4 Values in Design and Responsible Innovation

4.1 The "Naked Scanner" as the Holy Grail of Airport Security

In a shockingly honest confession – "Dear America, I saw you naked" – a former security officer of the US Transportation Security Administration admitted that the naked body beneath a passenger's clothes was fully visible when they passed through the airport scanner.[1] This confirmed a fear that many passengers had, namely that the security officers indeed saw their naked bodies and that they looked at them more than was necessary for security purposes. "Many of the images we gawked at were of overweight people, their every fold and dimple on full awful display," this former Transportation Security Administration officer wrote.[2] "Piercings of every kind were visible. Women who'd had mastectomies were easy to discern – their chests showed up on our screens."[3]

In the aftermath of the 9/11 attacks, security measures at airports were considerably tightened up. One question that received increasing attention was how to prevent nonmetallic hazardous items from being smuggled aboard planes. Several attempts had been made to carry nonmetallic explosive and flammable materials onto flights. The conventional security procedures were based on X-ray portals and could therefore detect only metals (e.g., knives and firearms), and the additional touch search – or "pat-down" – by security officers proved to be ineffective; it was further considered time-consuming and invasive for both the officers and the passengers. The whole-body scanner emerged as a solution to both problems. It could help to identify nonmetallic objects, and it relieved the passengers and security officers of the uncomfortable experience of a pat-down.

[1] Harrington, "Dear America, I Saw You Naked." [2] Ibid. [3] Ibid.

There are two types of whole-body scanners, and they differ in both the scanning and the imaging techniques they use (the two are interconnected). As regards the scanning technique, a distinction is made between backscatter scanners and millimeter wave scanners, both of which are capable of producing highly detailed images of a naked body. A backscatter scanner uses a low-intensity X-ray beam to create a three-dimensional holographic silhouette image of the body and any foreign objects on it, based on the backscattered (reflected) X-rays. The image produced is a highly detailed three-dimensional holographic silhouette of the body that depicts any foreign object on it. A millimeter wave scanner uses electromagnetic nonionizing frequencies to construct an image that reveals the exact shapes and curves of a human body, while also showing any objects on it. Because of the level of detail of the human body underneath clothes that whole-body scanners can reveal, they became known in the popular media as the "naked scanners" or, even worse, as the "virtual strip search."[4]

At first, whole-body scanners were introduced only to supplement the existing X-ray scanners, and were used when there were reasons to search a passenger more thoroughly. Many trials were started at airports throughout the world to assess the the security improvements they brought. Soon after the first trials, it became clear that the whole-body scanners were giving rise to serious privacy concerns, which caused some controversy. The controversy was exacerbated when it was decided that they should fully replace the existing X-ray scanners, after a "near-miss" incident in 2009 when a terrorist managed to carry eighty grams of a highly explosive material, sewn into his underwear, onto a Northwestern Airlines flight in the US.[5] His attempt to detonate the material on the plane failed, but the fact that he had managed to take it through the security portal (consisting of an X-ray scanner and a touch search) significantly increased the need for other airport screening methods to identify such explosive materials. Whole-body scanners were acclaimed as the holy grail of airport security.

The full replacement of X-ray portals with whole-body scanners also gave rise to serious controversies – for instance, regarding whether they would

[4] Cavoukian, *Whole Body Imaging in Airport Scanners*.

[5] Harrington, "Dear America, I Saw You Naked."

lead to severe violations of civil rights.[6] The large-scale use of whole-body scanners also added to the safety concerns.

4.1.1 Designing Scanners for Security and Privacy

Public security has become a matter of paramount importance for many governments around the world. Privacy has also been considered important, certainly in the second half of the twentieth century, but in the midst of the post-9/11 antiterrorism crisis, it seemed to suffer a major setback,[7] or at least was implicitly considered ethically less important and thus easily trumped in favor of security improvements. This was evidenced by a range of emerging legislation that increased the possibility of public surveillance, as well as by a number of technologies that were proposed to ensure security, for instance, CCTV cameras in the public sphere and the use of biometrics in surveillance and identification systems.[8]

The whole-body scanner is an example of the technologies that emerged as a solution to the increasing security concerns. During the trials, however, the problem of the visibility of the human body underneath clothing and the associated privacy concerns became clear fairly fast. In some places, these concerns led to the trials and further use of such scanners being abandoned. In India, for instance, the authorities decided to ban whole-body scanners after a trial in New Delhi airport, because "the images the machines produced were too revealing and would offend passengers, as well as embarrass the security officials."[9] Yet this was notthe case everywhere, as the benefits of the scanners seemed to be too great for them to be rejected altogether. In order to increase the privacy of the scanning process, several improvements were proposed – for instance, separating the officer in contact with the passenger being scanned from the officer operating the scanner, who reviews the image in a back room and does not see the actual passenger.[10] Other, more far-reaching improvements involved algorithmic changes to add

[6] See here, for instance, some concerns expressed by the American Civil Liberties Union: www.aclu.org/aclu-backgrounder-body-scanners-and-virtual-strip-searches?redirect=technology-and-liberty/aclu-backgrounder-body-scanners-and-virtual-strip-searches (consulted July 24, 2019).

[7] Cavoukian, *Security Technologies Enabling Privacy (STEPs)*, 1.

[8] Ibid., 4.

[9] Irvine, "Airport Officials Get X-ray Vision."

[10] Cavoukian, *Whole Body Imaging in Airport Scanners*, 3.

"privacy filters" to reduce the identifiability of the image – for instance, by blurring the face and the private parts of the human body – or to eliminate features that could be considered too personal.[11] Another privacy filter reduced the quality of the image produced of the body, following the argument that we do not need to use the imaging technology to its fullest potential, but only to see "objects hidden underneath the clothing of airline passengers."[12] This could be done by, for instance, reducing the quality of the three-dimensional holographic image that a millimeter wave scanner produces, while increasing the contrast between the body and any foreign object on it. Another adjustment made the screen portray only a generic or cartoon-like body, on which objects could still be indicated.[13]

Another important privacy concern associated with whole-body scanners is the question of whether the generated data is stored, and if so, how. As with other information technology applications, it is important to avoid "unnecessary or unlawful collection, use and disclosure of personal data" or creating methods and tools for individuals to access and control the data generated.[14]

Whole-body scanners were introduced primarily from the perspective of security, but they compromised privacy; privacy improvement was then added to *amend* this system. Privacy deserves, however, to be more important than a mere afterthought. Privacy needs to be built into the system, or, as Ann Cavoukian – a former Information and Privacy Commissioner in Canada – says, we should "design for privacy."[15] This approach promotes thinking about privacy from the outset, instead of trying to alleviate privacy concerns after systems have been put into use.

4.1.2 Designing Scanners also for Safety

When it comes to designing whole-body scanners, there have been many contentious discussions about security versus privacy. There is also another important aspect that deserves attention: that is, what are the health and safety issues associated with the different scanning techniques? Should we consider safety as a third concern, in addition to security and privacy? As

[11] Keller et al., "Privacy Algorithm for Airport Passenger Screening Portal"; Cavoukian, *Whole Body Imaging in Airport Scanners*.

