Editors’ Overview: Moral Responsibility in Technology and Engineering

Neelke Doorn · Ibo van de Poel

Received: 18 May 2011 / Accepted: 19 May 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

From its earliest issue back in 1995, responsibility has been a major theme in the Journal of Science and Engineering Ethics. Although the focus of attention and conceptions of responsibility vary considerably over the different articles, the prevalence of responsibility as a topic is remarkable. On a superficial reading, we could see it as the logical corollary of the importance of responsibility for the broader field of applied ethics. However, this can only partly explain why responsibility is such a widely discussed topic in the Journal of Science and Engineering Ethics. If we take, for example, the academic journals for medical ethics, responsibility is not half as frequently mentioned in the title or keywords.\(^1\) Apparently, there is something special about the relation between responsibility and technology and engineering.

To see what is special about responsibility in technology and engineering, one should perhaps start with the kind of paradigmatic situation that is usually the focus of attention in philosophical ethics, and even in most of the applied ethics literature. In this paradigmatic situation, there is usually an individual that is confronted with a difficult ethical choice. Although the choice at hand is morally complex, it is usually assumed that what raises ethical concerns are the possible actions of the individual and the direct consequences of these actions. Moreover, it is usually assumed that

\(^1\) A comparison the Journal of Medical Ethics, the Journal of Health Care Analysis, and Science and Engineering Ethics in the ISI database published in the period 2001–2010, shows that 60 out of the 221 research articles published in Science and Engineering Ethics have the word responsibility in the title, abstract or as keyword. For the Journal of Medical Ethics, this number is 112 out of 1.354 articles and for Health Care Analysis it is 25 out of 280.

N. Doorn (✉) · I. van de Poel
Department of Technology, Policy and Management, Delft University of Technology, P.O. Box 5015, 2600 GA Delft, The Netherlands
e-mail: N.Doorn@tudelft.nl
I. van de Poel
e-mail: I.R.vandepoel@tudelft.nl

Published online: 03 June 2011
such consequences are more or less certain. The ethical literature thus often assumes: (1) that it are individuals who act, (2) that the consequences of their actions are causally direct traceable, and (3) that these consequences are certain. None of these assumptions seem to apply to many of the ethical issues raised by modern technology and engineering.

First, engineering and technology development typically take place in collective settings, in which a lot of different agents, apart from the engineers involved, eventually shape the technology developed and its social consequences. Second, engineering and technology development are complex processes, which are characterized by long causal chains between the actions of engineers and scientists and the eventual effects that raise ethical concern. Third, social consequences of technology are often hard to predict beforehand.

Jonas (1984) has suggested that the three characteristics of technological action just mentioned (collectivity, indirect causation, uncertainty) require an ethics of responsibility rather than an ethics based on traditional ethical notions like consequences or duties. At the same time, it is clear that a lot of the traditional philosophical literature on responsibility still ignores the mentioned characteristics. Although philosophers have paid attention to the fact that consequences can be realized in a multiplicity of ways (e.g. Fischer and Ravizza 1998), their focus often is on rather direct consequences. Moreover, they tend to focus on responsibility for certain outcomes rather than on responsibility for risks. What has been recognized without doubt is the collective nature of action as witnessed by the vastly growing literature on collective responsibility (e.g. Smiley 2008; May and Hoffman 1991; French 1984).

There is yet another reason why responsibility is an important theme in engineering and technology. Technology increasingly influences the context in which humans have to act and so seems to co-shape their responsibility. An example of this mediating role is the use of automatic pilots in planes. Who is responsible if a plane with an automatic pilot crashes? Does it make sense to attribute responsibility in such cases to technology or is it possible to understand these complex situations so that all responsibility can in the end be attributed to humans? Examples like the automatic pilot also raise interesting questions about the responsibility of designers. It may be argued that they design the technological environment in which others have to act and to exercise certain responsibilities. Can, and should, designers design technological systems so that they enhance rather than limit the responsibility of the users of these systems; and what does such an obligation exactly entail?

