

"The point is to change it"

To the shoemakers and the ship-builders

John Tasioulas, University of Oxford

To the shoemakers and the ship-builders

John Tasioulas, University of Oxford

What is the contribution that the arts and humanities, including philosophy, can make to public engagement with the increasingly pervasive technology of artificial intelligence (AI)?

Choice

Perhaps the most fundamental contribution of the arts and humanities is to make vivid the fact that the development and deployment of AI technology is not a matter of destiny, but instead involves successive waves of highly consequential human choices. It is important to identify the choices, to frame them in the right way, and to raise the question: who gets to make them and how?

This is important because AI, and digital technology generally, has become the latest focus of the myth that social evolution is determined by independent variables over which we, as individuals or societies, are able to exert little control. A recent book by a leading industry figure employs the metaphor of a dehumanised force of nature to characterize the AI revolution — ‘a coming wave’ obeying a ‘a single, seemingly immutable law’, according to which a new technology ‘gets cheaper and easier to use, and ultimately it proliferates, far and wide.’ (Suleyman, 2023).

The humanities are vital to combatting this techno-determinism, which is profoundly disempowering for individuals and democratic publics alike. They can do so by reminding us of other technological developments that arose the day before yesterday — such as the

harnessing of steam or nuclear power — and how their development and deployment for good or ill were always contingent on human choices, and therefore hostage to systems of value and to power structures that could have been otherwise (Acemoglu and Johnson, 2023).

Ethics

Given the necessity for choice, the second contribution the arts and humanities can make is to emphasise the inescapability of ethics in articulating and making these choices.

Ethics is inescapable because it concerns the ultimate values in which our choices are anchored, whether we realise it or not. These are values that define what it is to have a good life, and what we morally owe to others, including non-human animals and to nature. The arts and humanities in general, and not just philosophy, engage directly with ethics — the ultimate ends of human life. And, in the context of AI, it is vital for them to resist various worrying contractions and distortions that the notion of ethics is apt to undergo. Thanks in part to the incursion of big tech into the AI ethics space, ‘ethics’ is often interpreted in unduly limited ways. For example, as a form of self-regulation lacking legal enforceability. It would be a hugely damaging capitulation to the distortions wrought by big tech to adopt its anaemic understanding of ethics as — at best — essentially self-regulation, or — at worst — corporate PR.

All forms of ‘regulation’ that might be proposed for AI, whether one’s self-regulation in deciding whether to use a social robot to keep one’s aged mother company, or the content of the social and legal norms that should govern the use of such robots, ultimately implicate choices that reflect ethical judgements about salient values and their prioritisation. Reclaiming a broad and foundational understanding of ethics in the AI domain, with radical implications for the re-ordering of social power, is an important task of the arts and humanities.

The dominant approach

The next question we might ask is: what is the shape of the ethical self-understanding that the arts and humanities can help to generate? The starting-point, I think, is to recognise that there is already a dominant approach in this area, that it has grave deficiencies, and that a key task for the humanities is to help us elaborate a more robust ethical outlook. I take the dominant approach to be that which is found most congenial by the powerful scientific, economic and governmental actors in this field.

Like anyone else, AI scientists are prone to the illusion that the intellectual tools at their disposal have a far greater problem-solving purchase than is actually warranted. This is a phenomenon that Plato diagnosed long ago with respect to the technical experts of his day, such as shoemakers and ship-builders. The mindset of scientists working in AI tends to be data-driven, it places great emphasis on optimisation as the core operation of rationality, and it prioritises formal and quantitative techniques.

Given that intellectual framework, it is little wonder that a leading AI scientist like Stuart Russell finds himself drawn to a preference-based utilitarianism as his overarching framework of ethics (Russell, 2019). According to Russell, the morally right thing to do is that which will maximise the fulfilment of human preferences. So, ethics is reduced to a technical exercise in prediction and optimisation — deciding which act or policy is likely to lead to the optimal fulfilment of human preferences.

But this view of ethics is, of course, open to serious challenge. Its concern with aggregating preferences threatens to override important rights that erect strong barriers to what can be done to individuals. And that’s even before we observe that human preferences may themselves be infected with racist, sexist or other prejudices. Ethics operates in the crucial space of reflection on what our preferences should be, a vital consideration that makes a belated appearance in the last few pages of Russell’s book. It does not take those preferences as ultimate determinants of value.

This crude, preference-based utilitarianism also exerts considerable power as an ideology among leading economic and governmental actors. This is less easy to see, because the doctrine has been modified by positing wealth-maximisation as the more readily measurable proxy for preference-satisfaction. Hence the tendency of GDP to hijack

governmental decision-making around economically consequential technologies such as AI, with the consequent sidelining of values that are not readily captured by market demand. Hence, also, the legitimization of profit-maximisation by corporations as the most effective institutional means to societal wealth-maximisation.

The three Ps: pluralism, processes and participation

The kind of ethics we should hope the arts and humanities steer us towards is one that ameliorates and transcends the limitations of this dominant paradigm. This humanistic ethics would have at least the following three features: the three Ps.

Pluralism

This humanistic ethics would emphasise the irreducible plurality of values, both in terms of the elements of human well-being and the core components of morality. This pluralism calls into question the availability of some optimising function for determining what is all-things-considered the right thing to do. It also undermines the facile assumption that the key to the ethics of AI will be found in one single master-concept, whether that be safety, trustworthiness, human rights or something else. How could human rights be the overarching framework for AI ethics when, for example, AI has a serious environmental impact that cannot be exclusively cashed out in terms of its bearing on anthropocentric concerns? And what about those human values to which we do not think of ourselves as having a right but which are nonetheless important, such as mercy, kindness, or solidarity? Nor can trustworthiness be the master value. It is at best parasitic on compliance with more basic values, hence it cannot displace the need to investigate those values. Equally, safety is an impoverished comprehensive rubric for AI ethics, since the concerns of the latter go well beyond protecting life and limb, even if highly speculative worries about the 'existential risk' to humanity posed by AI systems have tended to draw attention away from this fact.

Acknowledging the existence of a plurality of values, with their nuanced relations and messy conflicts, heightens the need for choice, and accentuates the question of whose decision will prevail. This sensitive exploration of a plurality of values and their interactions is what the arts and humanities, at their best, offer us. I say at their best because, of course, they often fail in this task. My own discipline, philosophy, has itself in recent years often propagated the

highly systematising and formal approach to ethics that I have condemned. Philosophers have much to learn from closer engagement with other humanities disciplines, like classics and history, and with the arts, especially fiction, which often penetrate to the heart of issues like the significance of personal relations, or the nature of human emotion, in ways that the more discursive methods of philosophy cannot. Exemplary here, of course, has been the work of philosophers like Martha Nussbaum, Charles Taylor, and Bernard Williams.

Processes not just outcomes

Of course, we want AI to achieve valuable social goals, such as improving access to education, justice and health care, and to do so in an effective and efficient way. The COVID-19 pandemic cast into sharp relief the question of which outcomes AI is being used to pursue. Did it help us, for example, reduce the need for our fellow citizens to undertake hazardous labour in the delivery of vital services, or is it primarily engaged in profit-generating activities, like vacuuming up people’s attention online and encouraging crass consumerism, that have little redeeming social value? The second feature of a humanistic approach to ethics is to underline that what we rightly care about is not just the value of the outcomes that AI can deliver, but the processes through which it does so.

Compare the use of AI in cancer diagnosis and its use in the sentencing of criminals. When it comes to cancer, what may be all-important is getting the most accurate diagnosis, and it is largely a matter of indifference whether this comes through the use of an AI diagnostic tool or the exercise of human judgement. In criminal sentencing, however, there is a powerful intuition that being sentenced by the robot judge — even if the sentence is likely to be less biased or more consistent than one rendered by a human counterpart — means sacrificing important values, such as accountability and empathy, relating to the process of decision.

This point is familiar, of course, in relation to such process values as transparency, procedural fairness and explainability. But it goes even deeper, because of the dread many understandably feel when contemplating a dehumanised world in which decisions that affect our deepest interests have, at least as their proximate makers, automated systems that can neither be held directly accountable nor empathise with the plight of the humans subject to their decisions (Campbell, 2020; Tasioulas, 2023).

Participation

The third feature relates to the importance of participation, whether as an individual or as part of a group of self-governing democratic citizens, in realising our ethical values. At the level of individual well-being, this takes the focus away from theories that equate human well-being with some end-state, such as pleasure or preference-satisfaction, that could in principle be secured by a process in which the person who enjoys them is entirely passive, for example, by pumping vast quantities of an antidepressant drug in the water supply. Contrary to this view, the exercise of our distinctive human faculties for reason, communication and social connection lies at the core of human wellbeing. This has immense relevance for assessing the significance of the increased deployment of AI systems in place of human endeavour.

One of the most important sites of participation in constructing a good life, in modern societies, is the workplace. According to a McKinsey study, around 30% of all work activities in 60% of occupations are capable of being automated (McKinsey Global Institute, 2017). Should we accept the idea that the large-scale elimination of job opportunities, due to automation, can be compensated for by the benefits that automation brings? The answer depends on whether the participatory self-fulfilment of work can, any time soon, be feasibly replaced by other activities, such as art, friendship, play or religion. If it cannot, addressing the problem with a mechanism like universal basic income, which involves the passive receipt of a benefit, will not be enough (Tasioulas, 2024).

Similarly, we value citizen participation as part of democratic self-government. And, arguably, we do so not just because of the instrumental benefits of democratic decision-making in reaching better decisions (‘the wisdom of crowds’ factor), including better decisions regarding the fair distribution of the benefits of new technologies. The need for democratic control, and the diffusion of power more generally, to ensure the fair distribution of the wealth resultant from new technologies is an important theme in Acemoglu and Johnson’s *Power and Progress*. But we also value participatory decision-making processes because they affirm the status of citizens as free and equal members of the community. This is an essential plank in the defence against the tendency of AI to be co-opted by technocratic modes of decision-making. These modes can erode democratic values by seeking to convert matters of political judgement into questions of purely technical

expertise. In this process the rhetoric of safety, and existential risk, has played a significant role in constraining democratic control over AI by encouraging an expert-based discourse that sidelines the experience and perspectives of ordinary citizens.

At present, much of the culture in which AI is embedded is distinctly technocratic, with decisions regarding the ‘values’ encoded in AI applications being taken by elites within the corporate or bureaucratic sectors, often largely shielded from vigorous democratic control. Indeed, a small group of tech giants accounts for the lion’s share of investment in AI research, dictating its overall direction. Meanwhile, AI-enabled social media poses risks to the quality of public deliberation by promoting the spread of disinformation, aggravating political polarisation, and so on. Similarly, the use of AI as part of corporate and governmental attempts to monitor and manipulate individuals undermines privacy and threatens the exercise of basic liberties, effectively discouraging citizen participation in democratic politics. On the positive side, we need to think seriously about how AI and digital technology more generally can enable, rather than hinder and distort, democratic participation, not just in the sphere of formal law-making, but also in corporate governance and beyond (Landemore, 2024).

Democratising the arts and humanities — The Lyceum Project

If the arts and humanities are to advance a humanistic ethical agenda in AI of the kind I have described, they must themselves be democratised. In a democracy, it is not enough to give people a vote while effectively excluding them from meaningful deliberation; and if they are to deliberate as equals, they must have access to the key sites where basic ideas about justice and the good are worked out.

The arts and humanities are prominent among those sites. Hence the wisdom of Article 27 of the Universal Declaration of Human Rights, which includes a right to participation in science and culture. We can see manifestations of this right, enabled by digital technology, in the resurgent citizen science movement (Vayena and Tasioulas, 2015). But we must also address the exclusion of our fellow citizens from the domains of artistic creativity and humanistic enquiry. This means that work on AI within the arts and humanities should not merely be accessible to a wider public, nor should it merely model civil and rational debate. It should also afford our fellow citizens the opportunity to articulate and critically assess their own views in dialogue with others.

Lest all this sound unduly abstract, I should add something about one of the more notable attempts I have made, as Director of the University of Oxford’s Institute for Ethics in AI, to practice what I am preaching.

On 20 June 2024, in partnership with Stanford University, ‘Demokritos’, Greece’s National Centre for Scientific Research, and the World Human Forum, a Greek-based think tank, the institute held the Lyceum Project, a one-day public event: ‘AI ethics with Aristotle’. The event took place at the Athens Conservatory, a stone’s throw from the original site of Aristotle’s school, the Lyceum. Around 500 people attended, from different walks of life, to listen to talks and panel discussions featuring both early-career and established philosophers, entrepreneurs, policymakers, computer scientists, and a speech by the Greek prime minister, Kiriakos Mitsotakis. As part of the event, the ancient historian Josiah Ober and I co-wrote a white paper on how an Aristotelian framework affords the best approach to the ethics of AI, illuminating such matters as work, democracy and global tech regulation (Ober and Tasioulas, 2024). The evening ended with a youth dialogue on the hallowed site of the ancient Lyceum that was followed by an inspiring performance of Sasha Waltz’s ‘In C’ by the Conservatory’s graduating class of dancers. These two concluding events served as powerful reminders that the future is in the hands of the younger generation and that humans have capacities for grace, beauty, and emotion far beyond any automated system (ox.ac.uk, 2024).

This event exemplified, for me, how philosophers might devise formats that foster inclusive and rigorous engagement with the largest questions confronting democratic publics. And it was gratifying to see that the Greek prime minister’s High-Level Advisory Committee on Artificial Intelligence, on which I was privileged to serve alongside another philosopher, Vassio Kindi, subsequently recommended an annual global AI ethics summit to be held in Athens, one drawing on the rich traditions of philosophical inquiry (Special Secretariat of Foresight, 2024). In a world in which a small number of powerful actors increasingly dominate the public discourse on questions of great human significance, it is important for philosophers to seize such opportunities for different voices to be heard.

This chapter is a shortened and updated version of the following blog post: John Tasioulas, ‘The role of the arts and humanities in thinking about artificial intelligence (AI)’, <https://www.adalovelaceinstitute.org/blog/role-arts-humanities-thinking-artificial-intelligence-ai/>

Notes:

- 1 Mustafa Suyleman, *The Coming Wave: AI, Power and the 21st Century's Greatest Dilemma* (Bodley Head, 2023), p.6.