[12] Cavoukian, *Security Technologies Enabling Privacy (STEPs)*, 7.

[13] Cavoukian, *Whole Body Imaging in Airport Scanners*, 6.

[14] Ibid., 1.

[15] Cavoukian, *Security Technologies Enabling Privacy (STEPs)*.

already mentioned, the two methods commonly used are millimeter wave and backscatter X-ray scanning. They differ essentially in that the former uses radio waves to detect energy reflected from the body to construct a three-dimensional image, while the backscatter X-ray uses low-intensity X-rays and detects the backscattered radiation.

X-rays are a form of ionizing radiation, meaning that they pose a potential health hazard (e.g., cancer risks) which depends on the type, intensity, and length of exposure. Millimeter radio waves are essentially nonionizing and therefore presumably safer than X-rays. It should be noted, though, that the doses of radiation associated with backscatter scanners are very low – "representing at most an extremely small cancer risk" – and are thus safe for most individuals who travel by air only a few times a year; the risks will, of course, be higher for frequent fliers and flight personnel.[16] At least equally important is the broader public health perspective: that is, given that a very large number of individuals are exposed to very many small doses per year (up to a billion scans a year in the US alone), there are concerns about "the long-term consequences of an extremely large number of people all being exposed to a likely extremely small radiation-induced cancer risk."[17] From an ethics-of-risk perspective (as discussed in Chapter 2), these concerns are particularly relevant when the ethical acceptability of this scanning method is assessed, because there is alternatives that has the same benefits but no such risks (i.e., the millimeter wave scanner). The ethical legitimacy of the risk imposed by backscatter scanners could therefore be questioned.

In this chapter, I will discuss how we can design for socially and ethically important features, such as security, privacy, safety, and more. In Section 4.2, I will discuss why ethics matter in engineering design. Section 4.3 focuses on persuasive technologies, or designing to promote a certain ethically desirable behavior. Section 4.4 moves the discussion to values and how they matter in the design, while Section 4.5 discusses how we can systematically design technology for values, specifically elaborating the two approaches of value-sensitive design (VSD) and Design for Values. An important question in the process of designing for values is how to address value conflicts and value trade-offs; this is discussed in Section 4,6. The concluding section, Section 4.7, discusses a new and influential policy approach, "responsible

[16] Brenner, "Are X-ray Backscatter Scanners Safe for Airport Passenger Screening?," 6.
[17] Ibid.

innovation," that front-loads ethical thinking in innovation in order to create not only technological artifacts but also socio-technical systems more responsibly. The role that values play in discussions of responsible innovation will also be further explored in this section.

4.2 Why Does Ethics Matter in Engineering Design?

The claim that technology is not a "neutral" practice based only on indisputable facts and figures has reverberated throughout this book. Engineering and technological design are essentially normatively or ethically laden.[18] An important example that is often mentioned in the literature concerns the "racist overpasses" described in a seminal paper by Langdon Winner titled "Do Artifacts Have Politics?" in 1980.[19] Winner discussed the extraordinarily low-hanging overpasses over the parkways in Long Island in the US state of New York. There are about 200 of these overpasses, and they each have a clearance height of only nine feet (2.75 m) at the curb. Even when one notices this "structural peculiarity," the reason for it is not immediately clear. In fact, the overpasses were "deliberately designed to achieve a particular social effect."[20] Robert Moses – a famous planner and builder who was responsible for many public works in New York between 1920 and the 1970s – apparently intended the overpasses to prevent public buses (which were twelve feet, or 3.65 m high) from using the routes. This was discovered years later by Moses' biographer, who pointed out that Moses had "social-class bias and racial prejudice."[21] White "upper-class" people, who owned automobiles could use the routes to access facilities such as Jones Beach, one of Moses' widely accredited public parks, while poor "lower-class" people – often racial minorities – who depended on public transit could not To reinforce this social and racial effect, Moses also vetoed "a proposed extension of the Long Island Railroad to Jones Beach"; the automobile therefore remained the only means to this park.[22]

This example illustrates how inequalities can be deliberately designed into construction, and how they can have an impact for many years afterward. While the example might seem a little extreme – reminiscent of the

18 Radder, "Why Technologies Are Inherently Normative."

19 Winner, "Do Artifacts Have Politics?" 20 Ibid., 123. 21 Ibid., 124.

22 See for Moses' biography Caro, *The Power Broker*. The references to this work here are based on Winner's article "Do Artifacts Have Politics?"

discussions of racism in the first half of the twentieth century – the types of questions it raises are still very relevant to engineering design, which always needs to take into account ethical issues such as security, privacy, and safety, as in the case of whole-body scanners. This process will then give rise to questions about how far of each of these issues should be taken into account, and whether the prioritization of one could compromise another. Should security improvements, for instance, be made at the expense of basic liberties such as privacy? Later in this chapter I will argue that the lens of values can help us to address such questions. However, before moving on to values and how we can systematically design technologies for them, I will first focus on an approach that aims to incentivize ethically acceptable behavior in design.

4.3 Designed for the Sake of Ethics: Persuasive Technology

In the previous section, I argued that the design of technical artifacts can have an ethically problematic effect. This, indeed, builds on the argument against the neutrality thesis in engineering design.[23] At a fundamental level, there are two strands of discussion within this thesis: first, about "whether technology itself or its influence on human life may be evaluated as morally good or bad," and second, about whether we may "ascribe some form of moral agency to technology."[24] A fuller discussion of agency will be found in Chapter 5, which concerns the ethics of autonomous technologies.

The discussion in this chapter is focused on how the design and use of technologies can involve ethical dimensions or, alternatively, how technologies can mediate human behavior, perceptions, and decisions.[25] Acknowledging this "moral" characteristic, we can design technologies in such a way that they can incentivize the "right choice." Such incentivization has been called "nudging." Stemming from social psychology, the concept of nudging initially had little to do with technological design; it was introduced by Richard Thaler and Cass Sunstein to mean designing an environment in such a way as to incentivize, or nudge, the user toward a certain behavior, choice, or attitude.[26] An example often used to illustrate the purpose and

[23] Flanagan, Howe, and Nissenbaum, "Embodying Values in Technology."

[24] Kroes and Verbeek, "Introduction," 1.

[25] Verbeek, *Moralizing Technology*, 2.

[26] Thaler and Sunstein, *Nudge*.

usefulness of nudging concerns the displaying of food choices in school cafeterias. Apparently, by merely rearranging the food, the consumption of certain items can be increased by as much as 25 percent.[27] Simply by placing carrot sticks, salad bowls, and fresh fruit at eye level and early in the cafeteria line (and before the deep-fried food), we can create an awareness of healthy food and nudge kids to eat more of it.

