Rawls's Wide Reflective Equilibrium as a Method for Engaged Interdisciplinary Collaboration

Potentials and Limitations for the Context of Technological Risks

Doorn, Neelke; Taebi, Behnam

DOI
10.1177/0162243917723153

Publication date
2017

Document Version
Final published version

Published in
Science, Technology & Human Values

Citation (APA)
Doorn, N., & Taebi, B. (2017). Rawls's Wide Reflective Equilibrium as a Method for Engaged Interdisciplinary Collaboration: Potentials and Limitations for the Context of Technological Risks. *Science, Technology & Human Values*, 1-31. https://doi.org/10.1177/0162243917723153

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Article

Science, Technology, & Human Values
1-31

DOI: 10.1177/0162243917723153
journals.sagepub.com/home/sth

Rawls's Wide Reflective Equilibrium as a Method for Engaged Interdisciplinary Collaboration: Potentials and Limitations for the Context of Technological Risks

Neelke Doorn[1] and Behnam Taebi[1]

Abstract

The introduction of new technologies in society is sometimes met with public resistance. Supported by public policy calls for "upstream engagement" and "responsible innovation," recent years have seen a notable rise in attempts to attune research and innovation processes to societal needs, so that stakeholders' concerns are taken into account in the design phase of technology. Both within the social sciences and in the ethics of technology, we see many interdisciplinary collaborations being initiated that aim to address tensions between various normative expectations about

[1]Department of Technology, Policy and Management, Delft University of Technology, Delft, the Netherlands

Corresponding Author:
Neelke Doorn, Department of Technology, Policy and Management, Delft University of Technology, PO Box 5015, 2600 GA Delft, the Netherlands.
Email: n.doorn@tudelft.nl

science and engineering and the actual outcomes. However, despite pleas to integrate social science research into the ethics of technology, effective normative models for assessing technologies are still scarce. Rawls's wide reflective equilibrium (WRE) is often mentioned as a promising approach to integrate insights from the social sciences in the normative analysis of concrete cases, but an in-depth discussion of how this would work in practice is still lacking. In this article, we explore to what extent the WRE method can be used in the context of technology development. Using cases in engineering and technology development, we discuss three issues that are currently neglected in the applied ethics literature on WRE. The first issue concerns the operationalization of abstract background theories to moral principles. The second issue concerns the inclusiveness of the method and the demand for openness. The third issue is how to establish whether or not an equilibrium has been reached. These issues should be taken into account when applying the methods to real-world cases involving technological risks. Applying the WRE method in the context of engaged interdisciplinary collaboration requires sensitivity for issues of power and representativeness to properly deal with the dynamics between the technical and normative researchers involved as well as society at large.

Keywords
technological risks, wide reflective equilibrium, social acceptance, moral acceptability, sociotechnical integration, responsible innovation, engaged interdisciplinary collaboration

The ethics of technology is inextricably linked with technological risks.[1] The introduction of new technologies in society is sometimes met with public resistance, especially if these new technologies pose new risks. It is increasingly recognized that decision-making on technological risks cannot ignore stakeholders' opinions on the desirability or undesirability of new technologies (Wustenhagen, Wolsink, and Burer 2007; Huijts, Molin, and Steg 2012). This prompts the question of how to decide on the acceptability of new technologies in the light of vociferous stakeholders' opinions and possible public resistance.

Recent years have seen a notable rise in attempts to attune research and innovation processes to societal needs, so that stakeholders' concerns are taken into account in the design phase of technology (Schuurbiers et al. 2013). Supported by public policy calls for "upstream engagement"