References:

- Acemoglu, D. and Johnson, S. (2023). *Power and Progress: Our Thousand-Year Struggle over Technology and Prosperity*. London: Basic Books.
- Campbell, J. (2020). *Causation in Psychology*. Harvard University Press, ch.3.
- Landmore, H. (2024). 'Can Artificial Intelligence Bring Deliberation to the Masses?', in R. Chang and A. Srinivasan (eds), *Conversations in Philosophy, Law, and Politics*. Oxford: Oxford University Press.
- McKinsey Global Institute (2017). *A future that works: Automation, employment, and productivity* | McKinsey. [online] [www.mckinsey.com](https://www.mckinsey.com/featured-insights/digital-disruption/harnessing-automation-for-a-future-that-works/de-de). Available at: <https://www.mckinsey.com/featured-insights/digital-disruption/harnessing-automation-for-a-future-that-works/de-de>.
- Ober, J. and Tasioulas, J. (2024). 'The Lyceum Project: AI Ethics with Aristotle'. *SSRN Electronic Journal*. doi: <https://doi.org/10.2139/ssrn.4879572>.
- Ox.ac.uk. (2024). *The Lyceum Project: AI Ethics with Aristotle* | Ethics in AI. [online] Available at: <https://www.oxford-aiethics.ox.ac.uk/lyceum-project-ai-ethics-aristotle-O>.
- Russell, S. (2019). *Human Compatible: AI and the Problem of Control*. London: Allen Lane.
- Special Secretariat of Foresight. (2024). A Blueprint for Greece's AI Transformation - Special Secretariat of Foresight. [online] Available at: <https://foresight.gov.gr/en/studies/A-Blueprint-for-Greece-s-AI-Transformation/>.
- Suyleman, M. (2023). *The Coming Wave: AI, Power and the 21st Century's Greatest Dilemma*. London: The Bodley Head.
- Tasioulas, J. (2023). 'The Rule of Algorithm and the Rule of Law', in C. Bezemek, M. Potacs, A. Somek (eds), *Vienna Lectures on Legal Philosophy vol 3: Legal Reasoning*. Oxford: Hart Publishing, pp.17-39.
- Tasioulas, J. (2024). 'Work and Play in the Shadow of AI', in D. Edmonds (ed.), *AI Morality*. Oxford: Oxford University Press.
- Vayena, E. and Tasioulas, J. (2015). 'We the Scientists': a Human Right to Citizen Science. *Philosophy & Technology* 28, pp. 479-485.

Section 3: Interventions

Philosophy protects the climate

Professor John Broome, University of Oxford

The influence of philosophy

Philosophy has an immense influence on public life. As an example, think of the influence of liberalism in politics. Liberalism is largely the creation of philosophers, including particularly John Locke and John Stuart Mill, and it is now the guiding principle of many countries' political systems. However, philosophy generally works its influence slowly, as ideas percolate from philosophical writing out into society at large.

It also sometimes happens that important public issues call directly for philosophical analysis. They cannot wait for the slow percolation of thought. For instance, in the 1980s advances in the technology of human reproduction demanded regulation. It was recognised that moral philosophy had an important contribution to make in determining the regulations. The philosopher Mary Warnock oversaw an official inquiry and report which led to the establishment of new UK legislation (Warnock, 1984). Her work was also influential elsewhere in the world.

Another example is climate change. Governments must decide how to respond to it. The need for action is urgent. Despite international efforts that have continued for more than three decades, emissions of greenhouse gas are still increasing. We are on track for warming of about three degrees by the end of this century, which is expected to be catastrophic. This is an emergency. Decisions need urgently to be made about how to drive down emissions. And good decision-making about climate change demands a philosophical contribution. Philosophy can help combat this most serious threat to humanity's future.

What has philosophy to do with climate change? One answer is that climate change itself is immoral in one respect. Some people in the world — mostly in rich countries — cause greenhouse gas to be emitted for their own benefit, and this does harm to other people — mostly in poor countries — who themselves contribute little to climate change. This is an injustice done by the rich to the poor, which is a sort of moral wrong. Wherever morality is at issue, there is a role for moral philosophy. Philosophical questions arise about the injustice of climate change. For example, how far is a country responsible for recompensing those who suffer from the emissions it released in the distant past?

I shall concentrate on a different contribution that philosophy can make. Decision-making about climate change involves judgements of value, and value is within the domain of moral philosophy. These judgements are generally made in practice using the tools of economics, by means of cost–benefit analysis broadly construed. For example, the *Stern Review* (Stern, 2007) — a major UK government report on the economics of climate change — estimated the benefits to be gained in 200 years from limiting climate change, and compared them with the present costs of limiting it. Cost–benefit analysis is a matter of valuation: the value — goodness — of benefits is weighed against the value — badness — of costs.

Philosophy allied with economics

Climate change affects all the world’s population over centuries. It requires methods of analysis that can cope with such numbers and such a long time. Economics possesses those methods, whereas moral philosophy has traditionally been concerned with the actions of individuals and relations between small numbers of people. But moral philosophy does supply underlying principles of valuation. Economics can apply these principles to the enormous problem of climate change.

The branch of economics that is concerned with value is known, oddly, as ‘welfare economics’. It encompasses cost–benefit analysis. Welfare economics is in effect applied ethics: moral philosophy applied to economic matters.

True, some economists believe their discipline to be independent of ethics. This belief was revealed in the reactions of some economists to the *Stern Review*. The *Stern Review* explicitly recognised that welfare economics rests on ethical assumptions. William Nordhaus

and Martin Weitzman — the two most distinguished US economists of climate change — each wrote a review of the *Stern Review*, and each denied that economics rests on ethics (Nordhaus and Weitzman, 2007). These reviewers insisted that values should be taken, not from ethics, but from the preferences of the public as they are revealed in markets.

This view about the source of value is an ethical theory. Some English speakers think that they speak their language without an accent, and only other speakers have an accent. They are so deeply imbued in their own accent that they do not recognise it as an accent at all. Nordhaus and Weitzman are so deeply imbued in their own ethical theory that they do not recognise it as an ethical theory at all. They accept it without thought, and suppose that anyone who takes the trouble to think about ethics is arrogantly trying to impose her own view on other people. Nordhaus wrote that the *Stern Review* 'takes the lofty vantage point of the world social planner, perhaps stoking the dying embers of the British Empire'. Weitzman accused the *Review* of 'relying mostly on a priori philosopher-king ethical judgements about the immorality of treating future generations differently from the current generation'.

These authors' own arrogant position would be more defensible if their ethical theory were a good one. But it is not, at least in the context of their two reviews. This context is a discussion of 'discounting', which is the practice of giving less value to future goods than to present goods. Nordhaus and Weitzman recommended deriving the rate at which future goods are discounted from the rate of interest in present financial markets. But the preferences of future generations are not represented in present markets. Even if it were right in general to derive values from people's preferences, it would still not be right to derive them from the preferences of only some of the affected people — the present generation — and ignore the preferences of the rest.

The idea that economics is independent of ethics must be rejected. It remains prevalent, especially in the US, but many economists side with Nicholas Stern in rejecting it. Philosophers who wish to influence the world's response to climate change would do well to ally themselves with these economists. Economics has traditionally been influential in policymaking. Governments rarely call for advice from philosophers, but they regularly call on their economic advisers. Moreover, economic theory offers crucial insights that must be recognised by policymakers if climate change is to be overcome. If philosophers can ensure that governments' economic advisers give good advice, properly based on principles of ethics as well as on economic theory, they will have made a major contribution. An alliance of philosophers and economists can together deliver good advice to governments.

An intellectual alliance already exists. The interface between economics and ethics is a well-populated field. Economists and moral philosophers have many interests in common. They include foundational questions about value: What is the source of value? Is it preferences? How do different values combine? Should different people's well-being combine additively as utilitarians suppose? Should the worse off have priority? And so on. They also include questions about particular values: What is bad about inequality, if anything? How should future goods be weighed against present goods? What value should be attached to changes in the world's population? How bad is a person's death? And so on. Many of these questions are directly relevant to climate policy.

In trying to answer them, philosophers and economists have always interacted. Indeed, they have sometimes been one and the same person; remember Adam Smith, John Stuart Mill and Henry Sidgwick. When the interaction breaks down, things go badly. When philosophers began to make the distinction between egalitarianism and prioritarianism in the 1980s and 1990s, they did not realize that the same distinction had been investigated by economists in the 1960s and 1970s, using different terminology. As a result, philosophers wasted effort in rediscovering what was already known.

The philosophers' ignorance arose partly because many of them could not read economists' writings. They did not understand their pervasive mathematical notation. Interacting with economists is demanding for a philosopher; it requires an investment in understanding their technical language and methods. But in value theory the investment is well worthwhile. Indeed, in the more quantitative areas of value theory it is essential.

The value of human life

Bad economics is a constant threat to good public decision making. The valuing of human life is an excellent example. This is a crucial issue for climate change, since it is widely agreed that the greatest harm climate change will do is to shorten people's lives. The US Environmental Protection Agency (US EPA) has recently produced a new estimate of the 'social cost of carbon', which is supposed to measure the harm done by emissions of greenhouse gas (US EPA, 2023). It estimates 'mortality costs' — the harm of killing people — as more than half of all the harm.

Economists have a way of setting a value on human life that conforms to their general ethical assumption that value should be based on people's preferences. They base the value of life on what people are willing to pay to reduce their risk of dying. Poor people are willing to pay less to reduce their risk than rich people are. Consequently, many economists value the lives of people who live in poor countries much less than the lives of those who live in rich countries. The US EPA does the same.

This is foolish. The amount of money someone is willing to pay to reduce risk to her life depends on two things: first, on the value she attaches to her life, and, second, on the value she attaches to money. Poor people attach a greater value to money than do rich people because they have a greater need for the things money can buy. This is the reason poor people are willing to pay less money for reducing risk to their lives. It is not because their lives have less value to them.

Money is not a good measure of value unless adjustments are made to correct for the differing values money has to different people. These adjustments are called 'distributional weights'. They can be made, but bad economics does not make them. The US EPA's valuation of life is bad economics.

Indeed, it is incompetent. Measuring value by means of people's willingness to pay, unadjusted, is known as the 'Kaldor-Hicks criterion' (Kaldor, 1939; Hicks 1939), and the US EPA explicitly uses this criterion. But it has been known for more than 70 years that the Kaldor-Hicks criterion leads to contradictory implications (de Scitovszky, 1941; Gorman, 1955; Blackorby and Donaldson, 1990; Broome, 2024). It is refuted by *reductio ad absurdum*, that is to say.

It is also immoral. In forming its response to climate change, the US EPA values the lives of Americans 35 times higher than the lives of Bangladeshis. Yet Americans on average emit 11 times as much greenhouse gas as Bangladeshis do. It is a licence for the rich to kill the poor.

Philosophy is not required to identify this fault in the US EPA's economics. All economists should have learned about it during their education. But actually, welfare economics is little taught in economics departments these days. The Kaldor-Hicks criterion, though refuted 70 years ago, is still regularly applied in cost-benefit analysis for policymaking. Contact with philosophy might help economists recognise the logic

of reductio ad absurdum. But the real role for philosophy in improving the economics of valuing lives is to think about what, actually, a person loses when she dies.

This has been a topic within philosophy since antiquity. The answer cannot safely be left to the preferences about risk that people happen to have, since those preferences are rarely thought out carefully. There is a group of economists who reject the Kaldor-Hicks criterion, and take the view that what a person loses when she dies is the rest of her life. They value this on the basis of how long and how good the rest of her life would have been. This value is codified by health economists in the form of quality-adjusted life years or qalys. The World Health Organization uses disability-adjusted life years or dalys. Practical measures of value such as these are very amenable to philosophical analysis.

A generalisation of qalys and dalys is to value a person's life by the total of well-being the person enjoys during her life, which is to say the well-being she has at each time, added up over her life. This could be called the total of 'well-being-adjusted life years'. Adding up this value across people implies that the harm done by climate change is the total of well-being it takes away from the world's population. This is a utilitarian measure of value.

Population ethics

Measuring value this way raises a problem that economists of climate change have so far entirely missed. They recognise that mitigating climate change saves lives. They forget it also affects births: it alters the number of people who are born and also the identities of the people who are born. This is because action to mitigate climate change alters people's lives significantly. Consequently, it affects how many children people choose to have, when they choose to have them, and whom they choose to have them with.

An action that mitigates climate change extends the lives of some people. It also causes some people to live who otherwise would not have lived at all. If value is measured by the total of people's well-being, extending lives counts as a benefit equal to the total of well-being it adds to existing lives. Creating new lives counts as a benefit equal to the total of well-being contained in the added lives. Action to mitigate climate change also causes some people not to live who otherwise would have lived, and this counts as a harm equal to the total of well-being those people would have enjoyed had they lived.

This is intuitively odd. Well-being can be added to the world in two ways. One is by extending the life of an existing person; the other by adding a person to the population, who then enjoys well-being. Measuring value by the total of well-being treats either of these ways as equally good. But they are not intuitively equally good. Indeed, many people’s intuition is that adding people to the population is not good at all, but ethically neutral, even if their lives will be good. The philosopher Jan Narveson wrote: ‘We are in favour of making people happy, but neutral about making happy people’ (Nareveson, 1973).

This issue has taken us into the realm of population ethics. It is not surprising that the question of valuing lives should force us to consider the value of creating lives as well as the value of extending lives. Since climate change will change the world’s population, good decision making about climate change must take population ethics into account. This is a place where philosophy can make a vital contribution. Once more, it is most likely to have influence if it works through economics. Indeed, some of the leading contributors to population ethics are the economists Charles Blackorby, David Donaldson and Walter Bossert.

The IPCC

My own experience of working with economists has been good. It helps that I was a professor of economics before I became a professor of philosophy. I left economics in 1995 but was drawn back to working with economists again a decade later when I returned to work on the ethics, and hence the economics, of climate change. I was a lead author of the 2014 assessment report of the Intergovernmental Panel on Climate Change (IPCC). I believe that my colleague Lukas Meyer and I are the only philosophers who have ever been lead authors for the IPCC.

Each IPCC report consists of three huge volumes — one from each of three working groups — together with several summaries. Each volume has a ‘Summary for Policymakers’, which is the most widely read and widely influential part of it. Meyer and I were in Working Group 3, which was concerned with how climate change can be mitigated. We joined about a dozen economists in writing a chapter entitled ‘Social, Economic and Ethical Concepts and Methods’ (IPCC, 2014). I found the economists much more tolerant of philosophy than I expected. Welfare economics has been little taught since the 1970s, but these economists at least recognised the contribution philosophy could make. So did Ottmar Edenhofer, the chair of the working group.

In the working group’s volume, I was given the space to describe aspects of value theory that underlie economic valuations of the harm done by climate change, and of the benefits that can be achieved by policies that limit climate change. I was able to discuss non-human values and cultural values, as well as human well-being, and how human well-being can be aggregated across time and across different people. I mentioned problems of population ethics and considered how lives should be valued. Meyer wrote about issues of justice including the question of how far a country should be held responsible for its emissions from long ago.

The Summary for Policymakers of this volume was written by a small subset of the authors, including me (IPCC, 2014). At the meetings of this group, I was with a few economists among many other social scientists and some lawyers. An affinity emerged between economics and philosophy as the more analytic disciplines among the rest. We wanted to write in clear, sharp sentences, stating our meaning exactly. Other contributors preferred to write long sentences containing many caveats and conditions, whose meaning was often not entirely determinate. I later saw the benefit of this vagueness. When we came to present the report to governments, it gave them less to disagree with and made it easier to cobble together a consensus. Still, it was striking how economics and philosophy converged on the aim of precision.

The Summary for Policymakers is highly compressed. Nonetheless it had enough space to state explicitly that climate change raises issues of justice and ethics. It states that climate policy can be aided by ethical analysis, and can take into account values of different sorts, including non-human values. It also makes it explicit that distributional weights should be applied to monetary measures of benefits and harm, to take account of the differing values money has to different people. These are not radical statements, but their presence is significant because each sentence in the Summary for Policymakers has been individually considered and approved by consensus at a meeting of the IPCC’s members. These members include the governments of every country in the world.