The same rationale of nudging also applies to the intentional designing of technology so that it will incentivize or persuade the user to behave in a way that is considered good." Persuasive technologies are intentionally designed to incentivize certain behaviors or attitudes that are considered desirable.[28] They aim to give "the user feedback on his actions (or omissions)" and try to "suggest" to the user "a desired pattern of behavior."[29] Among the good examples of such persuasive technology are gas pedals that "increase their resistance to the foot of the driver when the car is going above a certain speed so as to encourage a more economical use of energy."[30] A similar example that aims to promote more sustainable driving is a new design of dashboard in several hybrid cars that shows the fuel consumption in real time. The Honda Insight, for instance, "has a little display field, on which 'leaves icons' virtually grow, if the user is driving in an environmentally friendly manner"; if the driver accelerates and brakes a lot and thereby consumes fuel less responsibly, "the leaves will disappear again."[31] Thus, the dashboard displayencourages you to do the "right thing" while driving.

It is important to realize that "persuasion" here means that the choice can be made freely and voluntarily and on the basis of correct information; if the incentivized choice were imposed or forced (coercion) or based on wrong or incomplete information (manipulation or deception), the technology would not fall into the category of persuasive technologies.[32] In other words, persuasive technologies do indicate what the right choice is, but they leave the making of that choice to the user. If the intended choice is not made, persuasive technologies will remind the user of their wrong, or at least unadvised, choice, but all choices must remain open at all times.

[27] Ibid. [28] Fogg, "Now Is Your Chance to Decide What They Will Persuade Us to Do."
[29] Spahn, "And Lead Us (Not) into Persuasion . . .?," 634.
[30] Brey, "Ethical Aspects of Behavior-Steering Technology," 357.
[31] Spahn, "And Lead Us (Not) into Persuasion . . .?," 634.
[32] IJsselsteijn et al., "Persuasive Technology for Human Well-Being," 1.

Persuasive technologies are supposed only to encourage, and they do so to varying degrees ranging from subtle hints, such as a blinking light, to a more persuasive signals, such as an audible buzz.

While it is reasonable to assume that human beings are the best persuaders, technologies may have certain advantages: for instance, they can be more persistent, and they do not give up easily when ignored. Indeed, they can be *irritatingly persistent*: think of the seatbelt reminders that give an urgent buzz whose frequency and intensity continue to grow until you put your seatbelt on.[33] Even if the buzz were to stop after a few seconds (or perhaps after a minute), it would still be more effective thanan average human persuader. If the buzz did not stop, we might question the voluntariness of the choice it was incentivizing. That would then likely place the technology beyond the realm of persuasive technologies, as it would be coercing the user to choose the prescribed option,. which might affect the driving condition too negatively, maybe even increasing the risk of an accident. In fact, EU protocols insist on this voluntariness by emphasizing that seatbelt reminders “should not affect the drivability of the vehicle”; EU law requires a visual signal that persuades but does not irritate to the extent that it negatively affects the drivability of the vehicle.[34]

The development of these persuasive seatbelt reminders has a peculiar history. It starts in the 1970s, when some car manufacturers decided to design cars that would not start if the driver’s seatbelt was not fastened, or had a seatbelt that buckled mechanically when the car detected that someone was sitting in the driver’s seat. While this was clearly intended to improve the safety of the driver, not everyone appreciated being “mechanically forced to wear their seat belts,” and many drivers had their automatic seatbelts removed.[35] This design feature was, therefore, meant not merely to steer the driver’s behavior, but also to impose an action on them, without which a key functionality of the vehicle would not be available. One might ask: Is this an ethical problem? Indeed, the feature is only meant to protect the driver, and everyone is better off if there are fewer accidents and injuries.

33 Ibid., 2.

34 See here the EU seatbelt reminder guidelines and protocols: https://ec.europa.eu/transport/road_safety/specialist/knowledge/esave/esafety_measures_known_safety_effects/seat_belt_reminders_en (consulted August 6, 2019).

35 Brey, “Ethical Aspects of Behavior-Steering Technology,” 358.

The flip side of this imposed safety is that it removes individual liberties. Some people in the US took the case to court, arguing that their civil rights were being violated. They succeeded in having the regulations changed so that "wearing seat belts became again something that was mandatory but no longer mechanically forced."[36] Different seatbelt reminders were then introduced to either gently remind drivers to buckle up (i.e., with a blinking light) or earnestly urge them to do so (with an audible buzz).

The effect of persuasive technologies in removing users' liberty and autonomyhas been criticized as "technological paternalism," which conflicts with "the ideal of a free and autonomous choice of the individual."[37] Persuasive technologies have further been accused of creating a "responsibility vacuum," that is, a situation in which it is not clear who is responsible for an action and its consequences, especially when something goes wrong. In the next chapter, I will discuss both autonomy and responsibility in the broader context of autonomous or semiautonomous technologies.[38] Let us now turn to the question of how values matter in engineering design.

4.4 How Do Values Matter in Engineering Design?

In the previous section, I argued that technologies are not ethically neutral and that they can affect human perception, behavior, and actions; persuasive technologies aim, then, to steer these toward an ethically desirable course of action or attitude. We can, for instance, persuade a driver to wear a seatbelt for the sake of their own safety, or to drive more sustainably by not pushing on the gas pedal too hard or too often. This section extends the critique of the neutrality thesis by building on the reasoning that engineering design is essentially value-laden.[39] That is, technologies can embody and promote not only instrumental values – such as functional efficiency and ease of use – but also substantive moral and political values, such as privacy, trust, autonomy, and justice.[40] When a technology promotes certain values (e.g., safety and sustainability in driving), this enabling may in turn also hamper

[36] Ibid. [37] Spahn, "And Lead Us (Not) into Persuasion . . .?," 634.

[38] Brey, "Ethical Aspects of Behavior-Steering Technology," 363.

[39] Van de Poel, "Values in Engineering Design"; Van den Hoven, Vermaas, and Van de Poel, *Handbook of Ethics and Values in Technological Design*.

[40] Flanagan, Howe, and Nissenbaum, "Embodying Values in Technology," 322.

or reduce other values (e.g., the autonomy of the driver). In this section, we focus on what values mean and how they feature in engineering design.

As a working definition, we assume that values are things that are worth striving for for moral reasons, or "what a person or group of people consider important in life."[41] Values are, therefore, different from individual preferences, wishes, and desires, in that they relate to a common good that we wish to uphold for everyone. To say that something is valuable means not only that it is valuable to me as an individual, but also that "it is or should be of value to others."[42] Not all preferences and choices are explicitly linked with and expressed in conjunction with such values. Our focus here is on preferences and choices that are motivated by and stem from profound beliefs about people's conception of what is considered to be good. Value statements in engineering design indicate whether "certain things or state of affairs are good, valuable, or bad in a certain respect."[43] According to this definition, then, safety is to be considered a value, because we can assume that safety is relevant for moral reasons, and represents not a personal preference but a widely shared moral characteristic. In fact, increasing safety (or mitigating technological risks) has always been a primary concern in engineering; "the value of safety is almost always conceived as a ubiquitous though often implicit functional requirement."[44] In addition to safety, we might, for instance, think of privacy and security, as in the case of whole-body scanners discussed at the beginning of this chapter. As a matter of fact, security – not safety – was the leading design criterion for these scanners.

