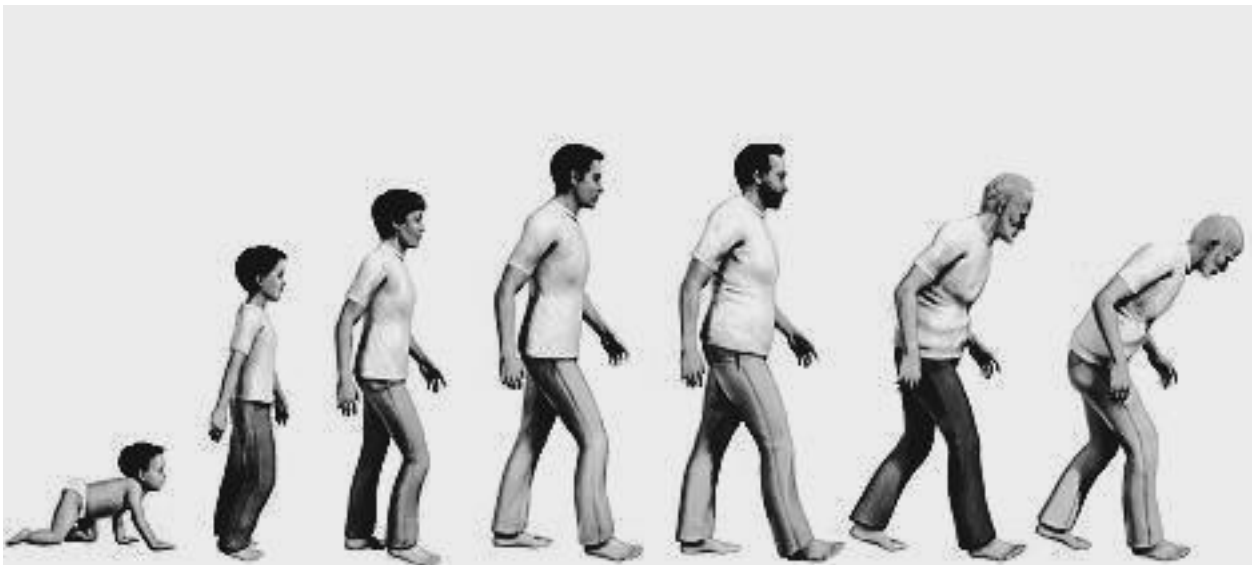


Personal Identity and Public Policy

Conveners:

Simon Beard (Centre for the Study of Existential Risk)

Timothy Campbell (Institute for Futures Studies)



November 1st – 2nd, 2016

Final Report

Contents

Workshop Summary	2
What are the ethical consequences of immortality technology? <i>Francesca Minerva and Adrian Rorheim</i>	3
Why philosophers fail to influence public debate – and how they can do better <i>Simon Beard and Michael Plant</i>	5
List of Papers and Presenters	7

Workshop Summary

Technological change presents certain extreme possibilities for the future of ‘the person’ that are hard to evaluate using standard economic or legal procedures.

- Firstly, there is the possibility of the total loss of all persons from the world, and hence some portion of space up to and including the entire future history of the universe.
- Secondly, there is the possibility that human persons will be replaced with persons, or person-like entities, of other kinds such as enhanced members of other species or artificial intelligences.

This workshop sought to grapple with these two possibilities by considering the ethical importance of persons in contemporary public policy debates. In so doing it examined both how existing policy evaluation procedures dealt with issues of personhood and what different philosophical theories about the nature and importance of personhood implied for public policy.

One aspect of research that was considered is what existing arguments about the nature of personhood in policy areas such as healthcare, criminal justice and human development imply about the nature of persons more generally. Papers from Jeff McMahan (University of Oxford), Tom Douglas (Oxford Uehero Centre for Practical Ethics) and Chase Besnarze (Midwestern University) considered these three policy

areas and argued that in each case it is important to consider personal identity as a matter of degree. Individuals develop over time and change in ways which mean that, even if we accept that a person retains their identity over time, we should still differentiate between their past and future selves. From the perspective of existential risk such arguments could imply that even within existing policy frameworks, persons exists on a spectrum whereby individual selves can be thought to constitute one and the same person to a greater or lesser degree. As Derek Parfit has put it, mere survival does not seem to be what matters morally. What matters in a person’s survival can hold to greater or lesser degrees. This creates possibilities for future evaluative mechanisms to extend the moral space beyond a one size fits all conception of personhood, and encompass higher and lower parts of this spectrum than are currently occupied by existing people. Hence there is a need to incorporate a wider range of ‘near persons’ in these frameworks.