So, for instance, every country, including the US, has agreed that distributional weights should be applied to measures of benefits and harms. Consequently, when the US EPA declined to apply distributional weights in measuring the social cost of carbon, it failed to honour a national commitment. As it happens, the Office of Management and Budget — another arm of the US government — has recently given its permission for US agencies to use distributional weights (Office of Management and Budget, 2023).

Did all this work in cooperation with economists achieve any practical influence for philosophy? The IPCC's 2014 report provided the scientific background for the meeting of the UN Framework Convention on Climate Change that took place in Paris in 2015. That meeting led to the Paris Agreement, which is now the governing document of the international community's fight against climate change. It is the framework regulating each country's contribution. So the IPCC's report as a whole could scarcely have been more influential in international politics. Philosophy's small contribution to it is a foot in the door.

References:

- Blackorby, C. and Donaldson, D. (1990). A Review Article: The Case against the Use of the Sum of Compensating Variations in Cost-Benefit Analysis. *Canadian Journal of Economics*, [online] 23(3), pp.471–494. Available at: <https://ideas.repec.org/a/cje/issued/v23y1990i3p471-94.html>.
- Broome, J. (2024). 'The Value of Life in the Social Cost of Carbon: A Critique and a Proposal'. *Journal of Benefit-Cost Analysis*, pp.1–17. doi: <https://doi.org/10.1017/bca.2024.21>.
- de Scitovszky, T. (1941). 'A Note on Welfare Propositions in Economics'. *The Review of Economic Studies*, 9(1), pp.77–88. doi: <https://doi.org/10.2307/2967640>.
- Gorman, W.M. (1955). 'The Intransitivity of Certain Criteria Used in Welfare Economics'. *Oxford Economic Papers*, [online] 7(1), pp.25–35. doi: <https://doi.org/10.2307/2662000>.
- Hicks, J.R. (1939). The Foundations of Welfare Economics. *The Economic Journal*, [online] 49(196), pp.696–712. doi: <https://doi.org/10.2307/2225023>.
- IPCC (2014). *AR5 Climate Change 2014: Mitigation of Climate Change*. [online] [ipcc.ch](https://www.ipcc.ch/report/ar5/wg3/). Available at: <https://www.ipcc.ch/report/ar5/wg3/>.
- Kaldor, N. (1939). 'Welfare Propositions of Economics and Interpersonal Comparisons of Utility'. *The Economic Journal*, 49(195), p.549–552. doi: <https://doi.org/10.2307/2224835>.
- Narveson, J. (1973). 'Moral Problems of Population'. *The Monist*, [online] 57(1), pp.62–86. Available at: <https://www.jstor.org/stable/27902295>.
- Nordhaus, W.D. and Weitzman, M. (2007). 'A review of the Stern Review on the economics of climate change'. *Journal of Economic Literature*, 45, pp.686–702; 55, pp.703–24.
- Office of Management and Budget (2023). 'Circular A4: regulatory analysis', pp.65–7. Available at: <https://bidenwhitehouse.archives.gov/wp-content/uploads/2023/11/CircularA-4.pdf>.
- Stern, N. (2007). *The Economics of Climate Change*. Cambridge University Press. [online] doi: <https://doi.org/10.1017/cbo9780511817434>.
- US Environmental Protection Agency (2023). *Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances*. Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NCEE&dirEntryId=356357.
- Warnock, D.M. (1984). *Report of the Committee of Enquiry into Human Fertilisation and Embryology*. London, HMSO.

The pathology of the prison

Dr. Sarah Bufkin, University of Birmingham

In HMP Wandsworth, 80% of the men are locked into overcrowded cells, most for more than 22 hours a day. The wings are rife with broken furniture, insects and rat faeces. Some cells reach over 35 degrees Celsius in the summer. Staff absences mean that many men struggle to regularly access the showers, attend medical appointments, or make it to classes or the library. In May, Chief Inspector of Prisons Charlie Taylor noted that he had encountered in Wandsworth ‘a degree of despondency that I have not come across in my time as Chief Inspector’. Seven inmates have taken their own lives in the past year alone.

Wandsworth is by no means an outlier. Similar problems are rife across the UK’s overcrowded prison system, where Taylor has issued seven ‘urgent notifications’ to highlight substantial failures since November 2022. Education, work and resettlement programmes were cut to the bone during the pandemic and have since been neglected. More than half of men and a staggering four-fifths of women in state custody reported experiencing mental health issues in 2022–2023, and yet mental healthcare provision is profoundly inadequate, with people in crisis spending too long in poor conditions before transfer to hospital. In turn, violence is rife inside, with many prison staff failing to turn on their body cameras when using batons or wielding chemical PAVA spray. Assaults are two-thirds higher in men’s prisons than they were a decade ago, and rates of self-harm among incarcerated women are eight times that of incarcerated men, reaching record highs last year.

From reading media reports over the so-called ‘prison crisis’, it can be all too easy to assume that the UK prison and probation system is caught up in an exceptional moment — a perfect storm of post-pandemic, post-austerity conditions that have deformed what would otherwise be the normal and humane functioning of the carceral state. Renovate the Victorian buildings, crack down on drug use, build new prison beds, and His Majesty’s Prisons will be back to their stated purposes as a tool of criminal justice and rehabilitation.

Within these debates, the failures are presumed to lie at the level of performance, not with the institution of the prison itself or with the practice of incarceration as punishment.

Yet this reading of the prison crisis only provides a narrow window into the brutalities that characterise contemporary carceral state. By carceral state, I refer to the system of institutions, agencies, technologies, laws, and discourses that use repressive state power to manage and control 'dangerous' or 'disorderly' populations. Such a state apparatus tends to encompass both the sites that immobilise people (such as jails, prisons, detention centres, asylums, etc) and the broader systems that decide who is to be surveyed, detained, charged, and locked up (such as the police, prosecutors, the courts, border control forces, and even welfare state agencies). This chapter contends that when we treat prisons like Wandsworth as mere aberrations from broader norms of criminal justice, we miss the systemic role that the prison plays within the carceral state and its broader project of racialised and class-based population control. Such an idealised account of the prison can indict and condemn poor prison policy, but it can't explain or diagnose why we continue to incarcerate greater and greater numbers of people, even as crime rates have fallen. Nor does it have much to say about who is most likely to be intensively surveyed by police, arrested, remanded to prison, and sentenced to custodial sentences — and whose lawbreaking rarely brings them into contact with the carceral state.

This is where critical philosophy takes a different path from what we might call normative political philosophy. Instead of first identifying abstract values or principles and then applying them to various concrete situations, critical theorists embrace a practice of social critique — they work from actual instances of oppression to diagnose how power asymmetries produce these harms and to gesture towards the sorts of emancipatory practices that might overcome them. In this respect, social critique is problem-focused rather than principle-focused; it aims to understand the why and how of oppression, rather than simply provide a normative objection. And critical theorists acknowledge that building a more just future is rarely a straightforward task of simply revealing to people that their practices don't align with their values (or that they ought to be valuing differently). It is not that the world cannot be changed, but that the work of doing so is never a simple project of asking people to become better moral agents or states to become better stewards. Existing power hierarchies place real constraints, both material and ideological, on what we can do and become, even as we might work to move beyond those limits.

The practice of social criticism, of course, is not the sole provenance of philosophers or political theorists. Nor is it an exercise exclusively carried out within the institutional remit of the academy. Within the discipline of philosophy, however, this model of social critique is often associated with the Frankfurt School, which took shape in and around the Institute for Social Research in Frankfurt during the 1920s. Under Max Horkheimer, the institute adopted a problem-oriented, interdisciplinary research agenda that sought to disclose the distortions and dysfunctions that capitalism produced — including the rise of fascism in 1930s Europe and in the Fordist US. As Horkheimer wrote in a famous essay on what distinguished critical from traditional theorizing, ‘every part of the theory presupposes the critique of the existing order and the struggle against it along lines determined by the theory itself’.

For contemporary Frankfurt School thinkers like Axel Honneth, what makes this critical approach distinctive is its commitment to the diagnosis of social pathologies. Such a method is not uncontroversial. After all, the language of ‘pathology’ has all too often been used to blame racially and sexually minoritised communities for their own oppression. But many critical theorists argue that it is worth holding on to social pathology as a concept for two reasons. First, it provides a broader and more robust ethical toolkit through which to understand what can go wrong in a social formation. Instead of only focusing on individual wrongdoing or on structural inequalities, social pathology critique indicts specific forms of life and social practices as unethical, alienating, inhuman or otherwise dysfunctional. It reminds us that there are certain worlds that we shouldn’t accept, that corrode or otherwise distort what we are capable of as human animals and planetary co-inhabitants. And second, this mode of critique focuses on the ways in which some societal problems tend to reproduce themselves. A pathological social formation degenerates; its problems become more entrenched and harder to dislodge. (One could think here of the continuing demand for intensifying fossil fuel extraction in the face of a climate catastrophe that is already upon us.) Like many illnesses or malignancies, social pathologies can only be successfully ameliorated by being tackled at their roots.

It is my contention, within this short and polemical essay, that the prison system today is best understood today as a kind of social pathology. Rather than view prisons as the ‘legitimate’ response to ‘diseased’ or otherwise ‘deviant’ individuals, I argue that the modern prison regime itself constitutes a pathological social development that not only fails to solve, but further entrenches broader problems of inequality, racialised and engendered violence, precarious employment, state neglect, and predatory practices

of dispossession. In making this claim, however, I take only indirect inspiration from the Frankfurt School, whose thinkers did little to problematise the prison or the carceral state more broadly. (The notable exception here is Otto Kirchheimer and Georg Rusche’s 1939 book, *Punishment and Social Structure*.)

Instead, I look to US-based abolitionists like Angela Davis, Ruth Wilson Gilmore, and Mariame Kaba, who ask different questions about the police and the prison. Rather than query the idealised ends a prison ought to serve, they investigate what prisons actually do to the people incarcerated in them and to our communities. Much of this abolitionist activism and scholarship has focused on challenging the racialised rise of mass incarceration in the US. But in this essay, I argue that such a conversation can also be productively re-deployed to tackle the expansion of the UK prison system. For all that we must recognise Britain’s disparate historical trajectory, I contend that those committed to sentencing and penal reform in the UK could learn from US abolitionists. Scholar-activists like Davis and Kaba remind us to ask which communities are more likely to end up behind bars and why — and to dare us to imagine a society that doesn’t see the police and the prison as the solution for a wide range of social problems.

Abolitionist social critique first seeks to denaturalise the prison — to show that it is not a ‘normal’ and natural part of a society’s response to crime, but instead a malignant development that produces more harm than good. This requires breaking the presumed link between crime and punishment. In an ideal society, we might hold that a person who breaks the law should be proportionately punished for that crime as a form of restitution. But over the last few decades, both the US and the UK have locked up many more people, for longer and harsher sentences, even as crime rates have fallen. In turn, most people who break the law won’t come into contact with the criminal justice system. It is increased state surveillance that produces convicts, not the act of offending itself. Nor do prisons today function as sites of rehabilitation, notwithstanding government rhetoric. Deep cuts to education, skills and work programmes; the horrifying decline in prison conditions; the decimation of offender resettlement services; and a lack of support for the formerly incarcerated once they are back on the outside — all these shifts reveal that the prison has become primarily a tool of incapacitation, as Ruth Wilson Gilmore puts it. We cage people to punish them and to remove them from society, not to transform or empower them.

Critics of the carceral state also challenge the idea that prisons make us safer. After all, moving people into state custody does little to solve the problems that cause law breaking in the first place, such as poverty, insecure housing, structural unemployment, drug addiction, mental health crises, gendered and sexualised violence, and racist pogroms. In theory, the prison should serve as a deterrent, especially to those who have suffered through it before. And yet recidivism rates remain incredibly high. In fact, evidence suggests that locking people up not only doesn't prevent them from reoffending, but that it increases their chances of breaking the law. Such studies should come as no surprise when we consider what prisons do to people. Incarcerating someone is to separate them from their family and their community, to expose them to routine forms of violence and indignity, to deny them healthcare and addiction treatment, to remove all sense of autonomy and purpose. Being incarcerated often means that people lose their housing and go into debt. When they leave, the state provides little support for re-entry and employers and landlords often discriminate against those with a criminal conviction. In turn, incarcerating too many people can scar the places they come from, cutting children off from their parents and breaking the social bonds of trust and support so vital to healthy communities. Prisons don't control crime. They help to produce it.

This forces us to ask, why have the numbers of the incarcerated, in both the US and the UK, grown so significantly over the past 50 years? Who benefits and who suffers from this aggressive expansion of state power? Just whose safety is the state concerned with — and who does it construe as the danger to be surveyed, policed, contained, and disposed of?

Here genealogical critiques can make a productive intervention. A genealogical critique problematises the status quo by revealing the contingent relations of force that produced a given institution, practice or discourse. Such historical reconstructions reveal that societies could have been organised otherwise and gesture towards the social functions that a given institution or practice serves. For example, scholars have shown how the development of the police and the prison system in the US has been intimately linked with forms of racialised, gendered, and class-based population management. Angela Davis has written extensively about how the afterlives of racial slavery and empire have shaped the modern carceral state, from the chain gang after abolition through to the mass incarceration crisis of the past 50 years. Abolitionists have re-conceptualised the late 20th century US carceral state as a 'prison-industrial complex', drawing on Dwight Eisenhower's warning about the growth of a 'military-industrial complex' in the wake of the

Vietnam War. In doing so, they draw strategic links between militarism, punishment and capitalism, arguing that a 'punishment industry' has taken root that profits from expanding and supplying the state's expansive custodial estate.

Perhaps no one has offered as clear a diagnosis of this punitive (and profoundly racialised) drive to lock people up in the late 20th century US than Ruth Wilson Gilmore, whose *Golden Gulag* explores why liberal California built upwards of 23 new major prisons, in addition to a range of other custodial institutions, between 1984 and 2005. Contrary to readings that attribute the mass incarceration crisis to the demand for cheap prison labour or the profit incentives of the private prison industry, Gilmore shows how it was the state, with the support of finance capital, large landowners, and de-industrialising rural areas, that drove this prison-building project. Ultimately, Gilmore argues that the growth of the US prison regime has been inseparable from the broader neoliberal project of 'organised abandonment' that stripped the welfare state for parts and left racialised, poor communities to fend for themselves as capital moved production abroad. For thinkers like Davis and Gilmore, then, it is not that today's prisons are failing to operate as they should and are therefore in need of reform, but that they are functioning as they are designed to — namely, as racialised, class-based, and gendered technologies of containment. They have become the primary means by which the state deals with racialised and class-based 'surplus populations' whose labour is both not needed and may never be needed under conditions of structural unemployment.