(Wildson and Willis 2004; Sismondo 2008) and "responsible innovation" (Owen, Bessant, and Heintz 2013; Van den Hoven et al. 2014), many interdisciplinary collaborations are emerging that aim to address tensions between various normative expectations about science and engineering and the actual outcomes (e.g., Fisher and Schuurbiers 2013). These initiatives extend the ethnographic approaches in science and technology studies (STS) that have become known as "laboratory studies" (Knorr-Cetina 1981) and explicitly aim to intervene in and engage with the processes that take place in research and innovation (Schuurbiers et al. 2013). Within the normative social sciences, several methods, with different aims and foci, have been developed to give shape to these interventions (Doorn et al. 2013a). One of the most elaborated is the method of midstream modulation (MM), originally developed by Erik Fisher and further tested by an international group of colleagues in a coordinated set of twenty laboratory studies, jointly referred to as the Social-Technical Integration Research (STIR) program (e.g., Fisher, Mahajan, and Mitcham 2006; Fisher and Schuurbiers 2013). The STIR program, which ran between 2009 and 2012, aimed to develop a method to increase the deliberative capacity of individual researchers by placing social scientists or humanities researchers "in the laboratory," thereby hoping to "modulate" the individual researcher "in the right direction," in the sense of including more sustainable procedures and considerations of stakeholder values in the development of technology (Doorn et al. 2013b, 238). At the level of the technological sector, within the broader category of constructive technology assessment, methods have been developed to increase the reflexivity of institutions and sectors in the society (Schot and Rip 1997; Rip and Robinson 2013). Although these methods are normative in their overall goal of contributing to "better technology in a better society" (Rip and Robinson 2013, 40), most researchers in the social sciences are reluctant to bring in their own normativity and avoid explicitly defining what "better" might mean, other than being responsive to societal needs.

In the ethics of technology, we see a similar movement toward interdisciplinary research efforts. Coming from a predominantly normative perspective, where normativity is derived from abstract theories and principles, increasing attention is being paid to contingency and the social constructedness of technology in the ethics of technology. Since the 1980s, and in the wake of STS, ethicists of technology have started developing conceptions of technology that also recognize technology as contingent, socially shaped, and contextually dependent, thereby also creating space to develop views on what morally desirable technology would look like (Brey 2010).

Consequently, in the ethics of technology, attention has shifted from the role of technology in society generally to specific types of technologies, the ethical values embedded in particular designs, and the ethics of engineering practice, often referred to in the literature as the "empirical turn" (Kroes and Meijers 2000; Achterhuis 2001). However, despite pleas to integrate social science research into the ethics of technology, effective models for assessing technologies are still scarce. These normative models should not only be empirically informed but also be able to make a moral assessment based on ethical norms and values (Brey 2010).

In medical ethics, the method of WRE is often presented as a promising approach in decisions on moral issues in specific cases and as such to provide guidance on the morally acceptable course of action in specific situations.[2] Although the WRE method has been developed neither for the medical ethics context nor for applied ethics in general, it may in principle also be useful for moral decision-making about new technologies (Doorn 2013; Cotton 2009; Van de Poel and Zwart 2010). More specifically, the WRE method seems particularly promising for integrating social scientific studies on public or *social acceptance* of technological risks in analyses of the *moral acceptability* of technological risks (Taebi Forthcoming; Van de Poel 2016). However, real applications are still missing and the WRE method itself is not without controversy. Nonetheless, we think that the WRE method may be promising, not only for ethicists of technology who aim to integrate social science research with normative analysis but also for normative social scientists. In contrast to the largely monodisciplinary twentieth century, in which scholars could remain with the boundaries of their own disciplinary methodology, social scientists are increasingly encouraged, through various funding schemes, to engage in interdisciplinary collaborations.[3]

In this article, we explore the main challenges posed by the use of the WRE method in risk-related decision-making. The section below presents the WRE method in more detail, followed by three sections that address three key issues: operationalization, inclusiveness, and the practical establishment of a reflective equilibrium. We use two case studies to illustrate our points. The first case concerns the diverse values at stake in the design of a nuclear reactor. Operationalizing these values reveals the value trade-offs that might motivate different technological choices in the design. The second case is an application of the WRE method as a method for ethical investigations in a real project: the Ambient Living with Embedded Networks (ALwEN) project.[4] The original aim of this case study was to explore how engineers distribute responsibility for

addressing technological risks, which illustrates the responsibility engineers assume for the potential risks that their work places on society.

