Birgit Beck · Michael Kühler

Editors

Technology, Anthropology, and Dimensions of Responsibility

J.B. METZLER

Introduction

"With great power comes great responsibility." The origin of this saying has been attributed to such diverse sources as Voltaire, Winston Churchill, Franklin D. Roosevelt, and the Spider-Man comic book. Although its roots may remain in the dark of history, this does not diminish its importance and topicality. Technological progress has undoubtedly brought great power to humanity which has to be balanced by responsible action. Today, due to ever expanding technological power and the progress in knowledge about the impact of our interactions with our surroundings, humans are commonly considered to be responsible for many instances that a few centuries or even decades ago would have been deemed plainly fateful. We now regard ourselves as responsible not only for direct consequences of attributable, voluntary, individual actions, but also in a more abstract way for the preservation of our natural environment, the prevention of climate change, the fight against diseases, poverty and hunger, a decent living standard for future generations, and ultimately even for either the conservation or the technological advancement of human nature itself, the last option including the creation of genetically enhanced children, cyborgs, artificial intelligence, so-called post-persons and eventually the overcoming of death and this world's end.

Praiseworthy and heroic as this recent human self-ascription of responsibility for virtually everything may seem at first glance, from a philosophical perspective, it reveals the need for both conceptual clarification and thorough ethical assessment. The notion of responsibility is currently subject to extensive conceptual scrutiny and figures prominently in diverse debates in applied ethics such as, e.g., discussions about genome editing, human as well as moral enhancement, environmental ethics and big data processing. These debates not only address questions of a widened scope of responsibility due to technological progress, but also cover critical assessments of how possibilities of intentionally "designing" specific aspects of persons, like in the case of genome editing, cyborgs, or human or moral enhancement, might threaten our traditional notion of individual responsibility if no longer the person but rather the technology involved is deemed to be responsible for a specific behaviour and its consequences. Taken together, technological progress, therefore, not only appears to widen the scope of our responsibility, but it also seems to be capable of putting into question our very understanding of (human) responsibility and, thereby, our traditional anthropological self-understanding as the only beings capable of moral responsibility.

However, the term responsibility may refer to a number of different ideas, and it is indispensable for a sufficiently precise discussion to distinguish some key notions. Firstly, prospective or forward-looking responsibility should be distinguished from retrospective or backward-looking responsibility. The former is mainly concerned with moral obligations regarding our prospective actions and their consequences, especially taking into account our increasing power and the growing impact our actions may have. This notion of prospective responsibility has recently gained more importance within the field of technology and finds its expression in ideas such as *responsible innovation* or *responsible design*. The latter is concerned with our age-old practice of holding people responsible or praising and blaming them for their actions and the consequences of them.

Secondly, causal responsibility should be distinguished from moral responsibility. While the former may be attributed to any event or living being if it causally contributed to some event coming about, the latter is typically considered to require agents and their actions. For example, while the neighbour's dog may be said to be causally responsible for our cat to disappear in a split-second, we usually do not consider the dog to be morally responsible and especially not praise- or blameworthy for it, since the dog is not an agent in the full sense. Only human beings, and more particularly persons, are traditionally thought to be agents in this full sense and, thus, capable of being morally responsible for their actions and the consequences of these actions because the latter can be *attributed* to them. Analogously, only persons are thought to be capable of taking over prospective responsibilities.

However, the attributability of actions and consequences is not the only basis for our social practices of holding persons morally responsible and assigning prospective responsibilities. For, it is not only possible but also a widespread practice to take or ascribe moral responsibility for events or consequences even if these cannot be (directly) attributed to us, and the same holds for taking or assigning prospective responsibilities. A trivial example would be parents who are not only thought to be responsible for their children in the prospective sense but who are also retrospectively held responsible for the latter's actions. Still, this social practice of holding people responsible independently of attributability not only raises severe moral questions about how fair this practice can actually be, but also seems to reduce the notion of moral responsibility to the idea of strict liability. Strict liability, in turn, seems to preclude praise- and blameworthiness, since for these latter practices of praising and blaming to make sense or at least to be fair, apparently, attributability has to be presupposed. Accordingly, while it might be fair to hold parents liable for damages caused by their children, it does not seem to make much sense to blame them for their children's actions. After all, it was not the parents who acted. The background principle at work at this point is the so-called control principle: We are morally responsible, especially praise- or blameworthy, only for outcomes and to the extent that what happens is under our control.