This critique of the US prison system is well trodden terrain. Less remarked upon is the ballooning numbers of people living behind bars in the UK. But I contend that the British prison regime can also be understood as a pathological state formation. It too functions as a form of surplus state power that has taken on a life of its own, sweeping up more targets and entrenching itself deeper within the institutional fabric regardless of whether it is necessary or useful. Today, the UK locks up a higher proportion of its people than any country in Western Europe. About 6,700 people in Britain are serving life sentences; by comparison, the remaining 51 countries in the Council of Europe (excluding Turkey) have a total of 9,002 people on life sentences combined. And the British state's carceral footprint has more than doubled in the past four decades, as the number of people in state custody went from around 40,000 in 1991 up to more than 88,500 people in early September of this year. The Ministry of Justice predicts that Britain could have as many as 114,800 people behind bars by March 2028, if current trends continue. Significant drivers of these skyrocketing

incarceration rates have been ill-thought-out and overly punitive sentencing laws, including the much-maligned Imprisonment for Public Protection sentences and mandatory minimums for certain crimes. Sentences have got longer, more people are spending time in prison for petty crimes, and the state has criminalised more and more activity, including forms of 'disruptive' protest.

This explosive growth in the prison regime, however, cannot be adequately diagnosed without reckoning with the long *durée* of Britain's imperial past or the inequities of its postcolonial present. In the wake of white backlash to postcolonial migration from the Caribbean, South Asia, and Africa in the 1960s and 70s, British authorities adopted many of these repressive techniques of population control it had developed in colonial counterinsurgency campaigns in Malaysia, Kenya, and Cyprus to discipline growing racially minoritised communities at home.

As Stuart Hall and his colleagues presciently showed in *Policing the Crisis*, the state and the media fed a racialised moral panic over young Black men and 'mugging' in the 1970s and, in doing so, built a popular consensus around the need for harsh 'law-and-order' measures, including long custodial sentences. Such a crisis was 'lived' and explained through race, Hall and his Centre for Contemporary Cultural Studies colleagues argued, but it had its roots in the broader economic crises of postwar British capitalism and the fast pace of social and cultural change that had left many British people feeling displaced and insecure.

Over the interceding decades, a revolving cast of racialised scares has driven the public appetite for tougher sentencing laws, from concerns over 'gang' activity and knife crime to Islamophobic worries over the 'Muslim terrorist' to the recent preoccupation with racially minoritised migrants and asylum seekers. As five senior former judges pointed out in a September briefing paper denouncing Britain's 'sentencing inflation', many of today's punitive sentencing requirements can be linked back to 'single-issue campaigns' that were mobilised by 'emotive media attention' and called for mandatory minimums and harsher state responses, thereby removing judicial discretion.

Structural racism, overpolicing, educational inequities, substandard housing, changing migration laws, and increased susceptibility to precarious under- and unemployment, in turn, mean that racially minoritised communities disproportionately suffer the repressive heel of the British state. Racially minoritised people, especially Black men, are much more

likely to be arrested, and they are more likely to be sent to prison than white people for the same offences. It should come as no surprise, then, that the racially minoritised are overrepresented in British prisons (27% of those incarcerated versus 18% of the general population). There are almost three times more Muslims in prison today than there were two decades ago. As the senior justices warned, 'Without urgent remedial action to address sentence inflation, this country will soon experience US-style mass incarceration.' The new Labour government, however, seems more invested in 'fixes' to the prison crisis that expand the footprint of the carceral state, rather than in radical solutions that would reduce the number of people we lock up and the communities we abandon.

Perhaps most importantly, abolitionists like Davis advance a fierce ethical indictment of incarceration as a deeply inhuman and inhumane practice. This is where critical theory diverges from applied analytical jurisprudence. Within analytical jurisprudence, philosophers of law debate how to best to understand legal concepts and theories of punishment, to ascertain how our practices might be better brought into alignment with those theories. Such a focused analytic exercise, of course, is important and can bring real insight into what we should (and should not) expect from the law. But critical theorists want to offer a more radical critique of punishment in modern societies than what is available within jurisprudence alone. On one hand, they want to diagnostically situate the rise of repressive state power within a broader picture of neoliberal capitalism and the making-surplus of whole communities. And on the other, critical theory mobilises a more robust ethical critique, one in which existing practices of penal 'justice' are indexed against a theory of human flourishing and found wanting. Whereas sociologists or criminologists might offer data on the (in)effectiveness of specific carceral policies, practitioners of social critique are unafraid to indict a broader form of life that sees the prison as a normal and natural response to socially produced insecurity and deprivation.

Social critique therefore raises the question of whether we want to live in a society that locks people in cages for years at a time, sometimes for decades, and calls it justice. That abandons people to overcrowded cells where they are trapped for 22 hours a day without purpose or hope. That exposes them to routine violence, medical neglect, unsanitary conditions, and the daily humiliations of prison life. That sanctions routine sexual coercion and the intensive state disciplining of gendered bodies. And social criticism calls many of us out for supporting policies that have abandoned entire communities to violence, poverty and neglect while pathologising them as 'deviant' or 'dangerous' and therefore making them

disposable. In doing so, critical philosophy of this kind dares us to envision a world without prisons, without militarised policing and immigration detention centres.

Such a political horizon might seem utopian, especially in a social formation in which global capitalism sets the terms of possibility. But abolitionists aren't as worried by the charge of utopianism. After all, while everyday political change tends to take place through incremental reforms, mobilising people often requires a radical and emancipatory vision — one committed to the idea that other worlds are possible. And there are concrete decarceral strategies that can move us further away from these repressive and dysfunctional responses to inequality, poverty, drug use, migration and crime without opening the doors on all the prisons tomorrow. As Davis points out, abolition isn't just about the negative work of dismantling repressive institutions and state practices; it is also about building new institutions that can make the prison obsolete. This vision of 'abolition democracy' involves making a radical investment of state resources into communities — from healthcare to education to housing to childcare to public spaces — and doing so on terms that give communities autonomy and input over that support. Social critique certainly doesn't have all the answers about how we get to this decarceral future. But it does the important work of denaturalising what Pierre Bourdieu called the 'inert violence in the order of things'. Abolitionists force us to ask what kind of human and humane world we might want to live in together, if we just had the conviction and the courage to work for it.

Call the midwife

Professor Roger Crisp, The Uehiro Oxford Institute

Professor Katrien Devolder, The Uehiro Oxford Institute

Dr Jonathan Pugh, The Uehiro Oxford Institute

Who was the first philosopher? Most people would say Socrates, and there is no doubt that he, and the character named after him in the dialogues of Plato, have played a huge role in the development of philosophy over the last two-and-a-half millennia. Socrates, then, in a sense gave birth to philosophy. But how could that be, given his frequent claims to knowing nothing except his own ignorance? Partly because those claims can easily be read as ironic, but also because his dialectical way of doing philosophy enabled others to put forward and develop their own ideas. Indeed, one of Socrates's most famous images is that of himself as a midwife. In this paper, we shall suggest that modern philosophers might be seen, and see themselves, in the same way, using the resources of philosophy to help others give birth to their own ideas.

At least as he appears in the evidence available to us, Socrates was a truly public philosopher. He believed he had a divine mission to interrogate everyone he encountered on fundamental moral issues, and through exposing their ignorance put them in a position to live more virtuous lives. His methods contrast with those of modern academic philosophers. He sought to discuss philosophy with all, not only his students or fellow philosophers. He did not see philosophy as a profession and indeed would take no payment from his interlocutors. Nor did he publish, believing that the only way to make progress in philosophy is to discuss it with real people. He argued that the written word cannot defend itself or be questioned and cannot teach the truth effectively.

Academic philosophy as we know it, then, might be said to have begun with the work of Socrates's most famous discussant, Plato, who founded a famous school for philosophy in Athens (the Academy), wrote down his views, and disseminated them. And when we consider the history of philosophy from the time of Socrates and Plato until now, it is hard to think that Plato's decision to record his thoughts was a mistake.

But someone might suggest academic debate in philosophy is pointless since there has been so little agreement, even on fundamental issues such as truth, the criteria of personhood, or ethical principles. Here it is worth remembering that philosophy is, probably, still at an early stage of its development: we often forget that we may well be living at the very beginning of an extraordinarily long intellectual history. Further, philosophy, including philosophical ethics, has made a good deal of progress. By the end of the 20th century, for example, the differences between various consequentialist and deontological positions were significantly clearer than they had been at the end of the previous century. And, through successful public engagement on a large scale, several philosophers have in recent years greatly advanced the understanding of ethics among the general public. These include, for example, Kwame Anthony Appiah in his *New York Times* column ‘The Ethicist’; Mary Warnock in her work on bioethical issues; Michael Sandel, in many television programmes and online courses; and Toby Ord, in his work on existential threats.

In addition, philosophers have, partly through their publishing, contributed to important ethical advances. For example: the abolition of slavery — Adam Smith; equality for women — Mary Wollstonecraft and J.S. Mill; education reform — John Dewey; the decriminalisation of homosexuality — H.L.A. Hart; the global distribution of healthcare resources — Larry Temkin; attitudes to global poverty and the treatment of non-human animals — Peter Singer; and human rights and economic development — Martha Nussbaum and Amartya Sen.

Philosophy, then, can change things, and work in practical ethics can sometimes have quite revisionary implications for the world of policymaking. This should not surprise us. Professional moral philosophy champions the rigorous scrutiny of even the most widely and deeply held beliefs and values, and some of its best work can substantially change our understanding of these beliefs and values, or what they might entail.

This willingness to challenge existing precepts and frameworks is in many ways a virtue of the field, and this sort of work in philosophy can of course sometimes have real world effects on policymaking. Yet this willingness can also sometimes be a limitation when it comes to thinking about the potential of philosophy to impact real-world policy. It is one thing to convince an academic colleague of a philosophical argument to change the world and change their current understanding of a practical problem; it is quite another thing to argue that it would be politically feasible or expedient for those in the policymaking sphere to follow suit.

At any rate, philosophy *can* assist with progress. Further, this is something we should expect (both expect to happen and *expect* from philosophers). But there is no doubt that a lot more could be done. One problem is that the explosion in professional applied or practical ethics happened during the 1970s, when publication had already become the primary mode of assessment of philosophical reputation, and hence the driver for academic activity. In the UK, for example, there have been eight major national research assessments since 1986, and only in 2014 was 'impact' considered (and then only in the form of somewhat artificial 'impact cases', rather than the direct engagement of individual academics with the public through publication, the media, committee work and so on).

Philosophers themselves have tended not to object to the focus on academic output, primarily because they like doing philosophy, and having their ideas taken seriously by their peers in the discipline. The result is that it would be difficult to significantly lengthen the list of 'engaged' philosophers given above. But philosophers should not be blamed for this, nor should members of the public for their ignorance of what philosophy is and how it might play a much larger part in public life. Both sides are the products of their times. If a finger is to be pointed anywhere, it might perhaps be towards Kenneth Baker, the education secretary in the Thatcher government, for not including philosophy in the 1988 national curriculum. Had the public become more aware of philosophy, there would have been more opportunities to benefit from interaction with professional philosophers, an interaction which would almost certainly have benefited professional philosophy as well as individual philosophers.

Let us operate on the assumption that clarity, understanding of arguments and their logic, and perspective are, in general, valuable in public life. Philosophers, then, because of the peculiar focus on argument in their training, have a huge amount to contribute, but they are not called upon as often as they might be. The current situation is in some ways analogous to a programme in which medical researchers are hired to make progress on treatments for various illnesses, and merely to publish their papers in journals, which are known to be read mainly by other medics. (Medics, of course, do get kudos from the success of their research, though the drivers behind their success often arise from the interest of drug companies. Philosophy is equally in the public interest, but not marketised.)

All, or at least most, philosophers, then, should be required by their employers to take part in public life. (This would of course require a national strategy, imposed by the Department of

Education.) This includes non-ethicists: they have much to contribute merely through their philosophical expertise. Further, for ethics in particular, 'ethical consultancy' programmes should be instituted, analogous in some ways to existing business ethics consultancies. Though they should, wherever possible, be non-profit, and their focus should not be restricted to the culture of the organisation that uses their services. Their function should not be seen as in any way moralistic or protreptic. The aim would be to help others to think through the issues they are facing with a more self-conscious focus on clarity and rigour.

It might be objected that these changes will slow down progress in academic philosophy. But, one might respond, the current situation may be inefficient, and fewer papers would not necessarily lead to slower progress in philosophy. And, through greater interaction, the philosophy that is published, especially on practical issues, would be more informed and helpful. Philosophy does move forward slowly, that is, carefully, and maybe things are better that way. And, even if the objection is correct, the advantages to the public of greater engagement by philosophers outweigh the drawbacks. Nor need we expect every philosopher (even the very best) to be required to spend the same amount of time on public engagement. Further, public engagement should be understood broadly. Some philosophers are better at, or more inclined to, working with the media, while others may prefer to serve on committees, publish accessible papers, or to work with organisations such as the Public Philosophy Network.

Earlier, we mentioned the potential for philosophy to lead to radical or revisionary conclusions. But it needn't do that, and in fact moral philosophy can also play a quite different, and arguably more powerful role in this sphere. Rather than using the tools of moral philosophy just to develop theoretical arguments supporting change (as do Appiah and the 'public intellectuals' listed above), the philosopher in the public sphere can also seek to use these tools to enhance understanding of the practical options that are on the policymaking table.

This is where philosophical obstetrics comes in. Let us return to ethical consultancy. The University of Oxford and some other leading universities have focused on 'research-led' teaching, that is, the teaching of undergraduates, in particular, who then go on to make a great contribution to their country (in most cases — politics may be an exception ...). Ethical consultancy is analogous to that, but it extends the net to include the public at large. As we noted, this already happens in certain businesses, and when it is done properly, rather than as a form of 'ethics-washing', it can improve corporate governance as well as the lives of

employees and other stakeholders. Here at the Uehiro Oxford Institute, we and many of our colleagues have been involved in work with government, both national and local, the NHS, schools, the police and prison service, as well as a wide range of other institutions through our Master of Studies (MSt) in Practical Ethics (University of Oxford, 2025), Bitesize Ethics (University of Oxford, 2024) series, and other programmes.

The consultancy role is perhaps of most use in areas of public debate in which there are a variety of conflicting moral considerations supporting mutually exclusive policy choices. This is particularly so when the public debate on the topic has become politicised in a manner that precludes different sides of the issue from acknowledging any weakness in their own position, or any strength in their opponent's. In such circumstances, progress can be extremely difficult, and the stalemate in the public and political debate may even lead to a kind of policy paralysis.

This is one important area in which the skills of the philosopher — not just the moral philosopher, but also the logician, the metaphysician, the epistemologist, if they have the right communication skills — can help to enable progress. By clarifying concepts, analysing the structure of the arguments that are employed to support different policy options, and pointing out where their different strengths and weaknesses may be, the public philosopher can help the policymaker to chart and navigate the different value trade-offs that they will unavoidably have to make in developing coherent policy on highly contested issues.

The benefit of philosophical input here is not that the philosopher can 'solve' the moral problem facing policymakers; unfortunately such neat solutions are rarely available in either politics or philosophy. Instead, it is to help ensure that the policy response to an issue with moral dimensions is well reasoned and systematically considers a wide range of morally relevant factors. Ultimately, then, philosophical input can help to ensure that policy provides moral justification to those affected by the policy. After all, to retain their legitimacy, it is crucial that policymakers can offer this group a reasonable answer to the question, 'Why did you decide that this was the best course of action?'