For the sake of readability, we will not elaborate on these cases in the main text but present them in text boxes. In the final substantive section, we place the WRE method within the wider scope of the normative social sciences, including some brief reflections on the position of a normative scientist in interdisciplinary collaborations.

The WRE Approach

John Rawls originally developed the concept of WRE in the context of political philosophy, but the model is increasingly used in applied ethics as a method of justification (Doorn 2010a).[5] Applied ethics is often about finding the morally justified action in a specific situation. The traditional approaches in applied ethics either start from an abstract theory and apply these to concrete cases ("theory-centered approaches") or they focus on the case itself and derive the morally justified action from the particularities of the specific case ("bottom-up"). Proponents of theory-centered approaches believe that applied ethics should apply general ethical theory to a particular case, which implies that we cannot reach a warranted conclusion about a practical moral issue until we have taken a stand on which theory we have most reason to accept (Peterson 2017). If we look at the current ethics literature, there is by no means an agreement on which particular theory to accept. Alternatively, we can analyze the moral aspects of a particular case without further reference to universal rules or general principles. However, without any reference to ethical theory or principles, it is difficult to justify why a particular action is morally justified. Many alternatives to the two extremes of top-down application of theory and bottom-up particularistic approaches have been developed in recent decades, and the WRE approach is one of the most advanced alternatives.[6]

Key to the WRE approach is that ethical justification does not give priority to either abstract theories or the particularities of a case. Instead, the approach seeks coherence between what people think on a more abstract level (theories, principles) and their judgments about a particular situation. In the WRE approach, a distinction is made between three levels of considerations: (1) descriptive and normative background theories, (2) moral principles, and (3) considered moral judgments about particular cases or situations (see Figure 1). The psychological assumption underlying the approach is that everyone involved wants to arrive at a stable and complete solution, in the sense that these considered judgments are more than a mere

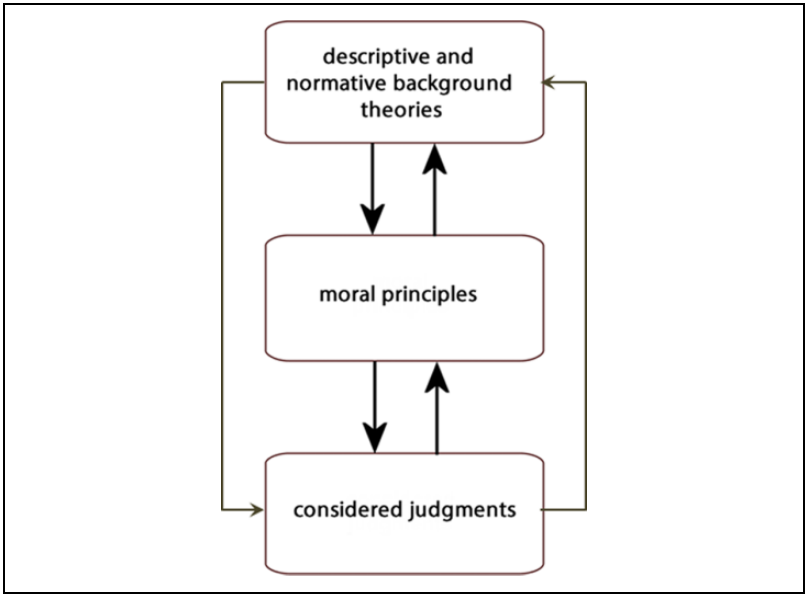

Figure 1. Wide reflective equilibrium model.

collection of accidental convictions. People will therefore aim for coherence between the considerations at the different levels. By moving back and forth between these levels in discussion, and revising those that do not fit well, people arrive at a so-called reflective equilibrium. Equilibrium exists if the different layers cohere and are mutually supportive; it is called reflective if the equilibrium is achieved by working back and forth between the different considerations and if all these considerations are appropriately adjustable in the light of new situations or points of view; and it is called wide if coherence is achieved between all three levels of considerations (Rawls 1999; see Figure 1 for a graphical presentation of the WRE model). The underlying idea of the WRE method is that an outcome, statement, or decision can be considered morally justified if all people involved are able to fit it into their own personal WRE; that is, if it coheres with the other, more abstract layers in the WRE. As each individual brings his or her own set of initial theories, principles, and considered judgments, each individual will probably arrive at a different WRE.