The articles collected in this issue have been presented at the conference “Moral Responsibility: Neuroscience, Organization & Engineering,” which was held in Delft, August 25–27, 2009. The central aim of this conference was to improve our understanding of responsibility in our current society. By drawing on philosophical and scientific expertise and insight, the discussion moved beyond the traditional discussion on free will and determinism, which has dominated the philosophical responsibility debate. One of the conclusions of the conference was that there is not one unitary, agreed-upon definition of responsibility but that there are many different meanings attached to the notion of responsibility, some being merely descriptive, others having an explicitly normative content. By reflecting on these
diverse meanings, this special issue of Science and Engineering Ethics provides a topical overview of how the theme responsibility in technology and engineering is currently approached by different scholars in the field.

Given the different meanings attached to the notion of responsibility, it is important to start with some conceptual clarifications. In fact, this is exactly what is done in the contributions by Davis (forthcoming), Kermisch (forthcoming), Van de Poel et al. (forthcoming), and Doorn (forthcoming). Davis provides the most fine-grained distinctions. Without revealing the full taxonomy here, we could say that the different senses of responsibility can be classified along two main dimensions. The first one is a temporal one, referring to the distinction between forward-looking responsibilities (that is, a sense of responsibility that refers to things that have not yet occurred) and backward-looking responsibilities (that is, senses of responsibility that refer to things that have happened in the past). It should be noted that some senses of responsibility contain both backward-looking and forward-looking elements. For example, responsibility-as-liability is typically attributed on the basis of past (causal) contributions that agents have made, but it also implies a duty to rectify, or at least compensate for, damage or loss.

A further distinction could be made between (more or less) descriptive notions of responsibility and normative notions of responsibility that imply an evaluation (e.g. “he is an irresponsible person”) or a prescription (e.g. “you should pay because you are responsible for what happened”). The prime focus of most contributions in this special issue is on such normative notions of responsibility.

The contributions to this special issue touch upon a number of themes and issues in relation to responsibility in technology and engineering. For the purpose of this introduction, we have consolidated them into three categories. The first category is what is also referred to as the problem of many hands, that is, the problem of attributing responsibility in complex collective settings. This problem is, as we will see, due to the collective nature of action in technology and engineering but also due to the other two characteristics of technological action discussed above, that is, long causal chains and uncertainty. The second theme, we distinguish, is responsibility for risks. Although there has recently been a special issue of this journal (Volume 16. No. 3) devoted to this theme, it is still an issue that hasn’t been discussed a lot (cf. Van de Poel and Nihlen-Fahlquist forthcoming). The third theme is the way technologies shape the contexts of human actions, and the consequences of this “mediating role” for especially the responsibility of designers of new technology.

The Problem of Many Hands

The notion of “the problem of many hands” was originally introduced by Dennis Thompson to refer to the difficulty of identifying the person responsible for some outcome if a large number of people are involved in an activity (Thompson 1980).

---

2 This distinction is a bit similar to, but does not coincide with, the distinction that Davis makes between future and historical uses of responsibility. Davis seems to reject the distinction between forward-looking and backward-looking responsibility as he seems to believe that most, or even all, senses of responsibility contain forward-looking and backward-looking elements.
Thompson originally formulated the problem in the context of the moral responsibility of public officials. Because many different officials, at various levels and in various ways, contribute to policies and the decisions of the organization it is difficult to ascribe moral responsibility for the organization’s conduct in final instance. For outsiders of an organization who want to hold someone responsible for a certain conduct, it is particularly difficult, or even impossible, to find any person who can be said to have independently formed and carried out a certain policy or taken some decision.

The various contributions show that there are different ways to conceive of the problem of many hands. Davis is skeptical about whether there really is a problem. He distinguishes between what he calls the problem of many hands and the problem of many causes. The first refers to what we above have alluded to as the collectivity of engineering action, the second to the complexity of causal chains. Following up on Thompson’s formulation, Davis suggests that the problem of many hands is the problem to identify the person responsible for harm. This is mainly a problem for outsiders, Davis argues, because insiders know very well who made what contribution. Conversely, solving the problem of many causes, on Davis’ reading, requires that we know what counts as a cause. It is typical of engineers, Davis argues, that they make it their responsibility to solve certain engineering problems. Solving these problems is something engineers do, or should do, simply because they are engineers, not because they caused the problem. Engineers have, in his view, typically constructed their profession to foreclose the metaphysical argument of many causes. Looking at engineering from an insider perspective, Davis argues, there is no technological bar to the sort of decisions engineers have taken responsibility for. In other words, the problem of many hands is less a problem for engineers themselves than it is for outsiders.

