This is the text accepted for publication in the *Journal of Applied Philosophy* (July 2023). Where possible, please cite the published version.

-

Nicholas Barrow

Robot Ethics

M. COECKELBERGH, 2022.

Cambridge, MIT Press.

vii + 191 pp, $16.95 (pb)

In *Robot Ethics*, Mark Coeckelbergh offers an insightful, critical, and distinguished tour of the ethical issues concerning robots. Intended as an introduction – and as part of MIT Press' essential knowledge series – *Robot Ethics* is an accessible and concise overview of a topic with imminent importance. Robotics will (and already has) fundamentally changed our lives, society, and environment. There is therefore a sense of urgency – and perhaps duty and responsibility – to evaluate its impacts before it is too late. This book serves as an exemplary resource, illustrating the wide array of ethical considerations robots pose without straying too far into the hypothetical, using an impressive array of conceptual techniques and clarity of writing. It is perfect for undergraduates (and would make a very good teaching resource), policymakers, or anyone interested in the ethics of emerging technologies.

Media and science fiction often depict robots as futuristic, set against the backdrop of some dystopian future where we have succumbed to our robot overlords. Coeckelbergh is very quick to dispel this representation: the robots of the future are here, now…just not as we imagined them. They are not monstrous 'but instead cute, helpful, entertaining, funny, and seductive' (p. 3), they are designed to be social, to 'fit in' with and help humans, and facilitate human-robot interaction. Nor are their threats existential (with the exception of perhaps killer drones). Instead, they raise more 'mundane' worries: for example, those concerning job security (as explored in chapter 2), deception (chapter 3), and elderly care (chapter 4).

Coeckelbergh distinguishes two dimensions to his exploration. Firstly, he is concerned with the ethical and societal questions raised by robots. He does this by using both a range of conceptual tools from the wider ethics literature and practical examples intended to help governance and responsible development policies. Secondly, Coeckelbergh is interested in using robots as a medium to inform deeper philosophical reflection on what is to be human. While discussing the immediate ethical concerns raised by robots is important, Coeckelbergh argues that a broader

perspective is needed to fully appreciate the societal implications of robots and why there is such widespread public fascination with them.

As an introduction, each chapter covers a different topic within the field of robot ethics. In the following, I provide a brief summary of each. After doing so, I critically engage with the final chapter where Coeckelbergh proposes an original thesis.

Chapter 2, 'Industrial Robots, Safety, and the Future of Work', considers the impact of intelligent robots taking away tasks from humans. Going beyond automation and the rise of capitalist injustices, Coeckelbergh offers reflections on universal basic income, the meaningfulness of work, and the physical and mental safety of workers that work in close proximity to potentially dangerous robots.

In Chapter 3, 'Robotic Home Companions, Privacy, and Deception', Coeckelbergh explores the implications of 'social robots', such as those designed to care for the elderly and children robots, that enter our homes. The primary concern in this chapter is privacy: what data do these robots collect? However, another theme is whether 'simulated' care can amount to human care, and whether it is wrong to deceive vulnerable users into thinking their robot companions *really* care.

The fourth chapter, 'Care Robots, Expertise, and the Quality of Health Care', asks whether robot nurses and surgeons can provide *good* care, prompting further discussion about what 'good care' consists of in general.

Chapter 5, 'Self- Driving Cars, Moral Agency, and Responsibility', tackles questions of whether robots can be moral agents (whether they can take moral ownership and be responsible for their actions). If they can, what values should they be aligned with? Which moral theory should guide their actions? What does it mean for a robot to be responsible?

The following chapter, 'Uncanny Androids, Appearance, and Moral Patiency', asks the inverse: whether robots can be the source of moral obligation. When (if ever) ought we to treat robots with moral consideration? What should ground this decision? What about robots that *look* like humans? What does this mean for moral standing more generally?

Chapter 7, 'Killer Drones, Distance, and Human Existence', evaluates the ethics of using robots in warfare. Further, it explores what the intuitions we have

about robot warfare might mean for warfare in general like the necessary and sufficient conditions for *just* war, ethics of killing, and whether machines could ever fully replace humans in warfare.

Coeckelbergh's concluding chapter 'Robotic Mirrors beyond the Human: Robot Ethics as an Environmental Ethics' begins by explaining how robots are mirrors of humanity. Thankfully, the author does not mean this literally: a quick Google search of 'robot' will quickly amass rows upon rows of white male robots that acutely misrepresent what humanity 'looks' like. Instead, it is the observation that we use robots as tools of comparison: to determine what we are. Coeckelbergh argues that this angle often results in negative comparisons: defining humans in terms of what they are not (not machines, not robots, not animals). Further, he suggests that by doing this we necessitate tension and competition whereby we defend the *human* against the *machine* – especially in the West whereby we think being human has some special status: robots should not take *human* jobs; robots are *human* tools; etc. Accordingly, we should go *beyond* the human. What does this look like? From a transhumanist outlook, it means accepting that humans are a type of robot and that they will eventually supersede us. And on a posthumous outlook, it means accepting human-robot cooperation and hybridity. The author thinks these positions are inadequate, suggesting that we consider robot ethics as an *environmental* ethics.

The author's aim is positive. Coeckelbergh's position is that as technologies of the future, robots ought to be aligned with future issues: in particular, ecological sustainability. This entails not just that we create environmentally friendly robots (weak) but that we consider what robotics should be striving for overall: ecological sustainability (strong). This involves asking what robots we ought to make (if at all), for what purpose, and in what capacity.

However, their approach requires further clarification. First, is this not a distinction without a difference? If we turned our attention to the author's preferred approach, considering what robots we ought to be pursuing in light of ecological sustainability (strong), would this not lead to creating environmentally friendly robots (weak)? It appears as though there fails to be a sufficiently explicit difference here. Second, whilst going *beyond* the Anthropocene it still seems to have it at its centre. Consider the following passage (p.207):

This is the text accepted for publication in the *Journal of Applied Philosophy* (July 2023). Where possible, please cite the published version.

-

Nicholas Barrow

> ‘What kind of artificial creatures do we need, if we need them at all, given the environmental predicament we are in today?’

Who is the author referring to when they say ‘we’? The author needs to sufficiently disentangle human goals from ecological goals. If, indeed, this is possible. In making robots aligned with values that are strictly ecological, why shouldn’t they involve human interests? The author seems to suggest that ecological ends have intrinsic motivation. Yet, by invoking ‘we’ they seem more concerned with ecological values for humans’ sake. Whilst I appreciate the intention, I think that the author needs to further clarify exactly what an ethics of robotics as an environmental ethics entails.

Overall, *Robot Ethics* is a clear and engaging introduction to the field of robot ethics. It also serves as more than an introduction, carrying through it a clear ideation of how robot ethics should be thought about and conducted in general. *Robot Ethics* is also grounded. It does not speculate too far into the hypothetical and remains (I think, to its merit) focused on practical and immediate concerns.

NICHOLAS BARROW
University College London