In engineering design, we can indeed design for many different values, including autonomy, accountability, transparency, justice, well-being, and sustainability.[45] Indeed, when we design for a value, that does not necessarily mean that we design for an uncontroversial and unified definition of that value. In fact, many of the above-mentioned values may have different interpretations, which will not necessarily converge on the same design specificities. Let me explain by discussing the value of sustainability. Prioritizing this value in a design could mean that the design should result

[41] Friedman, Kahn, and Borning, "Value Sensitive Design and Information Systems," 349.

[42] Van de Poel, "Values in Engineering Design," 974. [43] Ibid.

[44] Doorn and Hansson, "Design for the Value of Safety," 492.

[45] See for an overview of different values and different areas of application Van den Hoven, Vermaas, and Van de Poel, *Handbook of Ethics and Values in Technological Design*.

in "less degradation of our environment, less depletion of materials, and more social equity in our world," but it could also mean a "higher level of prosperity for people in developing countries."[46] While these requirements are not contradictory and can, in principle, be accommodated simultaneously, they might impose different demands during the design. An essential part of designing technologies for values is, therefore, the process of value specification or operationalization. During our discussion of this process, we will examine the different meanings and interpretations (i.e., specifications) of individual values and the potential value conflicts (where different values cannot be achieved at the same time). Sometimes, a conflict emerges as a result of different specifications of the same value by different stakeholders; this is highlighted in the discussion of responsible innovation in Section 4.7.

4.5 How to Systematically Design for Values

Value-sensitive design (VSD) is the first systematic approach to proactively understanding, addressing, and including values in the process of design. Scholars of VSD argue that the design process has value implications because new technology can shape our practice and hence promote or undermine certain values.[47] VSD originates from information technology and from the acknowledgment that important ethical values, such as user autonomy or freedom from bias, are being designed into computer systems; if the designer does not include them during the design process, it may prove difficult, or even impossible, to include them after the design has been completed. In the words of an important pioneer of VSD, Batya Friedman, we must create "computer technologies that – from an ethical position – we can and want to live with."[48] VSD for the first time systematically reviews these values at an early stage and in an iterative tripartite methodology consisting of conceptual, empirical, and technical investigations.[49]

Conceptual investigations aim to conceptually (that is, not yet empirically) identify the direct and indirect stakeholders, the values at stake, their potential conflicts, and the inevitable value trade-offs. Conceptual investigations

[46] Wever and Vogtländer, "Design for the Value of Sustainability," 513–14.
[47] Flanagan, Howe, and Nissenbaum, "Embodying Values in Technology."
[48] Friedman, "Value-Sensitive Design," 17.
[49] Friedman, Kahn, and Borning, "Value Sensitive Design."

are often the result of philosophical analysis at an early stage of design. Empirical investigations aim to answer the conceptual questions by investigating them among stakeholders, particularly focusing on how values are perceived and how choices are made when two values are in conflict. "How do stakeholders apprehend individual values in the interactive context? How do they prioritize competing values in design trade-offs? How do they prioritize individual values and usability considerations? Are there differences between espoused practice (what people say) and actual practice (what people do)?"[50] Technical investigations focus on the suitability of a certain technology for accommodating certain values: "a given technology is more suitable for certain activities and more readily supports certain values while rendering other activities and values more difficult to realize."[51] I will return to conceptual discussions of value conflicts and value trade-offs in Section 4.6.

VSD is similar to "design for privacy," discussed earlier in this chapter, but it aims to focus not on only one value, but on the range of different values at stake.[52] To returning to the case study at the beginning of this chapter, had a VSD approach been followed for the introduction of whole-body scanners, it would have required a conceptual investigation to first identify the set of values at stake at the design stage – that is, security, privacy, and safety, and perhaps more – and address the potential conflicts that could occur between them (e.g., between security and privacy). The empirical investigation would have engaged with stakeholders in order to understand how different stakeholders perceive each value – for instance, the value of privacy from the perspective of a passenger – while also addressing situations in which values conflicted. The technical investigation would then have revealed that a millimeter wave scanner is perhaps the one that can better help us address, or even bypass, the conflict between security and privacy; so this would be the technology that could accommodate both values simultaneously. In sum, a VSD approach might have helped to identify these conflicts in an early stage of development, which could in turn

[50] Ibid., 3. [51] Ibid.

[52] To be fair to what Cavoukian called "design for privacy," it was indeed also their aim to consider privacy in addition to security. So, essentially, their approach resembles the rationale of VSD; see Cavoukian, *Whole Body Imaging in Airport Scanners*.

have contributed to reducing the controversies that occurred as a result of privacy infringements.

A second approach to proactively including values in engineering design is the Design for Values approach, which can be seen as elaborating the VSD approach by extending the inclusion of moral values to other domains of technological design.[53] The Design for Values approach rests on the assumption that the "explicit and transparent articulation of values" is highly relevant to design and innovation, and it allows for designing for shared values.[54] The tool, methodologies, and procedures of Design for Values have mostly been developed in relation to the literature on the ethics of technology. The approach further rests on three main claims, as described in the seminal *Handbook of Ethics, Values, and Technological Design*, (1) values can be expressed and embedded in technology, (2) it is ethically important to think explicitly about such values, and (3) for values to have a serious bearing on design and development, they must be included at an early stage.[55]

While values are often general notions at a rather high level of abstraction, engineering design is often based on more concrete guidelines and instructions.[56] Therefore, values must first be translated into design requirements, a process that is also called value specification or operationalization, as mentioned in the previous section. To explicate this, Ibo van de Poel has introduced the concept of a value hierarchy, which aims to show "the relation by which higher level elements are translated into lower level elements in the hierarchy."[57] A value hierarchy can be represented as a triangle divided into three levels. The top level consists of values, and the bottom level is the design requirements that pertain to "certain properties, attributes or capabilities that the designed artefact, system or process should possess."[58] Between these two levels, there is an intermediate level of norms. Norms may include capabilities (e.g., the ability to preserve one's privacy), activities (e.g., moving quickly through the security procedure), or objectives

[53] Van de Poel, "Values in Engineering Design."

[54] Van den Hoven, Vermaas, and Van de Poel, "Design for Values in Nuclear Technology," 3.

[55] Van den Hoven, Vermaas, and Van de Poel, *Handbook of Ethics and Values in Technological Design*.

[56] This and next few paragraphs partly draw on the author's contribution to Elsinga et al., "Toward Sustainable and Inclusive Housing."

[57] Van de Poel, "Translating Values into Design Requirements," 253.

[58] Ibid., 254.

(e.g., minimizing the health risks posed by whole-body scanners). In this definition, norms do not, however, include specific targets to be achieved in design (e.g., a maximum level of radiation that passengers can be exposed to), which are called design requirements. It should be noted that this definition of norms differs from the definition of the term in policy (e.g., legal norms), and also from how the term is often used in engineering, for instance, in International Organization for Standardization norms. In the present terminology, such a specific requirement of design would rather be subsumed under the heading of design requirements, which themselves are informed by a specific norm. Therefore, this proactive step of translating values into norms serves the purpose of making abstract values more tangible and more applicable to designers' practical design requirements.