Coeckelbergh (forthcoming), though not explicitly mentioning to the problem of many hands, addresses the problem as well and he takes a different approach. For him, the main problem derives from a tension between individual and distributed action. He questions the adequacy of our traditional moral theories in situations of engineering disasters. On the basis of the explosion of the Deepwater Horizon platform and the resulting casualties and ecological damage, Coeckelbergh argues that the conditions for attributing moral responsibility laid down in traditional theories make demands on agency, control, and knowledge that are seldom met in engineering and technological action. On the basis of a Kierkegaardian notion of tragedy, he develops an account of moral responsibility that tries to avoid the dichotomy between freedom and fate, between activity and passivity. Accounting for experiences of moral responsibility in technological culture and engineering contexts requires that we recognize the tragic as a feature of the human condition. This “tragic” account of moral responsibility has implications for how we assess technological action. In line with Kierkegaard, it seems pointless to expect full, absolute control, Coeckelbergh argues. Since there will always be events and things
that we cannot control, just as there will be struggle and suffering as a result of that lack of control, the criterion of absolute control seems inadequate. It should be replaced with a criterion that allows for a more gradual assessment. The same holds for another condition for responsibility: knowledge. According to Coeckelbergh, this is not a matter of all or nothing, but of degree. Therefore, where technological action is concerned the question should not be whether or not a person is responsible, but to what extent a person is responsible.

Although Coeckelbergh’s discussion is primarily about backward-looking responsibilities, he also indicates how his reflections may be used to create more responsible technological action in the future (forward-looking responsibility). He argues for more epistemic transparency and the recognition that engineering is not an individual but a collective enterprise that stands in need of collective, pooled intelligence in order to contribute to responsible technological action. But here again, recognizing the tragic character of engineering also requires acknowledging that accidents might happen and that full control in technological action is untenable.

More conceptual clarifying work is done in the contribution by Van de Poel et al. (forthcoming). They take seriously the question raised by Davis; that is, the question whether there is indeed a problem of many hands. In their paper, the authors investigate how the problem of many hands can best be understood and why, and when, it exactly constitutes a problem. They take the example of climate change to illustrate how the ascription of backward-looking responsibility for collective harm may in some situations be difficult or even impossible. But is this also a moral problem?, the authors ask. Building on (Bovens 1998), they discuss three dimensions of the problem of many hands: a practical one (that is, the epistemological problem that Davis referred to), a moral or normative one, and a control dimension. Van de Poel et al. argue that the problem of many hands is best conceived as a situation where there is a gap in the distribution of responsibility in a collective setting which is problematic from a moral point of view. They content that whether a gap is morally problematic depends on what one considers the function of ascribing responsibility. This function, in turn, largely depends on what (ethical) theory with respect to responsibility one adopts. Deontologists, for example, emphasize retribution while consequentialists focus on efficacy. But difference senses of responsibility can also be associated with different functions. Responsibility-as-blameworthiness, for example, is typically attributed for reasons of retribution, whereas responsibility-as-liability often seems to be motivated by the desire to do justice to (potential) victims. So rather than one problem of many hands, there may be different varieties of the problem, each relating to a different sense of responsibility and a different function of attributing responsibility.

The different functions of ascribing responsibility in technology development and engineering, to which Van de Poel et al. refer, are central in Doorn’s contribution to this special issue. She distinguishes three perspectives for ascribing responsibility: a merit-based, a rights-based, and a consequentialist one. She discusses these three perspectives in the light of a recent trend in moral philosophy, and more specifically in the field of engineering ethics, to shift the focus of ethics from an abstract outsider’s perspective towards the practice in which moral
deliberation takes place. Doorn shows how this shift from an outsider’s perspective towards an insider’s perspective might have implications for our practice of ascribing responsibility. Based on recent discussions in engineering ethics literature, she develops two “appropriateness criteria” for responsibility ascriptions in engineering and technology development: (1) the approach should reflect people’s basic intuitions of when it is justified to ascribe responsibility to someone; (2) the approach should inform the direction of technology development and therewith improve engineering design. This second criterion requires that the approach can be applied to specific contextualized moral issues that are raised by specific technological and scientific developments rather than to more general abstract issues. Based on these “appropriateness criteria,” she discusses the pros and cons of the three perspectives.