Now, in expanding the scope of our control, recent technological developments apparently also expand the scope of our moral responsibility, which explains the first line of argument mentioned at the beginning. Yet, it may also be said that

recent technological developments do not only expand the scope of our control but also seem to be capable of reducing it in cases in which we rely on technology or even delegate control to it outright, like in the case of self-driving cars, autonomous weapons systems, or when we let algorithms make important decisions for us. Consequently, if such autonomous technology is, indeed, capable of "making decisions", it looks like our anthropological self-understanding of being the only beings capable of moral responsibility is put into question. However, if so, how should the idea of cars, weapons systems, or algorithms *being* morally responsible be spelled out in detail? Does it then also, for instance, make sense to *blame* a robot? If so, how exactly? Furthermore, assuming that we more and more delegate control to technology, how might this affect our self-understanding as responsible agents? Why should we take responsibility any longer in the first place if we may easily delegate it to technology? Yet, in doing so, could we lose some crucial aspects of what makes us (responsible) persons in the first place?

This volume aims at addressing these and related questions, both in terms of critical foundational discussions of the crucial concepts involved and in terms of engaging in a more "applied" debate about how pressing issues of recent technological developments in this regard should be addressed in practice. Taken together, the volume is meant to add to recent debate about the implications of technological progress for our self-understanding as (morally) responsible beings.

Contributions

In a concise and thought-provoking essay, Thomas Gil starts off the volume with entertaining the thought that we should abandon using the concept of responsibility altogether. The subsequent chapters, addressing various conceptual and practical challenges of the concept and its applications in our modern, technology-driven world, may be understood as attempts to prove this thought wrong. These chapters are divided into two sections:

1. Responsibility in Action,
2. Responsibility for Actions.

Responsibility in Action

The first section starts with the contribution of Janina Loh and Mark Coeckelbergh "Transformations of Responsibility in the Age of Automation: Being Answerable to Human and Non-Human Others". The authors note that especially in the realm of technology, big data, and new media it is questionable if our traditional understanding of responsibility is able to face current challenges—mostly due to its restricted focus on the autonomous, self-sufficient, individual human being as the genuine responsible agent. They engage with recent concerns about the possibility of ascribing responsibility to artificial systems or in cases in which we cannot

reduce responsibility to a limited and clearly defined group of responsible persons. In such cases, our conventional methods of identifying individual human agents as the solely feasible responsible agents—or as other important functions within the relational setup of the traditional concept of responsibility such as the addressee and the authority—frequently fail. Loh and Coeckelbergh argue that we need to question and move beyond a traditional understanding of responsibility in order to update it for these and further challenges.

In his contribution "Technology and Evolving and Contested Divisions of Moral Labour", Arie Rip explores the idea that responsibility is an open-ended concept, i.e. that its meaning or method of use is evolving. Yet, Rip wants to go a step further and argues that there is something like a responsibility language, and that it is evolving in terms of divisions of moral labour. In elaborating on the evolution of responsibility language, Rip shows how a corresponding division of moral labour has emerged. Accordingly, he concludes that if one wants to overcome traditional divisions of moral labour (for emancipatory reasons or because the present division of labour is not productive), other divisions of moral labour, including a suitably modified responsibility language, have to be envisaged and explored.

In her contribution "Infantilisation through Technology", Birgit Beck critically assesses persuasive technologies which are designed with the aim of influencing people's attitudes and behaviour in order to let them achieve their goals and realise their values more efficiently. It is commonly assumed that such technologies do not work in a malevolent, manipulative or coercive way, but rather function as self-administered nudges which promote rational, autonomous and perhaps even moral conduct and are therefore conducive to leading a good life. Beck argues that persuasive technologies—far from enhancing autonomy and the good life—might exert an influence on their users which she terms "infantilisation": Persuasive technologies might threaten our (self-)ascription of responsibility by treating average autonomous persons like children in need of normative and practical guidance. Beck assumes that this holds both on individualist and relational accounts of autonomy and might have a negative impact on exercising capacities of reflecting on and realising personal conceptions of the good life.