Van de Poel presents an example from chicken husbandry, specifically the design of battery cages, which he describes as "the most common system in industrialized countries for the housing of laying hens" because they make possible the production of eggs in an economically efficient way.[59] These cages, however, have been criticized for neglecting animal welfare because they do not provide a good living environment. The leading value here in the redesigning of the battery cages is thus animal welfare, which in policy-making has been translated into measurable requirements such as "egg production per animal, … egg weight and the mortality of chickens."[60] In addition to animal welfare, there are also other values at stake, such as "environmental sustainability," which relates to, for example, emissions (especially of methane and nitrogen dioxide) from the cages. Although translating the value of animal welfare was unfamiliar to engineers and designers, attempts were made to translate it into norms and requirements for the design or redesign of battery cages. For example, an interpretation of animal welfare at the level of norms could be expressed in terms of "enough litter," which in turn could be translated into specific design requirements such as "litter should occupy at least one third of the ground surface."[61]

Before moving on to a discussion of how to deal with value conflicts in design, let me first make two remarks that may help to put a value hierarchy and the relation between its three levels better into perspective. First, following Ibo van de Poel, I believe that values, norms, and design requirements have, in principle, a nondeductive relationship. That is, norms cannot

[59] Ibid. [60] Ibid. [61] Ibid., 258.

be logically deduced from values, and design requirements cannot be logically deduced from norms. Instead, the process of specification or operationalization of values is highly contextual and depends very much on the individual case. The relation between the levels has been described as one of "for the sake of"; a norm is presented for the sake of the higher-level value, and design requirements are for the sake of the norms.

Second, it is very likely that within each engineering context and for each specific case, more than one norm can be found for the sake of each value and different design requirements can be presented for the sake of each norm. The norm of "enough litter" in the example above can be interpreted in terms of the percentage of the ground surface that must be covered with litter, as explained above, but it can also be interpreted in terms of the absolute surface area that must be covered by litter, for instance, "at least 250 cm^2 of litter per hen."[62] The same goes for the translation of the value into the intermediate level of norms: "animal welfare" can be translated into "enough litter," but also in terms of "enough living space" or "the presence of laying nests."

4.6 How to Deal with Conflicting Values

A central aim of Design for Values is to identify conflicting values and then address them as far as possible. This will lead to the creation of socially and ethically more acceptable designs. However, addressing conflicting values raises both a fundamental and a practical question. At a fundamental level, if values are ethically important features – things we should hold paramount – can we foresee situations in which they could be trumped? Can we, for instance, compromise the value of human equality a little bit? Or to put it more bluntly: Is there a degree of racism that we might find acceptable if it would improve public safety and security? At a practical level, how can we address the issue of conflicting values? Identifying at the design stage that two values conflict does not always mean that one of the values must be left out completely (though that might sometimes be the case). It is worth contemplating how to deal with value conflicts. I will distinguish here between the strategies of "designing out the conflict" and "balancing the values." Let me first focus on the fundamental question of whether values must be considered to be absolute moral entities.

[62] Ibid.

4.6.1 Is a Little Bit of Racism Acceptable?

Understandably, a first response to this question is an appalled "Absolutely not!" Of course, I am not going to argue against this. Instead, I want to give an example of a situation in which such racism is the unfortunate result of certain policies concerning the use of facial recognition software. In the UK, for instance, the police have trialed the use of real-time facial recognition when people pass CCTVcameras. The police call facial recognition software an "invaluable tool" in the fight against crime, but it is also known that the software has difficulties in coping with "black and ethnic minority faces" since "the software is often trained on predominantly white faces."[63] An engineer might rightly argue that rectifying this is a matter of "simply" changing the technology, or refining the software so that it is sensitive to different types of faces, but such a change would require additional development time. The question then is whether trials should be stopped until the software is adjusted, or whether the false positives – leading to members of ethnic minorities being falsely accused of crimes – should be tolerated because of the "greater good" that will be served by using such software. As the BBC reported in 2019, the UK police have known about this issue since 2014, but decided to put the software into use before fixing the problem. Somehow, the problem of insensitivity to ethnic minorities failed to receive enough attention before the facial recognition program was implemented.[64]

In a similar example in the US, and in an attempt to show the racial biases of facial recognition software, the American Civil Liberties Union (ACLU) – an influential civil rights NGO – tested Amazon's facial recognition tool "Rekognition" on images of the members of the US Congress in 2018. Astonishingly, the results showed twenty-eight matches, falsely identifying the congressmen and congresswomen as people who had been arrested for crimes in the past. Among the falsely matched faces there was a disproportionately large number of members of color, including six members of the Black Caucus.[65] The ACLU conducted this research to show the undeniable and painful deficiencies of the facial recognition software and in order to lobby Congress to impose "a moratorium on law enforcement use of face

[63] White, "Police 'Miss' Chances to Improve Face Tech." [64] Ibid.

[65] Snow, "Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots."

surveillance."[66] Amazon acknowledged that the software needed to be better trained and adjusted, but it also defended it, emphasizing the great success of the tool in "preventing human trafficking, inhibiting child exploitation, and reuniting missing children with their families," among other things.[67]

To returning to the ethical question presented at the beginning of this section – namely whether compromising human equality is acceptable – we can easily agree that it is not. Yet, as sad as it sounds, the realistic observation is that we have technologies in use that have, for instance, racial biases. The fact that these technologies can bring tremendous "greater good" – by helping us to stop human traffickers and child abusers – seems to justify, or at least allow us temporarily to tolerate, their racial biases. One might argue that one of the aims of Design for Values and proactive thinking about values is that we should not be forced to make pragmatic choices when fundamental ethical questions are at stake.

4.6.2 Designing Out the Conflict

When this kind of conflict is spotted at an early stage, the designer may be better equipped to resolve it by choosing the technology that can best accommodate the conflicting values. This may, for instance, mean following the tripartite VSD approach, whose third step, the technical investigation, aims to identify which technology can best accommodate different values. In the case of facial recognition software, once the conflict was spotted, it should have been straightforward to design also for human equality: that is, the software should have been developed using images of people of different races, ethnicities, and skin colors, so that security could still be improved but without compromising human equality.

In the case of the whole-body scanner, such proactive thinking about values could have resulted in X-ray scanners being dismissed at the start because of their safety issues. It would also have helped us to develop imaging techniques that revealed few details of the human body but produced a high level of contrast between the human body and any objects on it. In this way, the values of security, privacy, and safety could all have been

[66] Alba, "Amazon Rekognition Falsely Matched 28 Members of Congress with Arrest Mugshots."

[67] Ibid.

accommodated in an appropriate way; that is, we could have optimized all three values. This is very much in line with how we optimize for engineering design criteria. Indeed, it is only a simplified version of an actual design question, in which a lot more will be at stake, including the price differences between scanning technologies, how fast they can be made available, how accessible each technology is for each country involved, what sort of adjustments in the software (e.g., privacy filters) can be designed and included, and so on. It is likely that a number of other values will also play an important role in these decisions. The argument here is that early engagement with these questions of value conflict could have allowed them to be designed out, resulting in an ethically more acceptable technology.