One of the characteristics of the WRE approach is that all of its elements are, in principle, open to revision. It means that if the considered judgments,

moral principles, and background theories do not cohere, any of these elements can be revised to seek a more coherent set of considered judgments, moral principles, and background theories. While this can be seen as an advantage, it also prompts the question of whether any considered judgment should be so powerful as to revise the more abstract layers (i.e., those of the moral principles and background theories). Similarly, the common view in applied ethics is that the original considered moral judgments are empirically informed; the moral principles and normative background theories serve to add the normative component. This immediately raises questions about how the empirical content relates to the normative outcome and whether this way of reasoning would be prone to the naturalistic fallacy, that is, to the fallacy of deriving normative conclusions from merely empirical data. After all, if the coherent set is ultimately meant to provide justification, empirical studies of what people accept cannot provide the final answer to what is morally acceptable. Let us briefly review these objections.

Concerning the first point, we think that this sharp dichotomy between empirically considered judgments, on the one hand, and the moral principles and background theories, on the other, rests on a mistaken view about how these different levels relate to one another. This is not only a top-down application from abstract theories to concrete judgments about particular cases, as the direction of influence also goes the other way. People's considered judgments may very well inform their moral principles and background theories and, consequently, they may also revise their principles in view of new information. For example, people who generally oppose government subsidies, a view that probably finds its origin in the layer of background theories, may lessen their resistance in the light of a lack of progress in the transition to renewable resources. Thus, it seems to go quite against the way people reason to assume that moral principles and background theories are fixed and only considered judgments are open to revision.

It is beyond the scope of this article to discuss the second issue of the naturalistic fallacy in any detail, but we concur with other authors who have argued that the use of empirical data in itself does not make a method prone to the naturalistic fallacy.[7] The main point is that one cannot derive normative conclusions from empirical data alone. However, the whole WRE approach is based on the inclusion of both empirical data *and* normative statements, and, as such, the method can avoid the naturalistic fallacy by including normative statements.

By providing a more detailed discussion of how the WRE approach works in practice, we argue that the method does indeed pose challenges,

but they are different from those commonly discussed in the applied ethics literature (i.e., the objections discussed in the previous paragraphs). In this article, we focus on three specific challenges that are especially relevant if the method is to be used in the context of technological risks: the first relates to the operationalization of abstract theories and principles in relation to a particular technology, the second relates to the inclusiveness of the method, and the third relates to the practical establishment of the WRE. By providing a detailed discussion of the challenges that could be encountered in actual applications, we present a more realistic picture of what the WRE approach method could possibly add to decision-making on technological risks.

Understanding Ethical Issues of Technology: The Conceptual Normative Analysis

One of the challenges in applying the WRE approach is to facilitate a *communication* between the levels, particularly between the two top levels. In more concrete terms, we need to be able to apply moral theory to a specific situation in which technology is designed, implemented, or used. This is not a straightforward task, because if general moral theory is to be applied in a meaningful manner, it should take into account many complex issues, such as uncertainty about risk and moral dilemmas (Beauchamp 2003a, 12). We could conceive of this step as the *operationalization* of the moral theory, so that it becomes applicable to the specific context and situation. The midlevel principles that will follow from this operationalization would then not be generally applicable principles—as presented by Beauchamp and Childress (2001) in their four principles of biomedical ethics—but rather *situational principles* relating to the specific situation of a technological design, development, or application.

Let us illustrate why we cannot operationalize moral theory without a profound understanding of the technology at hand by looking at the example of the design of a nuclear power reactor. In this example, we focus on the material dimension of the technology (cf. Bijker 1995; Latour 2005; Orlikowski 2007) and not on its discursive or institutional aspects, which feature more prominently in the social science literature inspired by the work of Beck (1992) and Giddens (1984) among others.