A second aspect of research that was considered concerns more speculative accounts about how we should respond to technologies that challenge our ideas about what constitutes a person, such as head transplants, brain uploading, cryogenics and human enhancement. Papers from Tim Campbell (Institute for Futures Studies) and Francesca Minerva (University of Ghent) contemplated how we should respond to and govern these technologies and in particular showed how our theoretical views about personal identity might influence such deliberations. It seems that whether we see

certain possible futures, for instance those in which humans are uploaded onto computers, as representing a utopia or a catastrophe will depend a great deal on what positions we take on the nature and continuity of persons across time.

This contrasted with a third aspect of research that was considered, posing arguments that directly addressed the relationship between philosophical arguments about personal identity and political institutions. Jeremy Williams (University of Birmingham) argued that personal identity was something of a special case within philosophy because philosophers arguing about its moral importance frequently appealed to metaphysical arguments, which are standardly excluded from secular political discourse - but there did not seem to be any good alternative within political philosophy for debating the political importance of personhood. Hanna Tierny (Cornell) and Max Suffis (Rice) both considered, partially as a response to this kind of concern, whether we should distinguish between different ways of talking about persons and whether we might differentiate between persons characterized as metaphysical entities and persons characterized by fellow members of the moral and / or social communities with whom we interact. Finally, Simon Beard (CSER) suggested that when we consider the value of individual lives we should differentiate between values that depended upon the fact that this life is lived by a particular person,

quality of life, from the more common conception of value that attaches to features of that life that are not related to the person living it, welfare.

The conference ended with a consideration of how philosophers with an interest in personal identity and the moral importance of personhood should interact with policy makers. Suggestions included building upon fields of policy making where the nature of persons is already debated, such as in healthcare allocation and the global burden of disease, and seeking out questions currently being posed in the policy sphere for which philosophers appear to be able to offer compelling and justified answers. It was proposed that philosophers wishing to embed thinking about personal identity should consider five key questions. What public policy issues depend upon personal identity? What positions about personal identity do those already working on these issues seem committed to? What are the implications of taking different positions regarding personal identity for these policy areas? Which of these implications seem to provide a feasible basis for policy making? How can a consensus for change be built around adopting policies that reflect the best available theory or theories?

The themes of the workshop were further elaborated on in the following articles, produced following the workshop by some of its participants.

What are the ethical consequences of immortality technology?

Francesca Minerva and Adrian Rorheim

This article was first published on Aeon on August 9th 2017

Immortality has gone secular. Unhooked from the realm of [gods and angels](#), it's now the subject of serious investment – both intellectual and financial – by philosophers, scientists and the Silicon Valley set. Several hundred people have already chosen to be [‘cryopreserved’](#) in preference to simply dying, as they wait for science to catch up and give them a second shot at life. But if we treat

death as a problem, what are the ethical implications of the highly speculative ‘solutions’ being mooted?

Of course, we don't currently have the means of achieving human immortality, nor is it clear that we ever will. But two hypothetical options have so far attracted the most interest and attention: rejuvenation technology, and mind uploading.

Like a futuristic fountain of youth, [rejuvenation](#) promises to remove and reverse the damage of ageing at the cellular level. Gerontologists such as Aubrey de Grey argue that growing old is a disease that we

can circumvent by having our cells replaced or repaired at regular intervals. Practically speaking, this might mean that every few years, you would visit a rejuvenation clinic. Doctors would not only remove infected, cancerous or otherwise unhealthy cells, but also induce healthy ones to regenerate more effectively and remove accumulated waste products. This deep makeover would 'turn back the clock' on your body, leaving you physiologically younger than your actual age. You would, however, remain just as vulnerable to death from acute trauma – that is, from injury and poisoning, whether accidental or not – as you were before.