It might be objected that this is all well and good, but that philosophy as an academic discipline will suffer because its practitioners are spending time on public engagement rather than academic research. This objection cannot be dismissed out of hand. But it is worth noting, first, that a great deal of philosophical research is funded by the public, who might not unreasonably expect philosophers to be prepared (indeed required),

like real midwives trained at public expense, to directly contribute to the public good. Further, to some extent, the value of research diminishes marginally, and it may be also that philosophers’ ability to think and to express their own views will be advanced through communication with non-specialists, with a focus on accessibility, avoiding jargon, and so on.

One example of putting this kind of vision of philosophy’s role in public life into practice is the Uehiro Oxford Institute’s Decision Aid for the Restitution of Cultural Artefacts (DARCA). The question of whether cultural institutions (like museums) ought to return certain items within their collections has been subject to increasing debate in recent years. There are two broad competing moral obligations that largely shape the debate in this sphere. On one hand, it is sometimes claimed that cultural institutions have a moral obligation to the publics they serve to preserve and showcase valuable cultural artefacts, because these items manifest important parts of humankind’s shared cultural heritage. Conversely, however, these same institutions are sometimes understood to be under a moral obligation to return certain artefacts because another party has a plausible moral claim to that item, perhaps because its return would constitute an appropriate form of reparation for a historical injustice involved in the object’s transfer of ownership. In view of this conflict, DARCA was developed to enable those facing the practical policy challenges of cultural restitution to adopt coherent and justifiable policy responses on a case-by-case basis.

DARCA was developed in close partnership with an interdisciplinary group of stakeholders. Across two workshops, and following a review of existing sector guidance, law, policy, and academic moral philosophy, the DARCA team identified a range of moral considerations that can be understood to affect the relative strength of these two moral obligations. Following iterated consultation with experts in the field, the team aimed to synthesise this body of work into an approachable and accessible tool that can enable individuals to think systematically through various considerations relevant to ethical questions about cultural restitution. To do so, the team developed a series of questions that can be used to assess how these factors might arise and interact in a particular case of cultural restitution. These questions form the theoretical basis of DARCA, and users answering these questions are invited to reflect on some of the philosophical complexities of each of the questions raised with carefully curated guidance notes. Upon completing DARCA, the user is provided with a summary of their answers, any further written justification they may have given for those answers, an overall assessment of the comparative strength of the argument for restitution, and an explanation of that assessment.

DARCA cannot solve the moral issue of cultural restitution, and that is not its intended goal. Indeed, it is intended to guide and supplement the user’s own moral reflection on this complex area of policy, rather than to replace it. However, after using DARCA, the user should have a better sense of where the moral arguments appear to lead in their case, why they do so, and what moral factors have made a difference. The role of philosophers in developing DARCA, then, has not been to provide policymakers with the ‘right’ answer; it has instead been to help them reach their own answer in a manner that is systematic, well-informed, and philosophically grounded. The role of the engaged practical philosopher, then, is primarily not to persuade, but to provide assistance to others in reaching their own, reasoned conclusions.

References:

- University of Oxford (2024). *Bitesize ethics 2024*. [online] Available at: <https://www.practicaethics.ox.ac.uk/event/bitesize-ethics-summer-programme> [Accessed 21 Feb. 2025].
- University of Oxford (2025). *MSt in Practical Ethics*. [online] Available at: <https://www.practicaethics.ox.ac.uk/mst-in-practical-ethics> [Accessed 21 Feb. 2025].

Ours to question why

Professor Cécile, Fabre University of Oxford

As I write this in November 2024, the war between Ukraine and Russia appears to have no end in sight, and neither does Israel's conflict with Hamas in Gaza. The UK is supplying military assistance to Ukraine and Israel. While the UK is not supplying 'boots on the grounds', it is contributing to the deaths of thousands of people. In living memory, its forces have directly killed individuals who had been declared the country's enemies, by order of the elected government of the day.

Of all the things that we do, killing another person stands in the greatest need of moral and legal justification: in the absence of such justification, it can constitute manslaughter or murder. A decision to resort to military force, regardless of who makes it and of its rationale, is in effect a decision to authorise soldiers — on our behalf — to kill enemy combatants and civilians, whether through deliberate acts or collateral consequences of military action. Thus, when parliament votes in favour of such action, it authorises soldiers to kill, not merely on our behalf but also, in some important sense, at our behest. When the government agrees to supply weapons to its allies, it provides the means for the act of killing. If we are to take the presumption against killing seriously, we — citizens, soldiers, elected representatives and government ministers — must also take seriously the task of morally justifying war.

Some view this task as hopelessly naive, futile, or both. It is naive, because states go to war if and when it serves their interests, and whatever philosophers have to say about the right and wrong of doing so has little purchase. It is futile, because what matters the most, surely, is not whether states are morally allowed to go to war and whether soldiers, so authorised, are morally allowed to kill enemy combatants and civilians: what matters is whether those acts are lawful.

And yet, a striking feature of recent parliamentary debates on military action is the highly moralised language employed by members of parliament (MPs) for and against war. In March

2003, the House of Commons voted in favour of the government's motion to wage war against Iraq, as part of a US-led coalition. Thus, arguing against the deployment of forces, Douglas Hogg MP, stated: ‘I do not think that any of the usual characteristics of a just war have been satisfied. If we were dealing with a situation in which Iraq had attacked another country or had mustered troops on the frontiers of another country, or if there were compelling evidence that Iraq was delivering to terrorists weapons of mass destruction with which they could attack another country, I would vote for war, but none of those circumstances exists.’

Or consider Peter Bradley's impassioned speech in support of intervention: ‘I believe in just wars. I believe that they are commissioned in defence of freedom, and against oppression. I also believe that, for them to be just wars, they must be the last resort. Diplomacy must come first, but if we are to prevail in defence of what we believe to be right, there must also be a limit to diplomacy. [...] If we are to set aside our prejudices and accept that doing nothing is not an option, if we accept that diplomacy is at an end and that Saddam continues to defy and threaten us, what is the alternative to war?’ There is none, Bradley concluded, siding with the government, though noting the difficulties of establishing, there and then in 2003, whether the conditions for a just war were met.

Ten years later, in August 2013, the House of Commons voted against the deployment of a British intervention force in Syria. MPs framed the decision whether to intervene in Syria in the language of rights and duties. When debating whether intervention in either Iraq or Syria was lawful, there seemed to have been little doubt in their mind about the moral legitimacy of the international legal framework within which Britain had to operate. As a constitutional matter, the decision to go to war is a royal prerogative: it does not require the approval of parliament, and any such approval (or disapproval for that matter) is merely advisory. Thus, in 2018, the then-prime minister Theresa May authorised military strikes against President Assad's regime (which had used chemical weapons against rebel forces and civilians) without seeking the consent of parliament. Nevertheless, as Hansard records, she addressed the House of Commons at length in similar terms to her two predecessors: ‘it was not just morally right but legally right to take military action [...] to alleviate further human suffering’.

Turning to the present, public debates, both inside and outside parliament, on the wars between Russia and Ukraine, and between Israel and Hamas, are suffused with moral language: the claim that Ukraine is morally justified in defending itself from Russia's unwarranted aggression is seen as beyond dispute; likewise, the claim that Hamas committed atrocities

against Israeli civilians on 7 October 2023 is widely endorsed (including by proponents of an independent Palestine); so is the view, at least in this country, that the humanitarian costs of Israel's retaliatory war in Palestine since then are disproportionate.

Both within parliament and, indeed, in the press, there is a remarkable degree of consensus on the conditions which a war must meet to be just and lawful. It must have a just cause — such as national defence against a military aggression, stopping a humanitarian disaster, or upholding the moral and legal prohibition on the use of weapons of mass destruction — and reliable evidence that there is such a cause. It must aim to bring about those just ends. It must be a proportionate and necessary means of achieving those ends, including in the way it is fought. More specifically, minimising harm to civilians is of paramount importance. Disagreements focus on whether those conditions apply in the specific cases under consideration.

Without explicitly acknowledging it, our MPs and prime ministers are engaging in philosophical arguments: they do what philosophers and theologians have done for millennia in both western and non-western traditions. Anecdotally, from my own experience as a philosopher of war, so do many soldiers. The question, then, is not so much whether foreign policy actors should engage with the philosophy of war, but rather whether they do so as well as they could. No less crucially, it is also what philosophers of war can learn from those public debates: if philosophy has a role to play in public life, so does public life in philosophy.

These conditions of a morally just war are familiar, intuitively powerful, and, at first sight, provide secure foundations for the laws of war. On closer inspection, and as philosophers of war have noted, they are much less straightforward than they appear. Let me highlight two serious difficulties. First, they raise thorny interpretative problems which, once brought into view, weaken prospects for consensus. Consider for example the claim that unwarranted military aggression is a just cause for war. In most cases, and paradigmatically so for Ukraine in February 2022 or Poland in September 1939, this may well be true. Suppose however that the aggressor-state invades the defender-state without shedding blood — as Russia did when it annexed Crimea in February 2014. Or that it seizes a remote and uninhabited part of the defender-state's territory. The aggression takes place without violating anyone's right not to be killed. The question, then, is whether and when violations of the comparatively less important right to territorial sovereignty warrant the use of lethal force against enemy combatants who have played a minimal role in those violations, as well as the collateral deaths of civilians.

Consider next the claim that war must be a proportionate means to achieving justified ends; that is, the bads it does must not be excessive in relation to the goods it brings about. To ascertain whether the war would be proportionate, thus, we must decide which goods, relative to which evils, are relevant. For example, suppose we go to war against an authoritarian state which has invaded us. There may be an opportunity to remove the authoritarian state's leaders from power and help its population install a democratic regime. It may seem that these are goods which ought to count against the bads our forces will inflict. Yet, it is doubtful that regime change is itself a just cause for war. If it is not justified to wage war to bring about this particular end, why should the outcome of a successful regime change be a factor when considering proportionality? And, if we do go to war against our unjust aggressor, the result could be that the aggressor's increased military spending leaves its civilians entirely destitute. Should those bads count against our decision to go to war?

Assume, next, that we can determine the relevant goods and bads. As the war goes on, policymakers must decide whether to carry on with the war or whether to sue for peace and withdraw their troops. This includes deciding whether the costs incurred so far count in the proportionality calculus, or whether they can be discounted as sunk costs. Suppose our elected representatives authorise the government to defend our homeland and anticipate a proportionate and thus acceptable cost of, say, 15,000 combatant fatalities. However, suppose the war goes badly: six months into the conflict, there are 10,000 combatant fatalities, for much smaller territorial gains than anticipated. Our military leaders estimate that a final push would deliver victory, but at a cost of 8,000 lives. Assume that they are correct: victory, then, in the end, would cost 18,000 lives.

There are two ways of thinking about the 10,000 lives already lost. On the one hand, they can be thought of as 'sunk costs'. Given that they can never be recovered, our leaders need not take them into account as they consider what to do next: rather, they must focus on the additional lives that will be lost if they proceed. As 8,000 is well below the 15,000 threshold, continuing with the war is not disproportionate. On the other hand, we can deem those 10,000 lives not to be 'sunk' and take the view that those costs are to be carried forward. If so, the war must stop, as it would fail to meet the proportionality condition. What should our leaders do?

What about the requirement to minimise harm to civilians? On the face of it, it is clear civilians should be protected from harm wherever possible. Matters are not so simple, however. For the problem is that some civilians contribute to unwarranted wars of aggression — for example,

those who work in the arms industry. There are others: civilian members of intelligence services; journalists and so-called ‘influencers’ who feed the war propaganda machine; and also civilian leaders, including (why not, after all?) civilian elected representatives who, in regimes which are or claim to be democratic, vote to authorise and facilitate the war.

Perhaps, then, the requirement to minimise harm to civilians only applies to innocent civilians — who do not participate in those ways in the war, let alone take up arms. They clearly are not legitimate targets; and whatever harm is done to them, as a collateral effect of our war, must be kept to a minimum. So far, so simple. Suppose, however, that complying with the requirement would entail a serious loss of life on our side. To whom must our leaders give priority, and on what basis? Answers to this most-difficult question are likely to depend on the kind of war we fight: whether it is a war of self-defence, or a war of intervention — and on whether the fatalities are soldiers (and whether these are conscripts or fight voluntarily) or fellow citizens.

There are no easy answers to the questions I have raised. The main lesson to draw from my remarks, if there is one, is that, in the light of a well-developed body of philosophical work on the morality of war, the appearance of consensus is likely to be deceptive — and indeed deceptively dangerous. The costs of being mistaken about, for example, what is a just cause and what is proportionate are measured in lives, not just in financial costs, votes, or reputation.

The second difficulty is that the morality of war is not as clearly aligned to the laws of war as is often supposed by many outside philosophy (and also, by quite a few within the just war tradition). The laws of war sharply distinguish between the legal status of a decision to resort to war, and the legal status of acts of killings within the war. Thus, take for granted that Russia unlawfully invaded Ukraine on 24 February 2022, and that Ukraine had a lawful cause for responding by force. Yet, although Russia’s war of aggression is unlawful, the individual acts of killing carried out by Russian soldiers against Ukrainian soldiers in pursuit of their regime’s unlawful ends are not, in themselves, unlawful. Indeed, they are not any less lawful than the acts of killing carried out by Ukrainian soldiers against Russian soldiers in defence of their country.

Morally speaking, however, this cannot be right. For if it is morally wrong to wage a war of aggression, then by implication it is morally wrong to kill in pursuit of that war. To be sure, it may well be that soldiers on the ‘unjust’ side have no choice but to fight — perhaps they are conscripted under life-threatening duress, or perhaps they are unavoidably unaware of their regime’s true ends. It is worth noting, however, that those reasons would not be

considered as exculpatory of the deliberate killing of innocent civilians. Yet, even if, for example, Russian soldiers are not blameworthy for killing Ukrainian soldiers, it remains the case that they are not morally permitted to do so, whereas (by implication) Ukrainian soldiers are morally permitted to kill in defence of their homeland.

Some readers might have no difficulty granting the point in this case — so deep is moral outrage, in this country at least, at this war. But it is worth highlighting what it implies: if, as some readers might think, the UK, the US and their coalition partners lacked a just cause for invading Iraq in 2003, then by implication and whatever the laws of war may say, soldiers of the coalition force were not morally permitted to kill Iraqi soldiers who were resisting the invasion. Theorists of the just war are familiar with such a view — indeed many explicitly endorse it. Outside academic philosophy, however, I wager that it is rather controversial — so much so, perhaps, as to cast doubt on the usefulness of philosophy to public policy, at least in this area.

Here is what non-philosophers may well be tempted to say. What follows is a rough summary of remarks I have often heard over the years:

'Well, many of our soldiers believed that they were doing the right thing, in fact the MPs who voted in favour of the war in 2003 believed they were doing the right thing. All those people acted in good faith. The invasion of Ukraine, or the Second World War for that matter, which philosophers of war keep invoking, are easy cases — too easy to serve as reliable guides given the complexities of most conflicts. Besides, suppose you're right and that our soldiers were not morally permitted to kill Iraqi soldiers who were defending their country. So what? Do we really want to revise the laws of war, so that killing enemy combatants in an unjust aggression is a war crime? Surely not: who would stand in judgement? How could we possibly enforce those laws? Besides, let us not pretend that Iraqi soldiers were saints: as well as defending their country, they were also defending a murderous regime so if the war was unjust, it was unjust on both sides, and it is really not clear what to make of that.'