When operationalizing concepts such as *intergenerational justice or safety*,[8] concepts which are often used to analyze energy technologies, at least three dilemmatic situations could occur with regard to conflicting values, namely, (i) values could conflict (spatially), (ii) values could have a temporal or time-related conflict, and (iii) one value could be interpreted

in two or more different and potentially conflicting ways; this is called an "internal conflict" of a value (Dignum et al. 2016). Let us review these dilemmatic situations by focusing on why an understanding of technology plays a crucial role in revealing these dilemmas. We will consider three future reactor types: the advanced boiling water reactor (ABWR), the high temperature reactor pebble-bed module (HTR-PM), and the gas-cooled fast reactor (GFR) with respect to our presented set of values. Text Box 1 presents these reactors and their key design features.

If we wish to design with safety as the leading value, the HTR-PM seems to be the best option. That is, unlike the ABWR, which is also designed to improve safety, the HTR-PM makes a meltdown risk physically impossible. If we wish to maximally satisfy the value of sustainability (i.e., resource durability), the GFR would be preferred because this type of reactor facilitates the use of thorium, a naturally more abundant material. The GFR would, however, compromise the value of security because plutonium, a dangerously radioactive element with a very long-lived isotope time, is one of the products of this reactor. When recommending one of the reactor types, we must be clear on which value we wish to ensure to the maximum extent and which other values are we willing to (partly) compromise; this is the first type of conflict mentioned above. The second type of conflict would occur if short-term and long-term assurance of a value do not go hand in hand. When we take long-term safety as the leading criterion, we wish to use the reactor that helps us eliminate most of the long-lived isotopes, namely, the GFR, while the GFR would certainly not be the best option from the perspective of short-term safety (and security). An HTR-PM would absolutely be the better option from the perspective of ensuring short-term safety for the reasons mentioned above. A moral analysis of the different nuclear reactors therefore involves a reflection on the relevance of each value in relation to the other values (Taebi and Kloosterman 2015).[9] The third and last type of conflict is an internal conflict. While everybody would easily agree that safety is an important value to uphold when designing nuclear reactors, there are different *conceptions* of the concept of safety, following the distinction that Rawls ([1971] 1999) has made between concepts and conceptions (see also Hart 1961). Safety could, for instance, be defined as reducing the system's dependence on human intervention; the Chernobyl accident resulted partly from operators' failure. The ABWR is a reactor that reduces reliance on operators' intervention by making reactors *passively* safe. But safety could also be interpreted as making a nuclear meltdown physically impossible. Starting the design from scratch, the HTR-PM has certain physical and mechanical features that do not allow

Text Box 1. The Case of Nuclear Reactor Designs.

When designing nuclear reactors, many important issues need to be considered. These include the possibility and probability of core failure or meltdown, the type of fuel needed, the amount of energy produced, the volume and lifetime of waste remaining after operation, and, last but not least, the possibility of using the reactor to manufacture the key ingredient of a nuclear bomb, namely, weapons-grade nuclear material (i.e., highly enriched uranium or plutonium). These issues essentially relate to certain underlying *values*. Indeed, there is the leading value of safety for designing nuclear reactors, but in addition to safety, there are at least four other key values, namely, security (i.e., sabotage), nonproliferation, sustainability (i.e., environmental impacts, energy resource availability), and economic viability (i.e., embarking on new technology and its continuation). The evolution of generations of nuclear reactors (numbered generation I, II, III, III+, and IV, respectively) has been analyzed in terms of these values. Generation I reactors are the "proof of concept" and include prototypes from the 1950s and 1960s. Generation II reactors are commercialized power plants from the 1960s, designed to be economical and reliable; almost all operational energy reactors in the world are of the generation II type. The following generations III, III+, and IV are either incremental design improvements or revolutionary new designs that started from scratch with a specific value as the leading design criterion.