Joschka Haltaufderheide is concerned with the impact of a recent technological breakthrough on the concept of responsibility in his contribution "CRISPR-Cas and the Wicked Problem of Moral Responsibility". The emergence of CRISPR-Cas genome editing has raised severe ethical concerns. It has been suggested that CRISPR-Cas leads to a game-changing shift in biotechnology and, in turn, in ethical perspectives on the issue. However, it is less apparent what this shift might be about and how the advent of CRISPR-Cas changes ethical debates on genome editing and moral responsibility. Against this background, Haltaufderheide analyses the advent of the CRISPR-Cas technology from a perspective of moral philosophy, focusing on the features of the technology and a particular concept of moral responsibility that is usually employed in bioethics, which focuses on individual imputability of intentional actions and their foreseeable results in case of harm.

Haltaufderheide argues that the easy accessibility, efficiency and effectiveness of CRISPR-Cas exert pressure on the traditional concept of individual retrospective as well as prospective moral responsibility and requires a shift in moral perspective. He proposes to think of moral responsibility in broader terms, including more systematic accounts of forward-looking or prospective responsibility, which exceed the individual level, and lays out an agenda for such an account.

In his contribution "Seeing the Turn: Microscopes, Gyroscopes, and Responsible Analysis in Petroleum Engineering", Eric Kerr considers the role of perception in petroleum engineering. Specifically, he looks at data analysis practices in geological surveying, wellbore navigation, directional drilling and related techniques. Kerr's analysis of the topic begins with a familiar argument in the philosophy of science that there are no stable, non-stipulative grounds for distinguishing between ordinary cases of perception and cases where the perceptual system includes microscopes. It extends this claim to the use of gyroscopes and the data they generate. Finally, he explores the implications of this form of perception for the idea of responsible data analysis and concludes that such responsible analysis, in the ideal case, does not mean passing over control to software but trusting one's own expert judgment built up over decades which relies on close familiarity with the objects of concern.

In the aptly named final contribution to the first section, "We are the End of the World: Stories of Anthropocenic Hyperarousal", Axel Gelfert starts by noting that the science of climate change has long had to negotiate the tension between the demand for hard numerical data and the need for imagining radically different futures. In recent years, the notion of the 'Anthropocene'—that is, of a new geological epoch brought about by the cumulative effects of humans on the Earth's geochemical cycles—has opened up fruitful space for theoretical exploration of this kind. In his contribution, Gelfert focuses on literary manifestations of this anthropocenic imagination, both in the form of recent climate fiction ('cli fi') and its genealogical precursors. Drawing in particular on novels by Arno Schmidt, J.G. Ballard, and Erwin Uhrmann—all of whom discuss how human subjectivity is altered under conditions of transformative environmental change—he argues that our collective response to the hyperobjectual relations of the Anthropocene is best described as a state of 'anthropocenic hyperarousal'.

Responsibility for Actions

The second section starts with Thomas Grote's and Ezio Di Nucci's contribution "Algorithmic Decision-Making and the Problem of Control". The authors note that in the legal sector, as well as in public policy or medicine, decisions are increasingly being delegated to learning algorithms. They argue that these delegation practices involve trade-offs in terms of control and scrutinise these trade-offs from a normative point of view. In particular, they focus on two (potential) sources for loss of control: (i) epistemic dependence and (ii) the accountability gap. By drawing on the literature on testimony and moral responsibility, in addition to

discussing some of the basic concepts of machine learning, Grote and Di Nucci argue that relevant loss of control might shape the motivational structure of decision-makers in a way that is ethically problematic. Therefore, even under the assumption that learning algorithms make fairer or more objective decisions than human experts, associated costs stemming from the loss of control might yet make delegating high-stakes decisions to learning algorithms ethically questionable.

Michael Kühler's contribution "Technological Moral Luck" takes up the challenge of a loss of control for moral responsibility. According to him, it is a pervasive feature of today's life that we rely more and more on technology when making decisions. For example, we often "blindly" follow the instructions of navigation systems when driving. Letting the navigation system "take control" is precisely one of the main reasons to use such a technology in the first place because we usually do not have the time to determine the best route ourselves, especially given the current traffic situation. Moreover, we may even have developed a tendency to see ourselves as less responsible or even to shun responsibility altogether because of this lack of control, like when we say that it was not really us who made the decision but the navigation system. In his contribution, Kühler addresses this claim about our diminished or even lacking moral responsibility when relying on technology in our decision-making. For, as he notes, if it could be shown that by relying on technology we, indeed, lose a morally relevant form of control, but that we are or should be held responsible for our decisions and their consequences nonetheless, this moral practice would include more and more cases of moral luck, i.e. we would be held morally responsible for things beyond our control. Kühler proposes to dub such instances *technological moral luck* and argues that the stronger we understand the underlying control principle for moral responsibility, the more we will have to accept that moral responsibility becomes a matter of moral luck if we still want to hold agents morally responsible when they rely on technology.