Philosophers of war can learn at least four lessons from these sorts of remarks. First, they need to pay greater attention to the conditions under which policymakers — members of parliament, cabinet ministers, defence officials, military personnel — are required to make those decisions that will inevitably lead to deaths. This requires that they seriously

consider empirical evidence about war and political decision making, as well as records of parliamentary proceedings and publicly available policy briefs.

Second, they need to be more attentive than they have been so far to the fact that, according to their own principles, no war is fully just or unjust. This obvious observation, which one often hears from non-philosophers, has profound implications for our moral assessment of decisions to go to war in the first place, and of individual military missions with that war.

Third, they need to contribute more than they have done so far to debates about the relationship between the morality and the laws of war, and to discussions about the dilemmas faced by public officials whose moral judgements are in tension with the law. 'Resign!', we might say. But what if the law offers the best moral compromise we can hope for under the circumstances? What about the costs of keeping one's hands clean?

Finally, philosophers of war need to do much more work than they have done so far on the rights and wrongs of decision-making under conditions of uncertainty. By uncertainty, I mean both uncertainty about the facts of the case (for example, does a satellite image of troop movements along a border constitute evidence that our neighbour is about to attack us?), and uncertainty about what is the right thing to do (for example, does the loss of 10,000 soldiers constitute a sunk cost or not?).

Moral uncertainty is particularly troubling. There is no settled consensus amongst philosophers on many of the issues discussed here. Our leaders do not have time to wait until we develop a plausible view on sunk costs before deciding whether, for example, to increase military support for Ukraine. But what they ought to expect from philosophers is some guidance on the moral implications of both views.

To be sure, philosophers have addressed those issues — just as policy actors are aware of the moral complexities of seemingly straightforward principles. All too often, however, they fail, it seems, to communicate their views outside academia. To make progress, what is needed from both philosophers and policymakers is greater reciprocal engagement. This could take many forms, from compulsory ethics teaching in military academies, to setting up more systematic and regular exchanges between philosophers of war, our elected representatives, and members of the armed forces. In our current geopolitical moment, ethical reasoning that is both rigorous and attuned to the realities of armed conflicts is more important than ever.

On the new demise of ethics

Dr. Elke Schwarz, Queen Mary University of London

This essay is about the disappearance of ethics in an era which is characterised by the capacity to inflict large-scale violence at machine-speed. In what follows, I argue that in a technologically saturated military environment, where artificial intelligence (AI) is fast becoming a foundational infrastructure, ethics has taken a leave of absence. This should worry us, because the erosion of ethics in the domain of war and conflict might also be a harbinger for the demise of ethical thinking more broadly in public life.

At the time of writing, the world is witness to two aggressive international wars that have been raging on with breathtaking levels of human and material destruction. These two wars, in Ukraine and Gaza, have accelerated the development and rollout of new AI technologies that strive for greater and faster delivery of force, and reports of AI-enabled autonomous weapons systems being fielded are beginning to proliferate. The systems of greatest concern are those in which the human decision-maker plays an increasingly marginal role. In such times, which stand under the tragic sign of expanding violence, escalating conflict, and rising geopolitical tensions, the evaporation of what it means to give ethical consideration to other human beings should give us pause for thought.

At a 2024 AI Expo conference in Washington DC, a panel of experts with military, government, and defence industry backgrounds contemplated the wars of the future. The panel's core message was that future wars will be both devastating and incredibly lethal, but they will also be precise and humane because of new AI-enabled weapons. With this, the experts seemed to suggest that responsible powers must acquire AI weapons to deter irresponsible powers from using such weapons, and if responsible powers must use such weapons, they will do so, of course, responsibly. The ethos of this peculiar justification is perhaps best summed up by the remarks of one of the panellists — the CEO of an AI company that specialises in offensive targeting — who proclaimed with a remarkable

degree of enthusiasm: ‘The peace activists are actually the pro-war activists, and we’re the peace activists!’ — proposing the argument that, if you don’t want war, somehow the responsible action is to gear up for one. You would be forgiven for not following this logic. It makes a somewhat muddled allusion to the mechanisms of deterrence theory, a Cold War game theory-inspired way to justify stockpiling that most absurd of all weapons, the nuclear bomb. It is not a coherent discourse. But it is one in which the erosion of the status of ethics is patently demonstrated. This erosion has taken place gradually, concurrent with the ascendancy of computational rationality across various domains of human affairs.

In his 1954 book, *The Human Use of Human Beings* (Wiener, 1954), the grandfather of cybernetics Norbert Wiener makes some incisive observations about the co-constitution of humans and autonomous machines. He saw the perils of cybernetic technology for society as twofold. First, its inner logic might be wielded by one group of humans as a form of power and control over others. Second, this logic might also lead humans to believe it would be better to outsource complex and difficult moral decisions to machine authority. As an eminent scientist of technology and mathematical thinker, Wiener was ill at ease with the way in which technological thinking was fast becoming a dominant mode of philosophical and social reasoning. An insight we appear not to be heeding today.

Indeed, cybernetics, especially in the form of computational processes, has captured the collective imagination of the west since the 1950s and 60s, prompting many to understand the world and everything in it through feedback logics and probabilistic reasoning. Cybernetics is, first and foremost, a theory of processes and mechanisms. It is a theory of functionality in which inputs are translated into outputs. In cybernetics, the world is conceived as a collection of information that is to be transmitted. Information is understood as calculable data and even humans are conceived as objects of information, rendering them functional elements in a system determined by process. Cybernetics is unique in that it enables material and mental phenomena to be conceived in the same computational register. This makes its technological logic especially alluring.

It is unsurprising, then, that certain, prominent strands of moral philosophy were seduced into this schema. Derek Parfit’s approach to applied ethics, for example, was thoroughly informed by the probabilistic logics of expected utility and game theory, and various modes of analytic moral reasoning about justifications for harm — from the Trolley Problem onwards — are characterised, in form, by approaching ethics as an algorithmic

puzzle. All this fits hand-in-glove with modern approaches to warfare, which are themselves more and more dependent on computational reason.

Every epoch has its own paradigmatic ways of conceptualising wars and of justifying the harms these paradigms produce. Ours stands under the sway of workflow efficiency. Current military doctrine and practice increasingly mirrors the logic of computational optimisation, prioritising cost-benefit calculations in the interest of speed and scale. All ethically salient considerations are swept up in this logic. The notion of permissible civilian harm, for example, is rerouted through the process of calculating 'collateral damage' — the percentage of civilians (men, women, children) that might be harmed if a situation necessitates it. A notoriously controversial calculus that clearly echoes the parameters of the Trolley Problem.

The Trolley Problem and its many variations are early iterations of algorithmic rationality. It is also fundamentally a problem about justifying harm. The basic scenario is fixed around the same dilemma: there is a runaway trolley, five people are tied down on one track, one person — either on another track or a footbridge or in some other perilous situation — will die. A decision must be made — will one person be sacrificed or shall five die? Each variation sets out to test to what degree different variables in the scenario matter to the moral outcome. New variables are introduced to change the moral calculus — a lever to divert the trolley; a tractor to rescue the five but which will then crush a bystander; the option to push someone off a bridge to stop the trolley. Technological instruments are always posited to make a difference to the justification of the lethal action. The role technologies play in shaping moral choices is almost never acknowledged. But crucially, this mode of reasoning implicitly constitutes an objectification of all that is human. Every subject becomes object in a world of computable things.

To be fair, Philippa Foot's original version of the problem was not intended to be used in such a formulaic manner, nor did Foot seek to spawn an entire 'Trolley-ology' industry. One might argue that utility-focused approaches to ethical reasoning originated precisely from a moral concern with social reform, refracted through a then radically new mode of accounting for egalitarian and universal values. (Although, it must be said, even in Bentham's era, the idea of smoothing the plurality of human life into a single calculus was controversial.) But the point at issue is this: ethical reasoning does not happen in a vacuum. It develops and changes alongside ways of envisioning and practising human life. In our time, these visions and practices stand under the sign of computation. Consequently, our forms of ethical reasoning are increasingly inflected by Bayesian statistics and game theory.

Nowhere is this more clearly felt than in the context of warfare. As military weapons become more computationally sophisticated, ethics is pressed further into a schema of technicity, to the point where ethics itself becomes technical, an exercise in which systematic process matters more than substance. Analytic approaches to ethics facilitate this. We can see this within abstracted conceptions of 'just war' reasoning, where the language of calculable utility dominates and the human is objectified, quite literally, in one of two ways: he or she is either the recipient of a moral decision and becomes a set of data points within a wider configuration of variables; or he or she is tasked with administering the moral calculation about who may die and in this way becomes a functional element in the wider data infrastructure. In their most abstract forms, analytic modes of just war reasoning become indistinguishable from technical discourse. They abound with mathematical syntax, taking on the pure form of algorithmic rationality itself: if X condition applies, then, *ceteris paribus*, Y must logically follow.

Algorithmic rationality seems tailor-made for thinking through ethical problems associated with autonomous, potentially lethal machines. The MIT Moral Machine project, for example, wants to identify the most agreeable ethical principles with which to programme self-driving cars, so that these cars can make the 'right' kill decision autonomously. Other research efforts by well-known technology companies are underway to design 'AI morality', so that humans can receive 'ethical guidance' on issues such as who should receive a kidney. When humans are rendered as objects of information, it is only logical that the most complex ethical calculations should be left to machines.

Abstraction marginalises the ethos of ethics altogether, especially in and for warfare. The aforementioned deterrence theory, popularised during the Cold War, is one such form of abstract reasoning about lethal technologies. It assumes that all actors are purely rational; that all actions and consequences of these actions are, to some degree, predictable; and that, based on this, the best outcome can be ascertained by drawing on expected utility and decision theory. This is ethics as economic science, as a technique for solving optimisation problems. In the context of nuclear deterrence, it overlooks entirely that the rational grounding of this approach rests on a threat of civilisational annihilation. In the contemporary context, with AI-enabled weapons, moral reasoning fits effortlessly into a workflow process, and it equally becomes ever more remote from recognising the ethos of the ethical problem itself, which is almost always an intractable social problem, resistant to optimisation efforts.

The abstraction is not accidental, but precisely the point. Abstraction decouples empathy from action. It decouples the consequences of actions from the moral weight of taking responsibility. It objectifies all those embroiled in the morally relevant calculus as data. It removes the most fundamental element of ethics — that of understanding one's responsibility in relation to a decision which affects others. It also works to render the 'rational' so fantastically technological that surely, only savvy experts can be trusted with engineering the difficult choices. This directs our gaze onto 'solvable' ethical problems and away from those that do not sit so comfortably within a techno-economic logic.

However, our morality, and thus our moral decision-making, is anchored in a shared history of human plurality and human social relations. It is this condition that makes us not just functional actors, but moral actors. Moral agency means the capacity not simply to have, but also to take moral responsibility in relation to others. And that means having the capacity and space to reflect on one's position and that of others in any given configuration of moral relevance, particularly in warfare. The ever-present potential for open-endedness in moral decision-making and responsibility-taking does not lend itself to the parameters of systems that function best with and within closures, and at speed and scale, as is the case with AI.

Analytic approaches to ethics are, of course, only one way of thinking about ethics as it relates to war and technologies of war. Instead of an algorithmic programme, ethics might be understood as a relational practice. Ethics as a practice is open-ended, deliberative, and ultimately concerned with broader questions concerning human relations, not technical processes. It rests on a foundation of human relations, anchored in both vulnerability and trust. As pivotal parameters for social and political life change, ethical guidelines, their associated practices, and our human relationship with ethical values and beliefs become subject to new inquiry. Ethical thought should inform the laws, guidelines, and rules that govern social and political life. But it must also serve to prompt critical reflection about the ways in which morality might become fossilised in these laws and rules, to avoid the risk that the form supersedes the content. I am reminded of Hannah Arendt's observation about morality and mores: 'What people get used to is less the content of the rules [than] the possession of the rules under which to subsume particulars.'

Having a set of codified ethical principles in place does not guarantee ethics in practice. This is tragically confirmed in our present moment when the law appears to be inadequate to curb the worst excesses of the ongoing conflicts, let alone stymie their expansion. Neither just

war principles nor international humanitarian law were intended as a checklist that could be satisfied, or optimised, in purely procedural ways. Rather, the intent with early just war thinking was to grapple with and understand that actions and circumstances need to be considered carefully in morally charged environments. That there are always broader ethical questions to consider in any war, questions about how we wish to live with one another, about the possibility for peace once fighting ends.

There is no easy answer to the question of how many innocent people may be killed to achieve a goal in warfare, nor is it sufficient to think of the ethics of war in mere numerical terms, counting only the dead. Those that make a decision to kill must wrestle with the weight of responsibility attached to such actions. For this, we must remain humans, not become objects ourselves, and we must tolerate the discomfort of an irresolvable ethical challenge, despite being embedded within a network of machine functionality. Herein resides moral responsibility.

The ethos of international humanitarian law or the just war tradition is still intuitive to most of us: those blameless should not have to suffer, violence should not be a first resort, and there should be a mandate for restraint to violence because otherwise peace cannot be restored. Where ethics becomes automated, this ethos is exchanged for variables that can be adjusted based on preference, interest and aim. In other words, it becomes subject to power, which ultimately is a feature of nihilism. But that is where we are today. The accelerated rollout of AI-enabled technologies of warfare, their imperative to make war faster and more lethal, and their logics of process and efficiency leave neither space nor reason for ethics to thrive, not even as a calculation. No wonder, then, that ethics is on its way out as an uncomfortable necessity, with the many intractable, irresolvable, difficult challenges that arise in war. Ethics is about meaning; AI is about function. The two are not easy bedfellows.

Two points are worth stressing in closing. First, AI's emphasis on speed is utterly incompatible with ethical deliberation. I have written elsewhere on how the iterative logic of failing and learning through accelerated processing power is both the foundation of AI as well as the technology industry itself. The motto ‘move fast and break things’ is not just a PR slogan; it is an ideology. Increasingly too, it seems, for the defence domain. With this shift, ethics is necessarily sidelined. The second point is that computational logics always objectify their environments, and that means objectifying humans. When an AI system identifies a human as a target-object, that human is immediately objectified, and thus

dehumanised. And from the history of mass violence, we know that dehumanisation plays a significant role in facilitating an erosion of moral restraint.

And algorithmic ethics cannot address either of these issues. Quite the contrary. Taking the ethos of ethics seriously would mean, first and foremost, not succumbing to the speed and scale mandate of killing as a systematic process. It would also mean that we ask more of those involved in AI-enabled war, not less; that we should consider maximally restrictive ideas about the use of violence rather than follow the logic of machine functionality. This means that those involved in the business of war should become habituated into thinking about humans as humans, not objects, depicted on an interface.

The ethical stakes are patently high, in warfare and otherwise. By functionalising the notion of agency and by marginalising the human as a decision maker, we diminish our ability to understand what moral agency, and responsibility, means in practice. To be sure, moral agency can be theorised and defined in different ways, and I accept that perhaps our idea of what constitutes agency should be revised in an AI-saturated environment. But agency and *moral* agency are not the same. Moral relations do not arise *ex nihilo*. As humans, we understand and are able to judge the contextual nuances of human relations in a way that technological artefacts are simply not able to. Technological artefacts can comprehend neither the complexities and incongruities of human life nor its meaning.