In principle, the best achievable nuclear reactor would satisfy all of the values referred to above. However, the safest reactor is not necessarily the most sustainable one, and the reactor that best guarantees resource durability could easily compromise safety and nonproliferation. Depending on which values are decisive, drastically different reactors could be recommended. Three reactors discussed here are the advanced boiling water reactor (ABWR), the high temperature reactor pebble-bed module (HTR-PM), and the gas-cooled fast reactor (GFR).

The ABWR is an evolutionary successor of the generation II boiling water reactor, with several safety improvements including ten separate internal pumps at the bottom of the reactor vessel, several emergency cooling systems, and the encasing of the reactor vessel in thick fiber-reinforced concrete. A key safety improvement of the ABWR is the development of *passive* models of safety, which do not require *active* involvement of an operator.

(*Continued*)

Text Box 1. (*Continued*)

The HTR-PM is a radical new design with substantial safety improvement, partly due to the shape and type of fuel chosen. The fuel is encompassed in two layers of pyrolytic graphite and one layer of silicon carbide, which make leakage of radioactive nuclides (i.e., fission products) substantially less likely, since those layers can withstand very high temperatures and can thus support the integrity of fuel spheres. These technical features make a core meltdown as the worst failure in a nuclear reactor physically impossible. So, the conception of safety as reducing the probability of meltdown is best warranted in reactors of this type.

The GFR deploys the major isotope of uranium, thus enhancing resource durability in order to meet the value of sustainability. Less than 1 percent of all naturally occurring uranium is deployable in conventional generation II thermal reactors, while fast reactors are capable of converting the major isotope of uranium (238U > 99 percent) to fissile 239Pu in order to use it as fuel. These reactors are the breeder reactors that breed (or make) new fuel (i.e., 239Pu). This 239Pu forms, however, a major challenge for GFR due to security and nonproliferation concerns. Of the reactors popularly referred to as fast breeders, the GFR is probably one of the best choices in terms of ensuring durability of resources (and hence sustainability), but we could also assume that, due to the plutonium that runs throughout the cycle, it is one of worst choices when it comes to considerations of security and nonproliferation of nuclear weapons. GFR could also be used in the *burner* (as opposed to breeder) mode, which would help us eliminate long-term isotopes in nuclear waste. This technological path would then help us ensure the conception of the value of safety as reducing the long-term concerns for future generations.

Note: This case is drawn from Taebi and Kloosterman (2015).

the core of the reactor to melt. This does not mean that there are no safety issues involved in this type of reactor—radioactive material could also leak from an HTR-PM into the surrounding environment, raising serious safety concerns—but it does mean that we are referring to different conceptions of the value of safety in an ABWR and an HTR-PM.

We have shown that there are different types of value conflicts that will only become visible in the process of operationalization if we fully understand the technological choices involved. It is in the process of operationalization that these and other trade-offs will come to light. While much of the public debate is about the choice between different energy technologies, this case study shows that, given a particular energy technology, different designs give the same energy technology quite different implications for intergenerational justice.

This suggests that operationalizing a moral framework is not just a matter of top-down application of an abstract concept, but rather an iterative process in which the top layer of the WRE model is informed by how different values can be interpreted with respect to a particular technology. In the second case study, this was done by elaborating the notion of "social acceptance" of the technology. In the original research proposal for the technical project, social acceptance was mentioned as the criterion for the success of the project: if, at the end of project, the technology achieved social acceptance, the project would be considered a success. However, in the absence of any clear definition, this notion of social acceptance did not in the first instance provide any guidance for the technical research project. Only after the parallel ethical research started did social acceptance become a central notion. The technical researchers were asked what they thought social acceptance meant in the context of their project. Many researchers took social acceptance to mean that the technology should improve the lives of future users. In addition, quite a number of researchers took social acceptance to mean that the technology should be morally acceptable in some objective sense determined by the clinical experts and the ethicist, thus related to moral acceptability rather than to factual acceptance. It was only by moving back and forth between the various interpretations the researchers gave to the concept of social acceptance that it steered the project in the direction of greater user involvement. This suggests that context matters, not just in the considered judgments of potential stakeholders, but also in how content is given to relevant values and frameworks that are used to assess the desirability of a technology.

