deemed taboo or very risky, or dubious, not the kind of thing that you talk about in polite company. That was basically the situation 100 years ago. Nowadays, they're broadly tolerated, even in countries where there's a dominant cultural belief that is religiously

conservative and opposed to extramarital sex. The practical reality is that most people ignore those religious strictures, right? I mean, it's definitely true in a country like Ireland, where most people are nominally Catholic, but routinely ignore the Churches' teachings on sex outside of marriage. Definitely true in Spain, Italy, and so forth.

MK: Same reality here in Brazil.

JD: What has caused that change in sexual morality over time? There are different theories out there about this. One argument is that it is the technology that has really changed, which is namely the widespread availability of effective forms of contraception, particularly for women. The pill or IUDs and things like that – that's what has changed sexual morality, because what that did was just massively reduce the social costs and risks of having sex outside the marriage. There are people who will counter that and say that "no, that is actually not what really caused the change. You could have had effective contraception but if the legal framework in a country didn't allow for the sale of effective contraception, that wouldn't allow the change in social behavior". I mean, that was true in Ireland. It was legally banned. You couldn't purchase contraception in Ireland until 1972. There was a Constitutional decision that allowed this. Same is true in the U.S. as well, right? We're having this interview a couple of days after this big controversy over the reversing of the abortion decision in the U.S. – the Roe v. Wade decision. But that Roe v. Wade decision is actually based on a right to privacy under the U.S. Constitution. One of the first cases in which that right to privacy was discussed was the case called Griswold v. Connecticut, which was about the right to access contraception for married couples. Basically Ireland followed the exact same trajectory: we identified a constitutional right to marital privacy and then allowed for the sale of contraceptives. So you couldn't have had the social change without that institutional change. How do you get the institutional change? Well, that might have been because of social movements. There were Planned Parenthood groups in the U.S., or social activist groups that were advocating for access to the contraception or educating people about it, and so forth. Or maybe it was like the Swinging 60s, it was like the sexual revolution of the 60s... kind of cultural revolution and sexual permissiveness emerged as a new attitude, which eventually affected the legal framework and that's what

changed. So there are lot of disputes about what exactly changed. Was it the technology? Was it changes to the law? Was it some kind of broader shift in attitudes? My sense of this is that it probably was all of these things. All these things contributed to the change. I think some were probably more important than others. I definitely think legal access to contraception is important. I also think that contraception itself is a key part of the story, and that if you didn't have that… if you didn't have some way to reduce the health risks and social risks of having sex outside the marriage, you wouldn't have had the change in social morality. So you need a technology change, but you also need certain kinds of institutional or legal changes around it. But it could be the case that other kinds of changes in morality aren't so technology dependent or driven. They could be more to do with cultural forces or cultural changes. I think the change in attitudes towards, let's say, same sex relations and same sex marriage is a good example of that. It doesn't seem to me to have been driven by any kind of technological change. It seems to me to have been driven by other kinds of cultural and moral changes. We adopted the more liberal, autonomy-based view of sexual morality as opposed to a religious conservative social morality and sexual morality, and that meant that suddenly the opposition to same sex relations became less defensible to most people. So it can vary depending on the case study that you're looking at.

So that is for the first part of the question. The second part of the question is on the different types of intelligence and the model for the future of society. I think I would disagree with the idea that that's the heart of my proposal. I think that that is a particular application of it or a way of thinking about it. I'm much keener on the idea that we should take future value change seriously and we should have a systematic inquiry into it, and that there are different methods that we can use to pursue that inquiry. That's what I think is really important. But I also then discuss in that paper how would I go about doing it, like there's a particular model of one causal factor that might shape value systems, and that might define the space of axiological possibility for future human civilizations. I'm not convinced that the proposed model is the correct model. I think it's just an interesting model that's worth exploring, and the idea behind it is based purely on an analogy with another theory that I came across. Ian Morris, a well-known archaeologist and historian, wrote this book *Forages, Farmers, and Fossil Fuels*, which is about different value systems over time. His main argument is that the technology of energy capture in a