Rejuvenation seems like a fairly low-risk solution, since it essentially extends and improves your body's inherent ability to take care of itself. But if you truly wanted eternal life in a biological body, it would have to be an extremely secure life indeed. You'd need to avoid any risk of physical harm to have your one shot at eternity, making you [among the most anxious people in history](#).

The other option would be mind uploading, in which your brain is digitally scanned and copied onto a computer. This method presupposes that consciousness is akin to software running on some kind of organic hard-disk – that what makes you *you* is the sum total of the information stored in the brain's operations, and therefore it should be possible to migrate the self onto a different physical substrate or platform. This remains a highly [controversial](#) stance. However, let's leave aside for now the question of where 'you' really reside, and play with the idea that it might be possible to replicate the brain in digital form one day.

Unlike rejuvenation, [mind uploading](#) could actually offer something tantalisingly close to true immortality. Just as we currently back up files on external drives and cloud storage, your uploaded mind could be copied innumerable times and backed up in secure locations, making it extremely unlikely that any natural or man-made disaster could destroy all of your copies.

Despite this advantage, mind uploading presents some difficult ethical issues. Some

philosophers, such as David Chalmers, [think](#) there is a possibility that your upload would appear functionally identical to your old self without having any conscious experience of the world. You'd be more of a zombie than a person, let alone *you*. Others, such as Daniel Dennett, have [argued](#) that this would not be a problem. Since you are reducible to the processes and content of your brain, a functionally identical copy of it – no matter the substrate on which it runs – could not possibly yield anything other than *you*.

What's more, we cannot predict what the actual upload would feel like to the mind being transferred. Would you experience some sort of intermediate break after the transfer, or something else altogether? What if the whole process, including your very existence as a digital being, is so qualitatively different from biological existence as to make you utterly terrified or even catatonic? If so, what if you can't communicate to outsiders or switch yourself off? In this case, your immortality would amount to more of a curse than a blessing. Death might not be so bad after all, but unfortunately it might no longer be an option.

Another problem arises with the prospect of copying your uploaded mind and running the copy simultaneously with the original. One popular position in philosophy is that the *youness* of you depends on remaining a *singular* person – meaning that a 'fission' of your identity would be equivalent to death. That is to say: if you were to branch into you₁ and you₂, then you'd cease to exist as *you*, leaving you dead to all intents and purposes. Some thinkers, such as the late Derek Parfit, have argued that while *you* might not survive fission, as long as each new version of you has an unbroken connection to the original, this is just as good as ordinary survival.

Which option is more ethically fraught? In our view, 'mere' rejuvenation would probably be a less problematic choice. Yes, vanquishing death for the entire human species would greatly exacerbate our existing problems of overpopulation and inequality – but the problems would at least be reasonably

familiar. We can be pretty certain, for instance, that rejuvenation would [widen the gap](#) between the rich and poor, and would eventually force us to make decisive calls about resource use, whether to limit the rate of growth of the population, and so forth.

On the other hand, mind uploading would open up a plethora of completely new and unfamiliar ethical quandaries. Uploaded minds might constitute a radically [new sphere of moral agency](#). For example, we often consider cognitive capacities to be relevant to an agent's moral status (one reason that we attribute a higher moral status to humans than to mosquitoes). But it would be difficult to grasp the cognitive capacities of minds that can be enhanced by faster computers and communicate with each other at the speed of light, since this would make them incomparably smarter than the smartest biological human. As the economist Robin Hanson argued in *The Age of Em* (2016), we would therefore need to find fair ways of regulating the interactions between and within the old and new domains – that is, between humans and brain uploads, and between the uploads themselves. What's more, the

astonishingly rapid development of digital systems means that we might have very little time to decide how to implement even minimal regulations.