Considered Judgment: Which Arguments to Include?

The second challenge concerns the question of inclusiveness. As the WRE approach is a coherentist method of justification, its justificatory force is derived from the coherence between the different elements and not from a

noninferential foundation. Since pure coherentism implies that each element has the power to revise other elements, some minimum level of reliability is required. After all, it would be unwise to revise a relatively stable belief on the basis of a temporary hunch. A challenge for the WRE method is therefore to select the relevant information that can be included in the establishment of a WRE. In the methodological literature on the WRE method, this question is reduced to the question of which considered judgments to include, but it could in principle also be extended to the question of which moral principles and background theories to include. Let us first focus on how this challenge is discussed in the current literature.

The primary objection to including considered moral judgments is that they do not have the stability required to warrant the revision of other, more deliberated elements. Although Rawls refers to our capacity of judgment to warrant neutrality and correctness of our considered moral judgments,[10] this capacity is still considered a vague and ambiguous term (cf. Audi 1993; Carr 1975; Ebertz 1993; Kekes 1986; Nielsen 1982b). The reference to "neutrality" in Rawls's early work suggests that only those considered judgments can be included that do not favor one particular person, group, or value. Yet, this criterion is not evident and also depends on what one wants to achieve by applying the WRE approach. The WRE approach is subject to different interpretations, which may serve different purposes (Van der Burg and Van Willigenburg 1998). Sometimes the use of the WRE approach is intended to reach consensus on more abstract principles, for example, when discussing a political system. In those situations, the neutrality of the approach may be important and knowledge of concrete facts may be excluded from the equilibrium. Other applications of the WRE approach focus on agreement in specific situations, for example, in medical ethics. In those situations, it is important that knowledge of the specific situation is included in the equilibrium, and it would be absurd to exclude this for the reason of avoiding bias. In the context of decision-making on technological risks, local elements may be especially relevant as these local particularities often determine whether something is considered safe or just.

The existence of different versions of the approach does not necessarily frustrate the applicability of the method, but it has implications for its justificatory power. Especially when it is used as a practical method, choices need to be made about the types of beliefs and arguments to include, and such choices are inevitably selective (Van der Burg and Van Willigenburg 1998). For Rawls (1974-1975), "considered judgments" are those judgments "in which our moral capacities are most likely to be displayed without distortion" (p. 8). Unfortunately, Rawls is not explicit about how

this criterion would work in practice; that is, does it also constrain the content of the considered judgments or does it only affect how people come to their considered judgments?

In the applied ethics literature on the WRE approach that has appeared since the 1980s, most philosophers have attempted to improve the reliability of the method by putting extra demands on the credibility of the initial empirical input, that is, on the credibility of the *content* of the initial judgments (cf. Beauchamp and Childress [1992] 2001; Nielsen 1982a). Quoting early work by Rawls, Beauchamp and Childress ([1992] 2001, 398), for example, argue that only those considered judgments should be included "in which we have the highest level of confidence and the lowest level of bias (. . .), for example judgments about the wrongness of racial discrimination, religious tolerance, and political oppression." However, when we place too much focus on the initial credibility of the considered judgments, the method becomes exclusive, in the sense that some judgments are not taken into account because of their alleged bias and, therefore, lack of credibility. Not putting any demand on the initial credibility of the considered judgments makes the method unreliable as a justificatory method; but imposing overly strict demands on the initial credibility of the considered judgments may lead to the reproduction of the dominant discourse, because only those considered judgments that are in line with the dominant repertoire of arguments are considered credible (Callon, Lascoumes, and Barthe 2009).