Another example of the designing out of a conflict concerns a storm surge barrier in the Netherlands, a country where 40 percent of the land area is below sea level and very prone to flooding.[68] In response to a major flood in 1953 that killed over 1,800 people, the Dutch government drew up the Delta Plan to prevent further dangerous flooding. Part of this plan entailed closing off a vulnerable area, the Eastern Scheldt estuary. However, in the late 1960s and early 1970s, environmental concerns became more prominent, and simply blocking off the estuary was deemed unacceptable because of the likely ecological consequences, such as the risks from desalination and a lack of tides, and local fishermen were concerned that there would be huge consequences for their employment. To put it in terms of the conflicting values, there was the value of safety on the one hand, and ecological and well-being (employment) issues on the other. However, an ingenious piece of technology enabled the conflict to be designed out by setting the flood defense system so that the estuary is open by default but can be closed off with movable barriers at times of increased risk. Thus, none of the values was compromised; technology helped to bypass the value conflict.

A final example is from the city of Delft, where I am at the time of writing this chapter. If you had been driving around Delft before 2015, you might have encountered the surprising phenomenon that one of the major highways around the city stopped in an almost apocalyptic way, with the beautiful natural landscape – the untouched area of Midden Delfland – in front of you. This is the A4 highway, which was intended

[68] This description of the case is based on Van de Poel, "Changing Technologies."

to connect the capital, Amsterdam, with one of the world's largest harbors, at Rotterdam, but a seven-kilometer stretch of the highway, which would have cut through the landscape, was simply missing. The traffic had to go through Delft in order to reach the other main highway to Rotterdam, the A13, which was understandably always congested because it had to carry the load of two highways.[69] The reason why the highway did not continue through Midden Delfland, as it was originally meant to do, was an unresolved value conflict between the improved accessibility of Rotterdam harbor (with shorter travel times and better access from the capital) and its associated economic values on the one hand, and on the other hand ecological values associated with the unspoiled Midden Delfland in one of the most densely populated areas in the Netherlands, a country already known for its high population density. This controversy lasted for half a century, but at last an engineering solution was found that, although expensive, managed to accommodate both values sufficiently. A major part of this seven-kilometer missing link has now been built – not on but slightly under the surface, in a sort of "semi-tunnel" that is open on top; wildlife can move freely between the two sides of the highway along the eco-ducts and aqueducts that connect them. When hiking in the countryside around the highway, you can barely hear or see the highway because it has been partially sunk into the ground. This could also be considered an example of designing out a conflict, but it differs from the estuary example in that it was not thought of at the design stage. Instead, it was followed fifty years of controversy that had approached the problem in a binary mode: Should the economy trump ecological values? Furthermore, the eventual solution was substantially more expensive than a conventional highway: The famous Dutch tabloid *Metro* called it the most expensive piece of highway ever built in the country.[70] One might argue that a compromise was made on the economic values in order to enable the highway to be built. This brings me to the next approach to dealing with value conflicts: balancing the values.

[69] The information about the highway is adopted from the website of the Dutch Ministry of Infrastructure and Water Management: www.rijkswaterstaat.nl/wegen/wegenoverzicht/a4/index.aspx (consulted August 12, 2019).

[70] Halkes, "De A4, het duurste stukje snelweg van Nederland."

4.6.3 Balancing the Values

It is likely that designers will often be unable to accommodate all the values that are at stake in a project, and that choices will need to be made. Let me discuss this by returning to the example of the car that would not start until the seatbelt was fastened and the car that physically imposed the seatbelt on the driver. The conflicting values were the safety of the driver on the one hand and the driver's autonomy on the other. The seatbelt that mechanically coerced the driver to wear it seemed to tilt the balance unjustifiably toward the value of safety. As discussed earlier, this was controversial and led to the safety regulations being changed, which pushed manufacturers to use safety systems that were less coercive. The seatbelt signs on the dashboards, as well as the buzz, dealt with this conflict in such a way that drivers' autonomy and freedom were respected but they were reminded of their safety, albeit sometimes in a rather annoying way.

Let us look at autonomy versus safety from a different angle. Could safety never trump human autonomy? Autonomy and freedom of choice are undeniably essential ethical values. John Stuart Mill argued that individual freedoms need to be respected as long as they do not interfere with or hamper another individual's freedom.[71] Freedom and autonomy can then be limited – according to this definition – if they interfere with another person's freedom. In a discussion of Design for Values, one needs to identifythe *scope* of a value, or to whom it relates. The question of safety versus autonomy can be considered within the confines of the car itself, but of course a car participates in traffic. If the driver of a car refuses wear a seatbelt, we could argue that they are compromising their own safety, and we could not immediately justify limiting their autonomy and freedom.

Now consider a drunk driver who is sitting behind the wheel: They jeopardize not only their own safety, but also that of other traffic participants. Would it now be more justifable to limit their freedom and autonomy for the sake of the safety of the other traffic participants? And should we build cars in such a way that drunk drivers will not be able to drive them? This is the basic rationale behind an "alcohol interlock" – an in-car breathalyzer that checks the percentage of alcohol in the driver's breath before the car is started. Excess alcohol use while driving contributes to about

[71] Mill, "On Liberty."

25 percent of all road deaths in Europe, and according to EU road safety documentation, "high risk offenders" – "who offend regularly and/or exceed legal blood alcohol levels by a large amount" – are responsible for a large share of these fatalities. Alcohol interlocks were proposed as a means to deal with this problem of repeat offenders, and they have indeed reduced recidivism by 28–65 percent, leading to substantially fewer fatalities.[72] Limiting people's autonomy seems to be ethically much more justified in this situation. A proposed "soft" version of this process has been to incentivize the possession of personal breathalyzers: When in doubt, use the breathalyzer before you drive. This could be considered a persuasive technology. Yet the objection has been made that the low level of accuracy of commercially bought breathalyzers means that they cannot necessarily guarantee that one is fit to drive. Moreover, the breathalyzers can create a false sense of confidence and sobriety, especially in young, inexperienced drivers, who are more sensitive to smaller quantities of alcohol than more experienced drivers, and this can in turn increase the likelihood of an accident.

In a systematic analysis, Van de Poel presents a number of strategies for dealing with value conflicts, including Cost–Benefit Analysis (CBA), direct trade-offs, satisficing (thresholds), and re-specifications.[73] In a CBA, "all relevant considerations are expressed in one common monetary unit," so that we can decide whether the benefits outweigh the costs.[74] In fact, this builds on the rationale of monetizing costs and benfits as discussed in Chapter 2; and the same objections apply to this strategy too. Let me explain the strategy and its challenges by returning to the alcohol interlocks. In a study carried out in 2005, the benefits and the costs of an alcohol interlock were estimated for several EU countries. The benefit-to-cost ratio varied from country to country, depending on the size of the country, the numbers of drivers and drunk drivers, the statistics of earlier casualties, and the monetary value assigned to each fatality (which was different in each country). In

[72] The information on the alcohol interlock is adopted from the web page of the EU on road safety: https://ec.europa.eu/transport/road_safety/specialist/knowledge/esave/esafety_measures_known_safety_effects/alcolocks_en (consulted August 12, 2019).