What about the personal, practical consequences of your choice of immortality? Assuming you somehow make it to a future in which rejuvenation and brain uploading are available, your decision seems to depend on how much risk – and what *kinds* of risks – you're willing to assume. Rejuvenation seems like the most business-as-usual option, although it threatens to make you even more protective of your fragile physical body. Uploading would make it much more difficult for your mind to be destroyed, at least in practical terms, but it's not clear whether you would survive in any meaningful sense if you were copied several times over. This is entirely uncharted territory with risks far worse than what you'd face with rejuvenation. Nevertheless, the prospect of being [freed from our mortal shackles](#) is undeniably alluring – and if it's ever an option, one way or another, many people will probably conclude that it outweighs the dangers.

Why philosophers fail to influence public debate – and how they can do better

Simon Beard and Michael Plant

This article was first published on Quillette on August 2nd 2017

We all know that philosophers are expert thinkers but most philosophers, and especially moral philosophers, want to change the world as well. As Plato noted, once one has ascended to the pinnacle of wisdom, or at least successfully defended a PhD thesis, it is hard to resist the temptation to come back down again and help to spread the light to others.

However, for most of us, the idea of actually succeeding at this is little more than a dream. Attempts to get heard often end up backfiring or simply proving a waste of time and energy. Even philosophers whose work is in areas of real public interest, such as applied ethics,

can struggle to get a hearing above the noise of pundits, preachers and politicians whose views, though ill-considered and even inconsistent, are far easier on the ear and offer people a sense of certainty in a baffling world.

At a recent workshop on [Personal Identity and Public Policy](#) held at Oxford, we considered what to do about this problem. Our shared interest was in what makes people – well, people. In particular, what makes me the same person when I am young as when I am old. The answer to this question is vital to many issues, from health care to criminal justice, emerging technologies to the diagnosis of death.

However, these are often issues on which people, including doctors, lawyers and scientists, have already made up their minds. So why should anyone care what philosophers think about them?

Whose problems are these anyway?

Here's an example. [Should people be punished for crimes they committed in the distant past?](#) It seems pretty obvious that we should only punish a person for a crime if we are reasonably convinced that they are the same person who committed that crime. However, on many views of personal identity, once enough time has passed between the commission of the offence and the punishment, then, even if the criminal is still alive, they will no longer be the same person that they were and so could not deserve punishment.

One critical issue for philosophers is that having considered a problem like this for many years, we tend to think about it in fundamentally different ways to people coming to it for the first time. For us, there is nothing wrong with asking questions such as 'is there really a moral distinction between punishing somebody many years after they have committed a crime, and punishing someone who never committed any crimes to begin with'. However, for many people such questions appear heretical at best and incomprehensible at worst.

Therefore, before we can hope to engage in genuine public debate, and still be taken seriously, we need to find ways of addressing problems that people actually have. To articulate views in a way that can have an impact it is necessary that they are located within an area of debate for which their relevance is clear and easy to understand. Furthermore, while philosophers like to deal with arguments and arguments alone, most non-philosophers deal mainly in conclusions. A view, no matter how well expressed and cogent it may be, whose implications are unclear or unacceptable to a mass audience may well be worth pursuing academically, but will not be of wider interest – at least not without a lot of hard work.

So, while, for philosophers, there is a simple matter of principle here, there is no chance of making any progress unless we recognise that the conclusion that rapists and murders should avoid being convicted of their crimes is probably a step too far. Best then to restrict

oneself, at least in the early stages, to cases in which our conclusions appear less outrageous – for instance to crimes that depended more upon the identity of the criminal to begin with, such as fraud or conspiracy.

Making friends – in high places

The next problem philosophers face is that, much as we hate to admit it, we don't have all the answers. Getting moral philosophy right is an important part of good decision making, but it is only one part. Public debates, however, tend to focus on a whole package, means, motivation and opportunity, and if philosophers cannot find ways of speaking to all these things our opinions will only ever play a marginal role.

A first question is whether philosophers tend to agree amongst themselves. Let's return to the issue of criminal responsibility. As I mentioned earlier, some scholars take the view that one is simply not the same person in one's old age as in one's youth, one is merely a 'successor self'. This view tends to be supported by those who believe that personal identity over time is a matter of 'psychological continuity', the degree to which our memories, intentions, beliefs, desires and personality traits vary over time. Given enough time almost all of us change psychologically, so philosophers who take this view find it easy to conclude that, in at least some cases, it is morally wrong to punish somebody for crimes from their distant past.