Using the example of the development of a new sewage water treatment plant, Doorn shows how the different approaches for ascribing responsibilities have different implications for engineering practice in general, and R&D or technological design in particular. The paper concludes with the observation that there may be a tension between the demands that follow from the different approaches, most notably between fairness towards potential wrongdoers and the efficacy of the responsibility ascription. Since it is impossible to reduce the different demands into one overarching criterion, there may be a problematic gap in the distribution of responsibilities, which may, again, constitute an occurrence of the problem of many hands.

Responsibility and Risk

With the discussion of the sewage water treatment plant, Doorn links the topic of responsibility to that of technological risk, which is the second recurring theme in this issue. In her contribution, Kermisch (forthcoming) analyses the relationship between these concepts of risk and responsibility. She notes that this relationship is not as simple as one might think. To state that someone should be held responsible for a risk when some damage occurs is to miss the political and controversial complexity, Kermisch argues. She contrasts the technical definition of risk (that is, the statistical expectation value of unwanted events) with a social scientific conception of risk. The former view is dominant in the traditional engineering paradigm, where a distinction is made between the quantification of risk (to be done by technical experts and engineers) and the responsibility for managing the risks. In this paradigm, the relation between risk and responsibility is therefore only an indirect one, Kermisch explains.

In this traditional paradigm, risks are supposedly realistic, out there in the external world. On the basis of the work of the anthropologist Mary Douglas, Kermisch shows how risks may also be viewed as being constructed through the responsibilities they engage. Since each culture has a propensity to select those dangers that contribute to the stabilization of the corresponding social organization, thereby “translating” these into risks, while neglecting other dangers, risks are the result of an interaction between social processes and the external world. According
to this view, the public allocation of blame and responsibility is functionalist in the same way as the conceptualization of risk is: it contributes to the stabilization of social organizations. The concepts of risk and responsibility are therefore inseparable insofar as the notion of risk is constructed through the cultural processes of responsibility and blame allocation, Kermisch explains.

As Kermisch argues, the public increasingly demands experts and decision-makers to behave responsibly in a prospective way as well, which includes an attitude of care (virtue-responsibility). In order to integrate this moral dimension in risk management, she discusses two additional frameworks in which the relation between risk and responsibility is fully intertwined: the polithetic conception of risk defended by Steve Rayner and the anatomy of risk developed by Jonathan Wolff.

A similar plea for a more virtuous approach to responsibility is made by Roeser (forthcoming). Her central message is that engineers should include emotional reflection into their work in order to develop morally responsible technologies. Roeser starts her contribution by questioning the “neutrality view” on technology; that is, the view that technology is value neutral and engineering a predominantly mathematical, quantitative discipline.

Similar to Kermisch, Roeser argues for the inclusion of moral considerations in the decision making on risks, and for Roeser, emotions are an indispensable source of insight into these moral considerations. She supports her claim with empirical psychological research, which shows that emotions are a major determinant in the risk perceptions of lay people. Roeser’s approach is original in that she argues for a revaluation of emotions. These are not biases that threaten objectivity and rationality in thinking about acceptable risks, as it is often argued, but they may contribute to a correct understanding of the moral acceptability of a hazard, according to Roeser. Emotions such as sympathy and empathy may provide insight in fair distributions of risks and benefits, whereas fear and worry can indicate that a technology is a threat to human wellbeing. In order to adequately address these issues, engineers should appeal to their emotions and include them in their work, Roeser argues.