given society shapes its core values, specifically values in relation to equality of various types and violence, whether violence is acceptable or not, and who is entitled to be violent and resolve conflict by violence. He acknowledges the simplicity of this, but he brackets the entirety of human history into three ideal types of civilization: foraging societies who get all their energy from hunting and gathering; agricultural societies who get all their energy from agricultural production; and fossil fuel societies that get their energy from the burning of fossil fuels. He then maps out how this has changed values in different civilizations. The logic behind his model is that in order to survive, all of us need to capture energy, in order to just keep living. For our kind of metabolic production, we have to capture energy and burn it in order to survive. That's what allows us to keep going. So energy just has its outsized importance in human life, so it kind of naturally follows that whatever mechanisms we have available to us for capturing and harnessing energy will have a big impact on our value system and on our set of moral norms and rules. At the very least, in a hunting and gathering society, given how limited the energy supply is and how potentially fragile it is, it's not something that we control perfectly. We have to chase after a prey. We're not growing vegetables in fields that we control. We have to go out and find them. Everyone has to be involved in that process of energy capture in various ways, right? And so there's kind of like an ethical and moral obligation for most people in that society to be involved in that process. Morris also argues that that encourages these societies to be more egalitarian because of how fragile and limited the energy supply is. They have to share in any large food supplies they have and so forth. Once energy production becomes something that we control more tightly, and we don't need everybody to be involved in doing it, then you start getting maybe a richer and more diverse set of value systems as a result of that.

As I said, I think there is a good logic to Morris' framework because of the outsize importance of energy capture in human life. My idea in the paper was that Morris is correct in one sense, that energy capture is important, but how do we capture energy? It's through the application of human intelligence, or intelligence broadly conceived, right? Either the intelligence of individuals working alone, or the intelligence of groups of people working together, like "here's a technique for doing this thing", and "here's a way of smelting iron", developing tools that we can use to till the fields, plow the fields or hunt down animals. We share that technique over time, that's what enables us to

effectively capture energy, which changes our value system. Same with fossil fuel societies. How do we even get to having fossil fuel societies? It's because of the application of human intelligence to the process of capturing energy. So I was trying to say that, as a step back from what Morris was saying, energy capture is certainly of outsize importance, but intelligence is even more important because that's how we are going to be able to capture energy. We solve the problem of energy capture through the application of intelligence. And I then argued that there are different forms that intelligence can take. I broke it down into three ideal types: one is the individual and another is the collective, which I mentioned already, which have been hugely influential in human history. What's happening at the moment, in our present era, that's different is now we are actually developing artificial forms of intelligence, not just human, not just collective, but machine intelligence as well, OK? And so I think that the development of the artificial intelligence society will have a significant impact on our social value system. That's what I've been writing about for the past decade. I don't think of human societies as fitting into these kind of pure ideal types. I don't think that we're ever going to have a society that's all about artificial intelligence, but rather kind of different mixes and prioritizations of it, right? So does artificial intelligence work in tandem with groups of humans? Or does it work with individuals? There are different mixes of intelligences that we can imagine.

So let's take an example here, and this might be more controversial, but let's compare China and the European Union when it comes to artificial intelligence policy. This is a very simplistic reading of it, but a simple way of looking at the difference between those two countries, or civilizations, let's say, is that at the moment China is facilitating or enabling the use of A.I. in a way that ignores a lot of traditional, individualistic and human-centric values, or things that are prioritized within the European Union, like privacy. It's willing to use AI as an authoritarian way of a kind of mechanism of social control through things like credit scoring. There's also widespread use of facial recognition technology, and widespread use of different forms of artificial intelligence to manage and control a society, and so there is a sense in which A.I. is imposed from the top down and there's a relative kind of free for all… that there are relatively few constraints on the application of it. Contrast that with the European Union, where the model is that we have to develop trustworthy A.I., and where the goal is to make sure that humans