What of philosophers who do not share this view? The main alternative is 'animalism', the idea that personal identity consists in being the same biological organism over time. On this view, it is almost impossible, barring certain radical medical interventions, that somebody is not the same person in their old age as they were in their youth. However, many who take this view find, as a result, that personal identity over time is not so morally significant as we might think. Sure, punishing somebody for a historical offence is not the same thing as punishing an entirely different person, but why should their continuity as a biological organism matter to us when so

many other morally interesting facts about them, such as their personality and behaviour, might have changed? So at least amongst philosophers, there is broad agreement about the conclusion that we should often not punish people for historical offences, even though there is less agreement on why this is so.

Building bridges outside of philosophy can be more difficult. While philosophers often find common ground with certain other groups, such as psychologists, sociologists and even criminologists, others who may have more of an impact on public debate, like economists and lawyers, have very well-defined conventions and norms.

It is invariably easier to influence a debate when someone is already interested in what you have to say, and more people are going to be interested in what philosophers have to say if this can easily express it in terms that are relevant to them and carry clear implications for the kinds of decisions they face. Often, it is only in trying to engage others that we find people who are prepared to listen and, through talking with them, find better ways to tell them what you have to say.

Getting one's hands dirty

So, if philosophy is to live the dream of influencing public debate then philosophers

must think a lot more about what we are saying, how we are saying it, who we are saying it to and why they might care. However, there is still one more thing that needs to be done, the hard bit, actually getting out there and saying it. This leaves philosophers with probably the biggest problem of all, where to start.

Is it better to write a book and become the 'go-to academic' on an issue in the hope that people will come and ask you about it, to talk to relevant policymakers and find out what they most want to hear or to take to the streets and shout at the top of one's voice? Of course, this is not a question that can be answered once and for all. However, one useful proposal is often to find those people who one can influence most easily and who carry most influence over others. If this is a well-informed general public, then write a book, or better yet a series of blog posts. If it is a small group of specialist policy makers then go to them directly – sometimes it can be surprising how interested they can be (especially if they took a class or two in philosophy at university). Finally, however, if what one has to say is too big and too important to be left for others to help communicate it, then it's probably time to get behind those barricades.

Anyone care to join us?

List of Papers and Presenters

Francesca Minerva (University of Ghent) – Personal identity and Future Technologies

Tom Douglas (Uehiro Centre for Practical Ethics) – Parfitian Survival and Punishing Crimes from the Distant Past

Jeff McMahan (University of Oxford) – Personal Identity and Allocation of Scarce Health Resources

Jeremy Williams (University of Birmingham) – Death, Personal Identity, and Public Justification

Hannah Tierney (Cornell University) – The Problem of the Beginning and the End

Max Suffis (Rice University) – Indexicality and Personal Identity

Timothy Campbell (Institute for Futures Studies) – Don't Lose Your Head: A Moral Assessment of Whole-Body Transplantation

Simon Beard (Centre for the Study of Existential Risk) – Personal Identity and Quality of Life

Chase Bednarz (Northwestern University) – Practical Causal Set Modelling for Time Relative Personal Identity

Other workshop participants:

- Wessel van Dommelen (University of Groningen)
- Patrick Kaczmarek (University of Oxford)
- Alberto Giubilini (Uehero Centre for Practical Ethics)
- Cullum Miller (NHS)
- Joshua Parikh (University of Oxford)
- Michael Plant (University of Oxford)
- Huw Price (University of Cambridge)
- Anders Sandberg (Future of Humanity Institute and Institute for Futures Studies)
- Christophe Salvat (ENS Lyons)
- Teruji Thomas (University of Oxford)

This workshop was sponsored by the Centre for the Study of Existential Risk and is part for the Templeton World Charity Foundation funded project ‘Evaluating Extreme Technological Risks’



The workshop took place on November 1st the November 2nd 2016 and was kindly hosted by Corpus Christie College and the Future of Humanity Institute, both at the University of Oxford