Roeser shows how emotions can also play a role in forward-looking responsibility. Emotions such as sympathy, empathy and compassion can make people aware how their actions can contribute to improving the situations of others. Moreover, these emotions may also help to motivate to act accordingly. Although backward-looking responsibility and its concomitant emotions are important in that they let people critically reflect on what they have done in the past and how they could have done things better, this should ultimately lead to enhanced emotional sensitivity concerning forward-looking responsibility. The latter is especially important in the context of technology development and design, because these concern things that are yet to come. Roeser argues for a conception of responsibility based on virtue ethics, which transcends formal responsibility of job descriptions, such as to make the engineers become more involved. Moral emotions contribute to such an involved attitude, which will ultimately lead to morally better designs, Roeser argues.
Acting with Technology

The third recurring theme in this issue is about the mediating role of technology and the question how artifacts and institutions can be designed so as to promote rather than destroy responsibility. For instance, must expert systems necessarily take some responsibility away from their users by creating epistemic niches for which nobody can legitimately accept responsibility, or can responsibility be taken into account at an earlier stage and be somehow designed into those systems?

The question how technology plays a role in our ability to discharge our responsibilities is discussed in the paper by Polder-Verkiel (forthcoming). She links the discussion of technological mediation to that of collective responsibility, more specifically, to the backward-looking conception of collective responsibility: under what circumstances can people in a collective be held responsible? The central question in Polder-Verkiel’s contribution concerns the role of internet mediation. On the basis of a comparison between the famous Kitty Genovese case, where a person was assaulted and killed while 39 neighbors supposedly watched without calling until it was too late, and an online case in which a young man committed suicide in front of his webcam, Polder-Verkiel explores the salient differences between the two situations. She argues that neither the number of onlookers, nor their physical nearness, nor their anonymity has moral relevance when it comes to holding them responsible. These factors do not directly affect any of the conditions under which we can hold people responsible, Polder-Verkiel argues. However, two other factors do have an effect: the perceived reality of the situation and the ability to act. Although these factors are not unique to internet mediation, the way in which the different factors are intrinsically connected to internet mediation does seem to have a diminishing effect on responsibility in online situations.

The crucial conditions that Polder-Verkiel identifies in her paper are related to the so-called epistemic condition (people should know that a particular action or inaction leads to bad outcomes) and control or freedom-of-action condition (people should be able to take action) of responsibility. In the philosophical discussion of responsibility, these two considerations are often mentioned as conditions that have to be met before we can hold a person responsible.

Given that internet (or ICT systems, in general) can create situations in which people are not able to discharge their responsibilities, how could we design technical artifacts such that people’s ability for responsibility is enhanced rather than constrained? This question is explicitly addressed in the contribution by Van den Hoven et al. (forthcoming). Van den Hoven et al. start with the observation that technology and engineering are usually treated as the source of ethical problems, and ethics is treated as a constraint on engineering and technological development. However, ethics can also be a source of technological development, Van den Hoven et al. argue. Their particular argument is that technical innovation may be an appropriate way to deal with situations of “moral overload” or “moral dilemmas.” Such situations are characterized by a plurality of moral demands which cannot be met simultaneously. The philosophical literature has discussed a variety of approaches to moral dilemmas but has tended to ignore approaches that aim at
enlarging the possible set of actions, so that the initially conflicting moral demands may all be met. One such an approach is technical innovation.

Van den Hoven et al. argue that if technical innovation implies an improvement on all value dimensions, and thus that it diminishes moral overload, it entails moral progress. Obviously, however, not all technical innovations do so. They therefore suggest that engineers have a responsibility to develop those innovations that reduce moral overload and so enable users, and other actors, to act responsibly.

Van den Hoven et al. propose the methodology of Value Sensitive Design to systematically incorporate moral values into the design of technical artifacts and systems. The central idea in this methodology is to look at design from an ethical perspective, rather than making moral values a by-product.