[73] It should be noted that one of the strategies Van de Poel discusses is "innovation," which is basically what is described in Section 4.6 as designing out a conflict. See for the full list of strategies and the extensive discussions Van de Poel, "Design for Values in Nuclear Technology."

[74] Ibid., 101.

Spain, for instance, a reduction of 86.5 deaths per year could be achieved, which would amount to a total annual benefit of €69 million (at the rate of €800,000 per death), while the benefits of saving 5.5 lives per year in Norway would amount to almost half of that sum (€32.5 million per year). This is because of the substantially higher human life value in Norway (€5.9 million per death, as opposed to €0.5 million in Spain). So while the benefit-to-cost ratios varied substantially (and were sometimes inaccurate, owing a lack of data in some countries), it was clear that the benefits outweighed the costs. Some car manufacturers, such as Volvo and Toyota in Sweden, offered the installation of alcohol interlocks in trucks as a dealership option. Many transportation companies (for both goods and passengers) added alcohol interlocks to their vehicles to protect the interests of the large number of people who could be affected if the drivers of those vehicles were drunk.

4.7 Responsible Research and Innovation

VSD and Design for Value are two categories among a larger list of approaches and strategies for "early engagement with science and technology" with the aim of front-loading societal and ethical concerns about new technologies.[75] I have discussed these two value-based approaches at length because they are directly relevant to the practice of engineering and technological design. Many "early engagement" approaches have their roots in technology assessment, an "interdisciplinary research field aiming at, generally speaking, providing knowledge for better-informed and well-reflected decisions concerning new technologies [in order to] provide answers to the emergence of unintended and often undesirable side effects."[76] A new approach worth discussing here – because of its strong influence on both academic scholarship and public policy – is responsible innovation. Stemming from the same rationale as technology assessment, responsible innovation is premised on the acknowledgment that science and innovation have produced not only knowledge, understanding, progress, and well-being,

[75] See for a discussion of five approaches Doorn et al., *Early Engagement and New Technologies*.

[76] The quotation is from Grunwald, "Technology Assessment and Design for Values," 68. Technology assessment has more recent advanced approaches in the forms of constructive technology sssessment and political technology assessment; see for discussions on these approaches Doorn et al., *Early Engagement and New Technologies*.

but also "questions, dilemmas and unintended (sometimes undesirable) consequences."[77] It aims to offer a governance and policy approach to understanding and addressing those impacts at the development and innovation stage. Responsible innovation is a notion coined in the EU, first in several national policies in the UK, Norway, and the Netherlands, and later in several of the EU's Framework Programs. In the policy sphere, it is often referred to as responsible research and innovation (RRI), and its objective is to steer the research and innovation process so that the societal and ethical aspects of new innovations are included from the outset. In other words, RRI aims to achieve ethically acceptable, sustainable, and societally desirable technologies.[78] The normative dimension of the approach lies in, for example, what "ethically acceptable" or "desirable" exactly entails. For instance, in EU policy, "ethically acceptable" has been taken by some scholars and practitioners to mean compliance with the EU's Charter of Fundamental Rights (e.g., on matters of privacy, transparency, and safety).[79] Indeed, there are different approaches and definitions of ethical acceptability in the presence of technological risk.[80] In the whole-body scanner example, for instance, the ethical acceptability of different scanning techniques was compared. The fact that millimeter wave scanners have roughly the same performance as X-ray scanners but substantially fewer health risks, calls into question the ethical acceptability of the latter, even if we assume that the risks posed by X-ray scanners are very low (as is indeed the case).

According to an EU expert group, RRI entails involving all stakeholders in the processes of research and innovation at an early stage: "(A) To obtain relevant knowledge on the consequences of the outcomes of their actions and on the range of options open to them and (B) to effectively evaluate both outcomes and options in terms of societal needs and moral values and (C) to use these considerations (under A and B) as functional requirements for design and development of new research, products and services."[81] These are no small tasks, and each of the steps can bring a host of challenges and

[77] Owen et al., "A Framework for Responsible Innovation," 27.

[78] Von Schomberg, "A Vision for Responsible Research and Innovation," 64.

[79] Von Schomberg, *Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields*.

[80] Van de Poel and Royakkers, *Ethics, Technology and Engineering*; Taebi, "Bridging the Gap between Social Acceptance and Ethical Acceptability."

[81] Van den Hoven et al., *Options for Strengthening Responsible Research and Innovation*, 3.

obstacles. Let me mention one that is essential to the whole process. It is difficult to anticipate all consequences of innovations, because some consequences will essentially be visible only when a technology has been implemented, and sometimes only after several years; this is referred to as the Collingridge dilemma.[82] Hence, in addition to being "anticipatory," responsible innovations need to be "reflective," "deliberative," and "responsive."[83] That is, we must try to anticipate the intended and potentially unintended consequences of innovations, while reflecting on their underlying purposes, motivations, and impacts; we must further deliberate with stakeholders throughout the process via dialogue and debate, be responsive to criticism, and be adaptable in the way we innovate.[84]

While responsible innovation shares some of the basic rationales of the value-based approaches in technological design, it also differs from it and is somewhat more comprehensive in scope.[85] First, responsible innovation focuses not only on the innovation itself, but also on the context surrounding of the technology – for instance, the institutions, laws, and legislation that should steer its further development. Responsible innovation thus places the technology against the backdrop of important societal issues. Second, the focus of responsible innovation is not only on technological innovations, but also on innovations in general, for instance in medicine: One might argue that many innovations depend largely on technical developments; think, for instance, of developments in prosthetics and their use in medicine, which depends not only on medicine and orthopedics, but also on various subfields of mechanical engineering. Third, and in conjunction with the previous two issues, responsible innovation is essentially an interdisciplinary effort, for no single discipline can encompass the complexities of the innovations needed to address the societal "grand challenges" of the twenty-first century. A comprehensive overarching view is very much needed, and it should connect engineering with the humanities and social sciences. Interdisciplinarity is a basic rationale that responsible innovation has

82 Collingridge, *The Social Control of Technology*.

83 Owen et al., "A Framework for Responsible Innovation."

84 See ibid. for more details of these four dimensions. See also Stilgoe, Owen, and Macnaghten, "Developing a Framework for Responsible Innovation."

85 Jenkins et al., "Synthesizing Value Sensitive Design, Responsible Research and Innovation, and Energy Justice."

adopted from technology assessment,[86] but responsible innovation has expanded its application substantially; for instance, several EU agencies fund interdisciplinary programs. In addition, EU research findings have also incentivized, and sometimes required, such interdisciplinary collaboration. In the realm of policy-making, more recent discussions of dealing with grand societal challenges have expanded this even further into transdisciplinary research that engages not only different disciplines, but also nonacademic stakeholders.[87] While transdisciplinary research is not a new notion in academia, the renewed interest in it in the area of policy (according to this definition) is certainly noteworthy.