AI-enabled weapon systems are such technological artefacts. And they are also technologies of violence. Instruments of violence are designed to inflict damage on human bodies and human property. Too often we discuss these systems as though they were somehow separate from this violence against human life; as though war is primarily an engineering challenge and as though they somehow carry within themselves their own justification for the violence they enable. Algorithmic ethics facilitates this configuration.

Rethinking ethics not as systematic process, but as practice rooted in human relations, is an urgent task. If we let ethics disappear, and with that the intrinsic, mutual value of human life, we are readying ourselves for a future of violence, not peace.

References:

- Wiener, N. (1954). *The human use of human beings: Cybernetics and society*. New York: Da Capo Press.

The world and his wife

Dr Musab Younis, University of Oxford

Our current political moment is characterised by such extreme and shocking violence that any measured reflection on the role of philosophy in public life can seem impossible. The elevated ideas that accompanied the rise of international organisations in the 20th century — peace, harmony, order, cooperation — appear as distant from reality as at any moment since the end of the second world war. The events of the year 2024, and in particular Israel's merciless genocide of Palestinians in Gaza, resolutely backed by the UK, the US and other western powers, were described by observers as a 'breaking point' and 'death blow' for international order.¹ And yet if brute force still governs the world, the language of force does not. Principles that we might call philosophical in nature are still being articulated constantly in public life. Across the world, in both popular and official discourse, we find references to progress, fairness, efficiency, justice, development and peace.

I want to suggest in this essay that philosophy has a role to play in politics even if its actual operation in the world is not easy to discern. Philosophy, understood capaciously — I leave aside in this essay debates about the differences between political theory and political philosophy — can challenge the common assumptions and limits that frame political discourse. By asking questions that go beyond the usual bounds of public debate, it can illuminate the unexamined and ahistorical assumptions that are to be found everywhere in public life. This type of philosophy is not the exclusive preserve of scholars working in universities. Anyone who has sought to name and describe a form of domination they experience, with the aim of resisting it, is — in my sense of the term — a philosopher. Many of those who lived under colonial rule, for example, developed forms of anticolonial thought that radically challenged the legitimacy of empire. Their arguments are expressions of political philosophy that show why imperial formations like Britain cannot be understood as isolated and singular national units.

Since the wave of decolonisation in the 20th century created a world of sovereign and supposedly equivalent states, many academic political theorists have been interested in the question of borders. Do national borders have any moral value? Do we owe more to fellow citizens of our own states than to the citizens of other states? But these questions are not often found in public discourse. What predominates in the mainstream media and among the largest political parties is a kind of certainty that sees the central actor of politics as a nation, distinct from and locked into endless competition with other nations. In a competitive world, ‘we’ must not lose out to ‘them’ — the vaguely hostile multitudes living beyond our protective borders but threatening constantly to traverse them.

This position corresponds broadly to a philosophical approach known as communitarianism. It emphasises the moral value of communities, which in the modern world are typically understood as sovereign states. It finds the borders that protect those states to be morally just. And the aim of foreign policy is basically to maintain or improve a country’s position vis-à-vis the positions of other states. Here is Winston Churchill explaining in 1914 that ‘we’, the British: ‘... are not a young people with an innocent record and a scanty inheritance ... We have engrossed to ourselves an altogether disproportionate share of the wealth and traffic of the world. We have got all we want in territory, and our claim to be left in the unmolested enjoyment of vast and splendid possessions, mainly acquired by violence, largely maintained by force, often seems less reasonable to others than to us.’ (Darwin, 2009). Churchill’s remarks were not intended to be self-critical. As a politician, he saw his role as helping to maintain Britain’s ‘disproportionate share’ rather than bring it into question.

Communitarian attitudes dominate conversations about immigration in Britain. If public opinion is divided on the question — polls show complicated and shifting attitudes, depending on how the question is framed — there is much less complexity in electoral politics. Politicians generally compete in their expressions of hostility to new arrivals. After winning the 2024 general election, the Labour party swiftly announced ‘new measures to cut historically high levels of net migration.’ At the Conservative party conference in October 2024, the contenders for party leadership all stressed their anti-migrant attitudes. ‘The world and his wife and their extended family...’ cannot come to Britain, said Robert Jenrick. ‘The age of mass migration must end.’ On X, he promised to ‘deport foreign criminals, get terrorists off our streets, and end illegal migration’. Immigrants who bring ‘foreign conflicts’ with them should not come to Britain, said Kemi Badenoch, who

eventually won the leadership contest. She added that ‘not all cultures are equally valid’ and that some immigrants bring with them ‘ancestral ethnic hostilities.’ James Cleverly, a former government minister, said that ‘dozens’ of asylum seekers would now be in Rwanda if he and his colleagues had remained in power. All these are strongly communitarian arguments. They animate much of the public and media discussion on migration through the use of terms like ‘secure borders’, ‘small boats’, ‘asylum seekers’, ‘illegal immigration’, ‘foreign criminals’, and ‘deportations’ — all of which stress the idea of an embattled political community. I am talking about Britain, but the story is a similar one in other Western countries like the US, France, Germany and Australia.

And yet political theory and philosophy, by which I mean critical and sustained reflection about politics that is not confined to the logic of policymaking, shows many alternatives to the communitarian position. Some argue straightforwardly for a radical cosmopolitanism that treats all people across the world as moral equals. Others focus on how nations have historically come to be imagined, and what is included and excluded from those acts of imagination. This latter view is, I suggest, a particularly useful one for thinking about the question of borders and immigration. With respect to Britain, it shows how the state was constituted through — and cannot be separated from — a form of hierarchical engagement with the world in the form of imperialism. A public discourse might paint Britain as an embattled political community. But this elides the fact that, as Churchill suggested, Britain is, on balance, the beneficiary of a global order that it constructed in its own interests. And this makes the idea of Britain as a bounded entity unconvincing in both historical and moral terms.

Such arguments have long been articulated by colonised peoples. ‘It was the sweat of the black man’s brow which laid the foundation for the present day opulence of Britain and White America,’ wrote an anonymous Sierra Leonean editorialist in 1938. The editorialist compared the relationship of Britain and Sierra Leone to that of ‘the cheat’ with ‘his victim’, and insisted the peoples of West Africa had a right to reparations after suffering for centuries from enslavement, colonisation, and forced transportation (“Rambling Talks,” *Sierra Leone Weekly News*, July 2, 1938.). Similar arguments across anticolonial writing saw Britain as a country whose present existence cannot be separated from its history as the centre of what was once the world’s most expansive empire. As Kojo Koram wrote recently: ‘Rather than saying Britain had an empire, it would be more accurate to say that the empire had Britain.’ (Koram, 2024).

A historically informed reflection on national identity opens up ways of thinking about politics beyond the reflexive defensiveness that characterises so much of contemporary discourse. It also throws into question the boundaries of the body that is imagined as the subject of politics. The categories through which our current political language is articulated depend on the idea of a separation between the domestic and international. Yet anticolonialists have long challenged that separation. In its place, they have insisted on the co-constitution of what came to be known as the global north and south. As Fanon put it: 'Europe is literally the creation of the Third World' (Fanon, 1963). Since the encompassing of the earth by western empires, national groups have always been imagined as elements of world-spanning projects. This is true for anticolonial nationalisms no less than colonial ones.

There are of course those who will bristle at any attempt to challenge the deeply rooted idea of national interest. 'Some people love to talk our country down,' said James Cleverly during the Conservative leadership contest. But 'this country has given so much to so many people.' It is easy to imagine the retort to this claim. The left is more patriotic than the right because its vision of 'our country' is more egalitarian and less beholden to elite interests. Yet each side of this debate, which dates back at least to the French Revolution, is circumscribed by the same spatial demarcation that refuses to accept how 'we' are connected to 'them', those beyond our borders who are excluded from the conversation even as they so often constitute its obsessive subject.

In the popular imagination, political theory is written by academics who are mostly based in universities in Europe and the US. But there are many traditions of serious thought about politics — including those that emerge from anticolonial traditions, social movements and radical political formations — that are not included in this definition of political theory. Those traditions are vital in understanding the injustices of the world and the necessity for its reorganisation and reconstruction. The events of recent years have acted as a sharp reminder of the rigid hierarchies that still govern international politics and the extreme difficulty for the vast majority of the world's peoples of resisting imperial power even at its most shocking and rapacious. Since the advent of European socialism, some people have criticised these hierarchies from the perspective of working populations of the imperial core. They have suggested, for example, that workers in Europe and North America would benefit from diverting funding from military uses to hospitals, schools and other social services. These arguments are convincing, of course. And yet — for all their well-

intentioned nature — they reproduce the idea that the people who really count are those who constitute part of ‘our’ collective national body.

When we consider the ways in which Britain was historically constituted as an empire, and is still constituted today as part of an imperial formation, it is much less easy to justify any such political demarcation that separates Britain’s population from the subjects of its ex-colonies. One of Edward Said’s key contributions was to show how national identities, especially those forged in a colonial context, implicate other peoples in ways that cannot be forgotten or ignored. This means that there is simply no realistic concept of what Britain is that does not take into account its expansion across so much of Africa, Asia, and the Caribbean, not to speak of North America, Australia and New Zealand, in a centuries-long process that created not just a single nation or state but many different states locked into a global — and extremely unequal — system of production and exchange. Palestine is often described as relevant to Britain because of Britain’s history of colonisation of the region. This is true. Yet we also have to insist — looking to Gaza, Sudan, the Congo, Haiti — that the history of colonialism still lives with us in the present, in a singular world that our prevailing discourse wrongly depicts as compartmentalised.

Notes:

- 1 An interview with seven leading international experts on genocide found that in relation to Israel’s destruction of Gaza “the question is not controversial—even for those who previously rejected the label” (NRC, May 14, 2025).

References:

- Darwin, J. (2009). *The Empire Project*. Cambridge: Cambridge University Press.
- Fanon, F. (1963). *The Wretched of the Earth*. New York: Grove Press.
- Koram, K (2024). Kemi Badenoch should read some Edmund Burke. [online] *New Statesman*. Available at: <https://www.newstatesman.com/ideas/2024/04/kemi-badenoch-should-read-edmund-burke> [Accessed 20 Feb. 2025].

Author biographies

Author biographies

David Archard

David Archard is Emeritus Professor of Philosophy at Queen's University Belfast, having previously taught at the Universities of Ulster, St Andrews and Lancaster. He has published extensively in applied ethics, political philosophy, and philosophy of law, especially on the topics of children, the family, sexual consent, and the public role of philosophy. His book *Children: Rights and Childhood*, now in its 3rd edition, was the first book philosophically to analyse the moral and political status of children.

He has been Honorary Chair of the Society for Applied Philosophy and is its Vice-President. For twelve years he was a Member of the Human Fertilization and Embryology Authority, and latterly its Deputy Chair. He was Chair of the Nuffield Council on Bioethics, the United Kingdom's de facto national ethics committee (2016-2023), is Chair of the Ethics Advisory Group of the 'Every Story Matters' exercise of the UK's COVID Public Inquiry, and a Member of the Clinical Ethics Committee of Great Ormond Street Hospital.

Shuk Ying Chan

Shuk Ying Chan is a Lecturer in Political Theory in the Department of Political Science at University College London. In her recent work, she has been interested in questions of colonialism and decolonization, empire and race, and reparative justice for historical injustices. Her monograph, *Postcolonial Global Justice* (Princeton University Press), was published in August 2025.

Prior to joining UCL in 2023, she was a Postdoctoral Prize Research Fellow at Nuffield College and the Department of Politics and International Relations at Oxford University. She completed her PhD in political theory at Princeton University. Her dissertation was awarded the 2022 Leo Strauss Award (for best dissertation in political philosophy) from the American Political Science Association. Originally from Hong Kong, she holds an MPhil in Political Theory and a BA in History and Politics from the University of Hong Kong.

Lorna Finlayson

Lorna Finlayson is a lecturer in philosophy at the University of Essex, and was previously a junior research fellow at King's College, Cambridge. She has published on a range of topics in political philosophy and its methodology, including feminism, political realism, and theories of ideology. She has also written for popular outlets such as the *London Review of Books*, the *New Left Review*, *The Guardian* and *The New Statesman*, on topics including universities, the Labour Party, free speech, animal rights, schools, and the NHS. She is the author of two books, *The Political Is Political* (London: Rowman and Littlefield International 2015) and *An Introduction to Feminism* (Cambridge: Cambridge University Press 2016). A third book, on the politics and philosophy of childhood, is under contract with Penguin Press.

Tariq Modood

Tariq Modood is Professor of Sociology, Politics, and Public Policy at the University of Bristol, and founding Director of the Bristol University Research Centre for the Study of Ethnicity and Citizenship. He has held over 40 grants and consultancies (UK, European and US), authored and edited over 35 books and reports, and published over 250 articles or chapters in political philosophy, sociology and public policy. He is the co-founding editor of the international journal, *Ethnicities*, and has served on the Commission on the Future of Multi-Ethnic Britain, the National Equality Panel, and the Commission on Religion and Belief in British Public Life. He was elected a Fellow of the British Academy in 2017.

Debra Satz

Debra Satz is the Vernon R. and Lysbeth Anderson Dean of the School of Humanities and Sciences at Stanford University, the Marta Sutton Weeks Professor Philosophy, and, by courtesy, Political Science. Her research focuses on the ethical limits of markets, objections to inequality, and issues in practical ethics. In 2004, Satz received the Walter J. Gores Award, Stanford's highest teaching honor.

Among her publications are *Why Some Things Should Not Be for Sale: The Moral Limits of Markets* (Oxford University Press, 2010); *Economic Analysis, Moral Philosophy and Public Policy* (Cambridge University Press, 2016) (with Dan Hausman and Michael McPherson); *Ideas That Matter: Democracy, Justice, Rights* (Oxford University Press, 2019) (with Annabelle Lever) and was a coauthor of the Institute for Fiscal Studies *Deaton Review: Dimensions of Inequality*, published by Oxford Open Series in 2024. A co-authored follow up to the *Deaton Review, Changing Inequalities: Causes, Concerns and Consequences* is forthcoming in 2026. She is a member of the American Academy of Arts and Sciences.

Adam Swift

Adam Swift is Professor of Political Theory /Philosophy in the Department of Political Science at University College London. He previously taught at Balliol College, Oxford — where he founded the Centre for the Study of Social Justice — and the University of Warwick. His book *Political Philosophy: A Beginners' Guide for Students and Politicians*, now in its 4th edition, has been translated into nine languages.

He has an M.Phil in Sociology from Nuffield College and is especially interested in interdisciplinary and collaborative work, often combining normative philosophy with empirical social science. He has published on a wide range of topics but has come to specialize in questions about social mobility, social justice, education and the family. His books include *How Not To Be A Hypocrite: School Choice for the Morally Perplexed Parent*, *Family Values: The Ethics of Parent-Child Relationships*, *Educational Goods: Values, Evidence and Decision-Making* and *How To Think About Religious Schools: Principles and Policies*.