The last contribution in this special issue is also related to technological mediation. Swierstra and Waelbers (forthcoming) develop a framework (“matrix”) that helps explore in advance how emerging technologies might plausibly affect the reasons behind people’s (moral) actions. Similar to Polder-Verkiel and Van den Hoven et al., Swierstra and Waelbers emphasize that technologies can influence people’s actions. It is therefore important to look beyond the technological, economical, and environmental aspects of technology, and to focus on its future social roles. These social roles are classified on the basis of a two-dimensional matrix. On the horizontal axis, the authors distinguish three basic types of reasons that play a role in practical judgment: what is the case, what can be done and what should be done. On the vertical axis, three classes of morally relevant issues are distinguished: stakeholders, consequences and the good life. For each cell in the matrix, technologies can simultaneously have a positive or negative contribution. For this issue’s topic, responsibility, the cell at the intersection of the “stakeholder row” and “ought column,” is the most relevant. Guiding questions identified with this cell are: Will the technology mediate our duties and responsibilities? Will the technology expand or lessen our duties and responsibilities? These and other questions could help professionals in understanding the potential social role of the technologies they develop. On the basis of the Google PowerMeter, an online monitoring tool that provides energy consumption information by utility smart meters and energy monitoring devices, the authors illustrate how this matrix can support the moral imagination of technology developers and engineers.

The authors argue that it is impossible to be fully certain about a technology’s future social role. This brings us back to contribution by Coeckelbergh, who argued—in similar vein—that it is our human condition to live with uncertainty. As it is impossible to strive for full control in engineering and technological action, so should we not expect to know exactly what social role a technology will have in the future, the authors argue. However, this should not exempt professionals from their responsibility to try developing new technologies in such a manner that the social role will be desirable.

Taking Stock

What could we conclude on the basis of these contributions, what are the implications for the practice of engineering? A first observation relates to the
notions of responsibility that are discussed in the contributions. As indicated earlier, one can distinguish between backward-looking notions of responsibility, which are often related to accountability and blame, and forward-looking responsibility, which can be understood as a virtue or a moral obligation to see to it that a certain state-of-affairs applies. Although some of the contributions (also) pay attention to backward-looking responsibility, in particular Polder-Verkiel, Coeckelberg and Van de Poel et al., most emphasize the forward-looking responsibility of engineers.

Interestingly, this trend is a reversal of the common pattern in philosophy, where the emphasis is usually on responsibility-as-blameworthiness. This suggests that in engineering practice, forward-looking responsibility is more important than backward-looking responsibility. As Davis suggests, engineers are usually less interested in blame for things that have gone wrong but rather tend to take responsibility for somehow improving technologies and ultimately the world.

Related to this is a suggestion that resonates in several contributions that engineering ethics should pay attention to the characteristics of engineering practice itself and adopt an insider perspective. Although their approach and their conclusion are distinctly different, both Davis and Coeckelbergh seem to argue for a more empirically informed perspective on responsibility. For Davis this means: taking an insider perspective and see what responsibility engineers take upon themselves, rather than focusing on holding responsible. Coeckelbergh’s main concern is the acceptance of the tragic element in the human condition. Also Doorn’s contribution can be read as a plea for an internalistic approach.

Another lesson to be learnt from this issue is the point that the neutrality view on technology is untenable. Kermisch and Roeser discuss this point in the light of risk management and risk acceptance. They show that an adequate addressing of risks requires more than merely quantitative considerations. Interestingly, both authors appeal to insights from virtue ethics to support their recommendations.

Van den Hoven et al., Swierstra and Waelbers, and Roeser all argue for explicitly including moral values in the development and design of technology. Although technology is not value-neutral, it is not necessary to either categorically embrace or reject technology, the authors agree. Technology should be developed with due attention for its moral implications. This is where the engineers’ and technology developers’ responsibility comes into play. Given the potential of technology, engineers and technology developers have a responsibility to use this potential for the better. However, where Roeser mainly focuses on the responsibility of engineers, Van den Hoven et al. and Swierstra and Waelbers take the responsibility of the end user into account as well. Technology can play a role in enhancing, or limiting, the end user’s capacity for responsibility. Engineers have a meta-task responsibility—to use the words of Van den Hoven et al.—to design technology in such a way that the responsibility of the end user is enhanced instead of limited. Roeser goes furthest in terms of practical recommendations. She argues for a reform of the engineering curricula, which should include courses that enhance the emotional and imaginative capacities of future engineers.
Open Access. This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References