4.7.1 Fracking, Responsible Innovation, and Design for Values

As already mentioned, responsible innovation follows to an extent the basic ideas of value-based approaches.[88] In fact, the EU document quoted above even mentions the role of moral values from the outset, in: the outcome of innovation should be assessed in terms of, for example, the moral values it contributes to. The link between value-based approaches in responsible innovation has been explicitly discussed in the literature.[89] Let me finish this chapter by illustrating how to connect responsible innovation with value-based approaches in engineering, using the example ofa proposal for the exploration of an unconventional gas field – containing shale gas – in the Netherlands. This is based on a case that my colleagues and I investigated in 2013 and 2014, in which we tried to consider responsible innovation as an endorsement of public values. The lens of values enabled us to take a fine-grained look at an interesting yet difficult-to-grasp public debate.[90] Let me explain this in more detail.

[86] Grunwald, "Technology Assessment for Responsible Innovation."

[87] The OECD has, for instance, developed a report that sums up the challenges of transdisciplinary research in order to help incentivize this type of research; see OECD, *Addressing Societal Challenges Using Interdisciplinary Research*.

[88] Taebi et al., "Responsible Innovation and an Endorsement of Public Values."

[89] Van den Hoven, "Value Sensitive Design and Responsible Innovation"; Van den Hoven, "Responsible Innovation"; Taebi et al., "Responsible Innovation and an Endorsement of Public Values."

[90] Correljé et al., "Responsible Innovation in Energy Projects."

The Netherlands has Europe's largest conventional gas field in its northern province of Groningen. Gas is, therefore, deeply embedded in the country's energy infrastructure, both for direct use in households and for electricity generation. Because of the familiarity with and wide acceptance of natural gas, it was assumed that a proposal to extract gas in an unconventional manner would be uncontroversial in this "gas country." What is particularly significant about shale gas is that its extraction method is based on hydrological fracturing – popularly called fracking – in which water mixed with certain chemicals is injected at high pressure into the hard shale layers in order to crack them open, allowing gas to be extracted. While the technology used for fracking is not new (conventional gas fields also use the method), the amount of water used and the chemicals added to it carry particular health and safety and also environmental concerns. Because of these concerns – in conjunction with the seismic risks associated with conventional gas fields, which became more serious in the Groningen gas field around the time when the application for a permit for shale gas exploration was being reviewed by the government – the debate spilled over from one type of gas extraction (for conventional gas) to another type (for unconventional shale gas); thus, the proposal for the exploration of shale gas became unexpectedly controversial.[91] The controversy led the government to impose a moratorium on onshore land extraction of shale gas. The public debate about shale gas exploration was very lively and rich in content, and it provided a fruitful empirical ground in which we could try to understand the development within the framework of responsible innovation.[92]

In our analysis of the public debate, we reviewed the arguments presented by the vociferous proponents and opponents of shale gas, looking for important concerns voiced by different stakeholders in terms of public values. Our first finding was that the public debate often took place at the level of norms – following the value hierarchy that Van de Poel presents for the translation of values into norms, which in turn can be translated into requirements.[93] We found that substantive and procedural values featured in this debate. In the first category, we distinguished, for instance, between

[91] Cuppen et al., "When Controversies Cascade."

[92] The findings below are based on: Dignum et al., "Contested Technologies and Design for Values."

[93] Van de Poel, "Translating Values into Design Requirements."

environmental friendliness, public health and safety, and resource durability and affordability. Procedural values were related to the process of decision-making and included values such as transparency, accountability, and justice.

A second finding of this empirical research was that, contrary to popular belief, proponents and opponents seemed to uphold pretty much the same set of values.[94] The conflicts arose over the operationalization of these values among the different stakeholders – that is, over which norms were presented for the sake of a value. This led us to conclude not only that inter-values conflicts can occur – so that a guarantee of one value may be at the expense of another value – but that there are sometimes intra-value conflicts (i.e., conflicts within one value) when, for instance, different stakeholders operationalize the same value differently. The value of public health and safety, for example, was upheld by almost all stakeholders. While some stakeholders regarded this value as important in relation to the composition and volume of the chemicals added to the water, and emphasized that the surface water coming back from the well needed to be disposed of properly, others found safety to be important in relation to the possible seismic effects that the extraction of gas could bring about. Yet another group of stakeholders related the value of safety to the possibility of replacing coal with gas, which is a less polluting energy resource, both in terms of greenhouse gas emissions (and the associated climate change concerns) and in terms of the direct effects on health engendered by the combustion of both resources for energy production.

To sum up, the lens of values enabled us to zoom in on the details of the controversies. More specifically, we tried to clarify the controversies that emerged as a a result of the inappropriate inclusion of these values. Responsible innovation is to be considered an endorsement of these values. Three remarks are in order. First, our study focused on a retrospective analysis of values in a controversial debate. The aim of responsible innovation is to proactively take account of societal and ethical issues in innovation, or to design for values. This study was meant only to contribute to the further conceptualization of the notion of responsible research and

[94] It should be said that the opponents in general emphasized the procedural values more than the proponents did. See for a detailed discussion of this debate: Dignum et al., "Substantive and Procedural Values in Energy Policy."

innovation, which should help in future proactive approaches to responsible innovation. Second, the shale gas controversy is just an example showing the relevance of discussions of responsible innovation in conjunction with design for values. As has been mentioned, we chose the case because of its empirical richness. Indeed, the basic rationale of this discussion is also broadly applicable to other technological innovations. Third, while RRI was developed in, and has been mostly applied in, Europe, it has been rightly asked whether its scope should remain limited to Europe.[95] There are interesting examples of the role RRI could playin developing countries, such as South America.[96]

4.8 Summary

This chapter has built upon the argument that technological developments are not morally neutral. Especially in the design of new technologies, there are several values that engineers should be aware of and proactively engage with. By discussing the whole-body scanners that were primarily introduced for security at airports in the post-9/11 era, the chapter has identified three main values at stake: security, privacy, and safety. It has discussed two approaches that can help us to proactively include values in design, namely VSD and Design for Values.

When including values at an early stage in design, we can run into situations in which values potentially conflict. The chapter has discussed strategies for dealing with value conflicts, distinguishing between (1) designing out the conflict and (2) balancing the two (or more) conflicting values in a sensible and acceptable way. Indeed, the chapter has not pretended to include all situations of this sort. Complex and ethically intricate situations will emerge in an actual design process, and technologies are – by their very nature – difficult to predict. Not only can this lead to unforeseen consequences, but it can also change the way a value conflict is resolved, tilting the balance in favor of a different value at a particular stage of the design. Furthermore, values are not static and can change: for instance, new values can emerge during the development and implementation of new

[95] Asveld et al., *Responsible Innovation 3*
[96] Vasen, "Responsible Innovation in Developing Countries."

technologies.[97] Values can also change in their moral meaning or relevance. How to design for changing values is yet another important topic that deserves attention.[98]

We have also discussed approaches that focus on innovations in their societal context, and specifically RRI, particularly in public policy. RRI and value-based approaches are very much connected in their goals and ambitions. Conceptualizing RRI in terms of endorsing values can facilitate a fine-grained way of looking at complicated debates and developments in terms of which values are at stake, how those values have been operationalized by different stakeholders, and which values potentially conflict.

[97] Cuppen et al., "Normative Diversity, Conflict and Transition."
[98] Van de Poel, "Design for Value Change."