In his contribution "Would Moral Machines Close the Responsibility Gap?", Peter Remmers notes that questions about *moral machines* are ever-present in contemporary discourse about robots and artificial intelligence. What would a machine that acts morally be like? Is it actually possible to build one? Should we work toward this goal and why? Would observable moral behavior be sufficient for a machine to count as a moral being, or would it need some 'subjective' foundation in its inner workings? In other words: Would it be enough for a 'moral machine' to behave *as if* it were sensitive to our moral affairs, or would it have to be able to sense and understand our moral concerns? Still, Remmers does not aim at a thorough overview of the discussion around the idea of moral machines and its role in the related problem of responsibility. Instead, he focuses on a few selected aspects connecting moral machines to the so-called responsibility gap. By introducing a number of clarifications and a certain perspective on the issues at hand, he answers the question of whether a technological issue should be addressed by approximating machines to humans in the negative. A reflection and debate on these and further similar issues would, therefore, have to shift its focus

from the isolated technological artifacts to the much wider contexts of application and socio-technological interactions.

In his contribution "Can We Forgive a Robot?", Michael Nagenborg takes the question whether it makes sense to praise or blame robots for their behaviour one step further. If we hold a robot fully responsible for its actions, how should we deal with that robot if it did something wrong? For example, if a robot murders a human being, should it be punished just like a human being who commits the very same crime? Can we actually punish a robot? While Nagenborg agrees that we need to think through how we can react to the wrongdoings of a robot *if* we are willing to hold the machine responsible for its actions, he explores a different perspective in his contribution. Instead of asking, if we can punish a robot, he asks if we can *forgive* a robot. The background of his inquiry is that forgiveness plays a crucial, yet often neglected role in human-human interactions. Therefore, it seems reasonable to assume that forgiveness will play a similar role in a society where humans and robots coexist and, at least, some of these robots are held responsible for their actions. His chapter is a speculative exercise to grasp what it could mean for human beings to live together with such machines and to demonstrate that "forgiving" provides us with an excellent lens to think through human-technology relations.

In his contribution "Artificial Intelligence in Extended Minds: Intrapersonal Diffusion of Responsibility and Legal Multiple Personality", Jan-Hendrik Heinrichs starts off with the question whether an artificially intelligent tool can be a part of a human's extended mind and highlights the interaction of two opposing streams of thought in this regard. One strand of thought can be identified as the externalist perspective in the philosophy of mind, which tries to explain complex states and processes of an individual as co-constituted by elements of the individual's material and social environment. The other strand is normative and explanatory atomism which insists that what is to be explained and evaluated is the behaviour of individuals. He argues that counterintuitive results turn up once atomism tries to appropriate insights from psychological externalism and holism. These results are made visible by technological innovations, especially artificially intelligent systems, but they do not result from these innovations alone. As Heinrichs attempts to show, they are rather implicit in situated cognition approaches which join both theoretical strands. This has repercussions for explanatory as well as ethical theorising based on situated cognition approaches. It is a fairly rare constellation, he concludes, that new technological options, namely artificial intelligence, raise doubts concerning a philosophical theory, namely extended mind theory.

In the final contribution to the volume, "Reproductive Medicine and Parental Responsibility", Tatjana Noemi Tömmel addresses the question whether and how assisted reproduction impacts parental responsibility against the background of three different perspectives: First, she takes a look at the relation between "anthropology" and "technology", namely the fear that reproductive medicine might change our concept of humanity. Second, she inquires whether it is really technology that alters the normative relationship between parents and children.

Do reproductive techniques demand new or special ethical *principles,* or do they confront us merely with new *situations*, which can be subsumed under existing ethical principles? As she defends the latter position, she focuses in the third part on parental responsibility in general, arguing that not the technology of assisted reproduction but parenthood as such poses *unique ethical problems* demanding philosophical attention. This allows her to come back to anthropological consid- erations, seen from the perspective of "responsibility". Although she does not want to make a case for virtue ethics in general, she argues that parenthood demands virtues, namely respect, love, and responsivity.