work well with A.I. and that A.I. doesn't in any way dominate over human values or human preferences. So within my model of these three types of civilization, one that prioritizes individuals and their intelligences, one that prioritizes collectives and their intelligences, and one that prioritizes A.I. – I'd say that China is probably slightly closer to the ideal A.I. society, whereas the European Union is closer to the individual society. A.I. has to be balanced with those things. So I think that will have an impact on social values systems – how we fit artificial intelligence into the existing intelligence infrastructure.

MK/MV: Right. Our last question and the shortest one: we would like to know if you have been following the comments made to your proposal, and if so, if they have motivated any changes or plans to develop it?

JD: As with most things that I have written, it surprises me that anyone reads it, and I'm not always aware of them doing so. So the Axiological Futurism paper, as far as I know, I know you've taken up the idea. I'm only aware of maybe two other people who have taken up the idea. Jeroen Hopster has written a bit about it[8], and I'm writing a paper with him now, so we agreed to kind of pool our efforts on that. There's a couple of people, like Hin-Yan Liu, in Denmark, who has written about it a bit.[9] In fact, I developed this legal disruption framework with him and a bunch of other people, focusing on how AI, in particular, disrupts legal systems and a group of other people. That was actually the original inspiration for the Axiological Futurism paper. But beyond those three people, or three groups of people and yourselves, Jeroen Hopster, Hin-Yan Liu and his collaborators, I'm not aware of anyone else who has taken up the idea, so I'm happy to hear about more of them, and if anyone wants to contribute to the idea or develop it or change it in some way, I'd be happy to hear what they have to say. I should say that in the interest of humility here that the idea I have on the paper isn't novel. There is actually a Dutch project about value change. Ibo van de Poel has been writing about this for some time.[10] He has a slightly narrower way of thinking about it than I do but there are a lot of similarities and overlaps. I am currently contracted to write a book about this topic with MIT Press, so I will certainly be monitoring

---

[8] For example, Hopster (2022).

[9] For example, Liu (2021) and Liu and Maas (2021).

[10] For the most current papers, see Poel (2021) and Poel and Kudina (2022).

closely in the next year or so any developments in the literature, the people who have been commenting on it or who have been writing some of the ideas and see how I can develop and refine the proposal further.

MK/MV: Thank you John for the interview.

References

BAILEY, 2004. Transhumanism: The most dangerous idea? Why striving to be more than human is human. Available at: https://reason.com/2004/08/25/transhumanism-the-most-dangero/. Accessed: Aug. 31, 2022.

DANAHER, J. Why internal moral enhancement might be politically better than external moral enhancement. **Neuroethics**, v. 12, p. 39-54, 2016.

DANAHER, J. Axiological futurism: The systematic study of the future of values. **Futures**, v. 132, p. 1-14, 2021.

FLORIDI, L. **The 4ᵗʰ Revolution**: How the infosphere is reshaping human reality. Oxford, UK: Oxford University Press, 2014.

HOPSTER, J. Future value change: Identifying realistic possibilities and risks. **Prometheus**, v. 38, n. 1, p. 113-123, 2022.

LIU, H-Y. AI challenges and the inadequacy of human rights protections. **Criminal Justice Ethics**, v. 40, n. 1, p. 2-22, 2021.

LIU, H-Y.; MAAS, M. M. 'Solving for X?' Towards a problem-finding framework to ground long-term governance strategies for artificial intelligence. **Futures**, v. 126, p. 1-22, 2021.

POEL, I. v. de. Design for value change. **Ethics and Information Technology**, v. 23, p. 27-31, 2021.

POEL, I. v. de; KUDINA, O. Understanding technology-induced value change: A pragmatist proposal. **Philosophy & Technology**, v. 35, A.N. 40, 2022.