James Wilson

James Wilson is Professor of Philosophy at UCL, and co-Director of the UCL Health Humanities Centre. He has pioneered a socially engaged and reflexive approach to philosophy, which not only brings existing philosophical research to bear on practical problems, but also examines how detailed understanding of practical problems reveals gaps and conceptual problems in existing philosophical theories. He has published extensively in applied and engaged ethics, including influential articles on health inequalities, healthcare resource allocation, paternalism, methodology in normative thinking, intellectual property, and the right to privacy. He is author of *Philosophy for Public Health and Public Policy: Beyond the Neglectful State* (OUP, 2021).

Wilson is joint editor of the philosophy journal MIND, and is a trustee of the Royal Institute of Philosophy. He has been a member of the National Data Guardian's Panel of adviser since 2016, and Deputy Chair of the Metropolitan Police Service's research ethics committee since 2020.

Jonathan Wolff

Jonathan Wolff is Emeritus Alfred Landecker Professor of Values and Public Policy, Blavatnik School of Government, University of Oxford. Previously he was Dean of Arts and Humanities and Professor of Philosophy at UCL. His books include *Robert Nozick* (1991), *An Introduction to Political Philosophy* (1996, 4th edition 2023), *Why Read Marx Today?* (2003), *Disadvantage* (with Avner de-Shalit 2007), *Ethics and Public Policy* (2011, 2nd edition 2019), *The Human Right to Health* (2012), *An Introduction to Moral Philosophy* (2018, 3rd edition 2025) and *City of Equals* (with Avner de-Shalit 2023). He is particularly interested in the relation between values and public policy, and has worked with The Home Office, Department of Culture Media and Sport, Railway Safety and Standards Board, The Nuffield Council on Bioethics, and the World Health Organisation. He is a Fellow of the British Academy and President of The Royal Institute of Philosophy.

Clare Chambers

Clare Chambers is Professor of Political Philosophy and a Fellow and Dean of Jesus College, University of Cambridge. She is the author of *Freedom & Equality: Essays in Liberalism and Feminism* (OUP, 2024); *Intact: A Defence of the Unmodified Body* (Penguin, 2022); *Against Marriage: An Egalitarian Defence of the Marriage-Free State* (OUP, 2017), which won the 2018 David Easton Prize of the American Political Science Association; *Sex, Culture, and Justice: The Limits of Choice* (Penn State Press, 2008); *Teach Yourself Political Philosophy* (with Phil Parvin, Hodder, 2012); and numerous articles and chapters on political philosophy, gender, and bioethics. Clare Chambers is also the editor of *The Routledge Handbook of Philosophy of Sex and Sexuality* (with Brian D. Earp and Lori Watson, Routledge, 2022) and co-Editor-in-Chief of *Res Publica*, the journal of legal, moral, and social philosophy (with Sune Laegaard).

Clare Chambers regularly appears on BBC Radio, and her research has featured in print and online media as diverse as *The Guardian*, *The New Statesman*, *El País*, *la Repubblica*, *iNews*, *The*

Times of India, the *Times Literary Supplement*, *The American Conservative*, *The New Humanist*, *Metro*, *Slate*, *Philosophy* 24/7, *Philosophy Bites*, *Daily Nous*, and *Aeon*. She was a Council Member of the Nuffield Council on Bioethics from 2020-25.

Anca Gheaus

Anca Gheaus is an Associate Professor in the Political Science Department at the Central European University in Vienna. She is a political philosopher interested in justice and the normative significance of personal relationships. She co-authored the book *Debating Surrogacy*, which came out at the Oxford University Press in 2024, and co-edited *The Routledge Handbook of the Philosophy of Childhood and Children*, published in 2018 by Routledge. She is also the author of numerous journal articles and book chapters, primarily on issues concerning childrearing, gender justice, non-ideal theory and methodological issues in political philosophy. At the moment she is working on a monograph on child-centred childrearing, under contract with Oxford University Press.

Katharine Jenkins

Katharine Jenkins is Professor of Philosophy at the University of Glasgow. She previously taught at the University of Nottingham, and held a Junior Research Fellowship at Jesus College, Cambridge. She has published widely in social philosophy, especially feminist philosophy and social ontology, and is the author of a monograph, *Ontology and Oppression: Race, Gender and Social Reality*, which explores how social categories such as races and genders exist and how these categories are bound up with systematic injustices. She has also published a short book for a general audience, *Feminist Philosophy: A Very Short Introduction*. She is the co-director of the Society for Women in Philosophy UK, and (for the period of 2024-2028) a Co-Editor-in-Chief of the academic journal *Hypatia: A Journal of Feminist Philosophy*.

Winnie Ma

Winnie Ma is a J.D. candidate at the University of California, Berkeley, School of Law, with particular interests in civil rights and ethical AI innovation. As a law student, she has worked at UC Berkeley's Samuelson Law, Technology & Public Policy Clinic, and has clerked at the Youth Law Center, whose mission is to transform the United States' juvenile justice and child welfare systems. Formerly, she was Lecturer in Philosophy at King's College London, specializing in the epistemology and ethics of AI, including medical algorithms, and the ethics of belief of stereotyping more generally. She has taught courses ranging from epistemology to the politics of health and medicine at King's College London in the Departments of Philosophy and of Global Health & Social Medicine, as well as the medical school. And she was formerly Research Associate and Project Manager at the Sowerby Philosophy & Medicine Project, working to introduce philosophy into the curricula that train clinicians and encourage interdisciplinary dialogue and collaborative research.

Martin O'Neill

Martin O'Neill is Professor of Political Philosophy at the University of York, having previously taught at the Universities of Cambridge and Manchester. He has published widely on the theory and practice of social and economic justice, in journals including *Philosophy & Public Affairs* and *Political Philosophy*. Among his books is *The Case for Community Wealth Building* (Polity Press, 2019), co-authored with Joe Guinan. He currently serves on the executive committee of the British Philosophical Association, and is a member of the Council of the Royal Institute of Philosophy.

Martin is a member of the Trustee Board of the Democracy Collaborative (a 'think-and-do tank' based in Washington DC), and previously served as a member of the UK Labour Party's Community Wealth Building Unit. He has consulted for international organisations including the International Labour Organization (ILO) and the World Economic Forum (WEF), and his work has been supported by funders

including the Institute for New Economic Thinking (INET), the Joseph Rowntree Foundation, the British Academy, the Leverhulme Trust, and the Interdisciplinary Social Research Foundation (ISRF).

John Tasioulas

John Tasioulas is Professor of Ethics and Legal Philosophy at the University of Oxford and inaugural Director of Oxford's Institute for Ethics in AI. He has previously taught at Glasgow, King's College London, and University College London, and held visiting appointments at the University of Chicago, Harvard University, and the University of Notre Dame. He has published extensively in moral, political, and legal philosophy, with a focus on human rights, punishment, international law, and the ethics of AI and digital technology. He is the author of *On Justice and Mercy: Essays in Moral and Legal Philosophy* (forthcoming, OUP) and the editor of *The Cambridge Companion to the Philosophy of Law* (CUP, 2020).

He has acted as a consultant on human rights to the World Bank and served as a member of the International Advisory Board, Panel for the Future of Science and Technology (STOA), European Parliament and as a member of the Prime Minister of Greece's High-Level Advisory Committee on AI. He is a Senior Fellow in Schmidt Sciences's AI2050 program and is currently engaged in an AI2050 project on humanistic AI ethics together with Professor Hélène Landemore (Yale University).

John Broome

John Broome is Emeritus White's Professor of Moral Philosophy at the University of Oxford, Honorary Professor at the Australian National University, and an Emeritus Fellow of Corpus Christi College, Oxford. He was previously Professor of Philosophy at the University of St Andrews and Professor of Economics at the University of Bristol.

He works on the philosophy of normativity and reasoning, and on the philosophy of climate change. He was a Lead Author of the Fifth Assessment report of the Intergovernmental Panel on Climate

Change. His most recent books are: *Climate Matters* (2012), *Rationality Through Reasoning* (2013), and *Normativity, Rationality and Reasoning* (2021).

Broome is a Fellow of the British Academy, the Australian Academy of the Humanities, and the Royal Society of Edinburgh. He is a Member of the Swedish Royal Academy of Sciences and the American Academy of Arts and Sciences. He is an Honorary Fellow of Trinity Hall, Cambridge, and holds an Honorary Doctorate from Lund University.

Sarah Bufkin

Sarah Bufkin is Assistant Professor of Political Theory at the University of Birmingham. She works on Black Atlantic political thought and Critical Theory, with a specific focus on racism and racialization, imperial techniques of government, and racial capitalism. Prior to Birmingham, she was an Examination Fellow at All Souls College, Oxford, where she completed her doctorate in political theory. Her writing has been published in *Cultural Studies*, *Philosophy & Social Criticism*, and *Identities*, among other journals. Sarah is a book review editor at *Perspectives on Politics*, an editor at *Contemporary Political Theory*, and a member of the editorial advisory group at *Soundings*. She is currently writing a book theorizing Frantz Fanon for the afterlife of empire. More broadly, she is interested in anticolonial thought, feminisms, critical social theory, continental philosophies of self and Other, and critiques of the carceral state.

Roger Crisp

Roger Crisp is Director of the Uehiro Oxford Institute, and Professor of Moral Philosophy at the University of Oxford. He has long-standing interests in the nature of ethics, ethical theories, and real-life ethical problems. With Tony Hope and others, he helped to establish in 1995 one of the first UK Clinical Ethics Committees, in the Oxford Radcliffe Hospitals Trust. He is the author of *Mill on Utilitarianism*, *Reasons and the Good*, *The Cosmos of Duty: Henry Sidgwick's Methods of Ethics*, and *Sacrifice Regained: Morality*

and Self-interest in British Moral Philosophy from Hobbes to Bentham. He edited the *Oxford Handbook of the History of Ethics*, and translated Aristotle's *Nicomachean Ethics* for Cambridge University Press. He is currently translating and commenting on three of the Platonic dialogues concerned with the death of Socrates, as well as thinking, when time allows, about Buddhism and personal identity.

Katrien Devolder

Katrien Devolder is Professor of Applied Ethics and Director of Public Philosophy at the Uehiro Oxford Institute and GB Fellow at Reuben College, University of Oxford. She has published two monographs: one on compromise positions in the embryonic stem cell debate (OUP, 2015) and one on the ethics of human cloning (Leuven University Press, 2001). She has also published numerous papers on ethical issues pertaining to gene editing, gamete donation and genetic parenthood, life extension, compromise in bioethical debate, moral complicity, animal ethics, conscientious objection, and euthanasia in prisoners. Her most recent work focuses on the concept and ethics of laziness. She also produces (conducts, films and edits) interviews with academics to make complex ethical debates accessible to a wide audience which can be viewed on YouTube's 'The Practical Ethics Channel' or listened to on Apple Podcasts ('Thinking Out Loud').

Jonathan Pugh

Jonathan Pugh is a Senior Research Fellow at the Uehiro Oxford Institute. He is also the Theme Lead for Values and Society at Reuben College, University of Oxford. His research interests lie primarily in issues concerning personal autonomy in practical ethics, particularly topics pertaining to informed consent. He was previously a member of the UK Pandemic Ethics Accelerator, and the Parfit-Radcliffe Richards Fellow at the University of Oxford. He is the author of *Autonomy, Rationality and Contemporary Bioethics* (OUP, 2020), and

he has also published work in philosophical and medical journals on a number of topics in neuroethics, research ethics, public health ethics, the ethics of psychiatry, and clinical ethics. He is currently the Principal Investigator on the project 'Predictive Technologies, Ethics, and the Future of Insurance', and the academic lead on the 'Decision Aid for the Restitution of Cultural Artefacts' project.

Cécile Fabre

Cécile Fabre

Cécile Fabre is Senior Research Fellow in Politics at All Souls College, Oxford. She holds degrees from La Sorbonne, the University of York, and the University of Oxford, and had held positions at the LSE and the University of Edinburgh. Her research interests include theories of distributive justice, the rights we have over our own body, and the ethics of foreign policy. She has published papers in (amongst others) *Ethics*, *Law and Philosophy*, *The Journal of Political Philosophy* and the *British Journal of Political Science*. Her books include *Social Rights under the Constitution* (OUP 2000), *Whose Body is it Anyway?* (OUP 2006), *Cosmopolitan War* (OUP 2012), *Cosmopolitan Peace* (OUP 2016), *Economic Statecraft* (Harvard UP 2018), and *Spying Through a Glass Darkly* (OUP 2022). She delivered the Tanner Lectures on Human Values at Stanford in 2022 on the ethics of preserving cultural heritage, and is expanding the lectures into a monograph. She is a member of Academia Europaea and a Fellow of the British Academy.

Elke Schwarz

Elke Schwarz is Professor of Political Theory at Queen Mary University London, UK. She holds a PhD from the London School of Economics and Political Science (LSE), an MA in Conflict Studies from the War Studies Department at King's College London (KCL) and a Bachelor in Business Studies from Belmont University (USA). She has published extensively on the ethics of technology and warfare, with a specific emphasis on new and emerging military technologies, including military Artificial

Intelligence (AI), autonomous weapon systems, drones and robots. She is the author of *Death Machines: The Ethics of Violent Technologies* (Manchester University Press). Elke is currently Vice-Chair of the International Committee of Robot Arms Control (ICRAC) and also an Associate with the Imperial War Museum (IWM) and an RSA Fellow. She held a 2022/23 Fellowship at the Center for Apocalyptic Studies (CAPAS) and was a 2024 Leverhulme Research Fellow.

Musab Younis

Musab Younis is Associate Professor of Political Theory and Jarvis Doctorow Tutorial Fellow in Politics at St Edmund Hall, University of Oxford. Previously, he worked at Queen Mary University of London (QMUL) as Senior Lecturer in the School of Politics and International Relations. He completed his doctorate at St Catherine's College, Oxford, and has held visiting positions at the European University Institute and the University of London Institute in Paris.

His research explores political theory in relation to race and empire, with a focus on the history of anticolonial thought. He is the author of *On the Scale of the World: The Formation of Black Anticolonial Thought* (University of California Press, 2022), which won the Sussex International Theory Prize and was named as one of the best academic books of 2023 by *The New Statesman* magazine. In addition to his academic writing, he is a regular contributor to the *London Review of Books*. His essays have also appeared in *n+1*, *Prospect*, *Baffler*, and *The Guardian*.

Acknowledgements

This volume was co-edited by Professor David Archard and Matt Davies.

Ben Brown, Harry Farmer, Mahi Hardalupas, Martin Davies, and Victoria Briggs provided substantive contributions and editorial support. Copyediting was carried out by Rebecca Ghani and design work by Eureka! Design Consultants. The production of this volume would not have been possible without extensive project management support from Ben Brown, and important contributions from other Nuffield Foundation colleagues including Raf Bogan, Michelle Wuisan, Nicola Webb and Victoria Briggs. The editors are grateful to all of the above for their work on this project.

Tim Gardam, former CEO of the Nuffield Foundation, provided the initial impetus for this project. The editors would like to thank Tim and his successor as Nuffield CEO, Gavin Kelly, for their role in making this publication a reality.