A challenge of the WRE method is therefore to find a mechanism that both provides some minimum level of credibility to the considered judgments, on the one hand,[11] and, on the other hand, allows for the incorporation of the "broadest evidence" available when moving from the considered judgments to the levels of principles and theories (Daniels 1996, 2-3). Whereas most philosophers try to make the method reliable by putting additional demands on the credibility of the initial input, some philosophers give priority to the inclusiveness of the WRE method. Instead of trying to warrant credibility in the *content* of the arguments or information that go into the WRE method, they propose seeking credibility in the *process* of applying the method. In other words, instead of putting demands on the content of the considered judgments that go into the process, they formulate criteria indicating what "good reasoning" amounts to and, in doing so, aim to provide sufficient warrant for a reliable and inclusive outcome ("good reasoning–justified outcome strategy").[12] This "good reasoning–justified output strategy," which is also very much in line with Habermas's work on discourse ethics (1990b) and communicative action

(1990a), has an advantage over the "credible input–justified output strategy" in that it allows for the inclusion of a broader set of elements, which is the cornerstone of the method.[13]

This discussion in the applied ethics literature on inclusion and exclusion of particular arguments has its counterpart in social scientific studies of the public understanding of science. Empirical studies increasingly question the alleged divide between "biased" and "irrational" laypersons' knowledge and the "unbiased" and "rational" knowledge of experts. In his classic study of Cumbrian sheep farmers and their response to the scientific advice after the Chernobyl disaster, Wynne (1992) showed that laypeople are capable of reflection on the epistemological status of their own "local" knowledge in relation to "outside" knowledge.

Also with respect to the alleged credibility of science in the modern world, empirical research points to a much more nuanced picture of the status of scientific expert knowledge. While the abstract philosopher's appeal to "credible input" seems to presuppose some objective standard of credibility, the historian Shapin suggests that credibility should be seen as the outcome of a contingent social and cultural practice (1995, 257). Hence, credibility is not something that exists "out there" in the world before people engage in some deliberative process but is rather something that is created during this deliberative process, something that is also highly dependent on particular conditions, resources, and tactics (Nukaga 2016).

Hence, whereas the majority of philosophers still emphasize the need to put extra demands on the credibility of the information that goes into the method, the social scientific literature seems to give priority to the inclusiveness of the judgments that the method should allow.[14] As disputed knowledge often plays a constitutive role in controversies on the implementation of new technologies (McCormick 2007), excluding one source of information upfront will probably not resolve any conflict. New technologies often involve uncertainties and complexities that in many cases are underplayed in expert assessments (Jaeger et al. 2013; Stirling 2011; Collingridge 1980; Doorn and Hansson 2011), which suggests that expert knowledge itself has its own bias. Hence, based on the social science literature, the "good reasoning strategy" seems to be the preferred method and this was also chosen for the application of the WRE method in the case study presented in Text Box 2.

In this study, the participants were asked to identify the risks that their technological application would pose and also the tasks that needed to be performed to address these risks. The participants were then asked which of

Text Box 2. ALwEN Project.

The ALwEN project is an research and development project aimed at developing a prototype support tool based on Ambient Technology to monitor and assist the activities of the elderly in the retirement home setting. Four universities, two independent industrial research institutes, one clinical partner, and a consortium of twelve small- and medium-sized enterprises cooperated in the ALwEN project. In parallel to the technical project, an "ethical parallel study" was conducted to investigate how the researchers distributed the responsibility for addressing ethical issues in the project, particularly ethical issues related to technological risks, where we broadly defined the latter as all undesirable aspects related to the particular technology. The ethical parallel study started with an observation period, followed by a series of interviews with members of the technical project team. The interview results were in turn used as input for an interactive workshop organized around the WRE model. Based on the interview results, some striking issues were selected and explored in more detail during this workshop. During the workshop, participants were asked to distribute these tasks over the different phases or activities within the ALwEN project. Meeting support software was used to allow for anonymous discussions in the hope that this would give participants more freedom to speak freely and not simply give the "desirable answer" for the sake of project's atmosphere. This was done in several rounds, with discussion taking place in between. At the end of the workshop, there was significantly more agreement on the scope of the project and the question of what should be done and by whom.

In the ALwEN project, the WRE framework was used both with a constructive and a justificatory aim in mind. As a constructive method, the WRE framework was used to encourage discussion about the scope of the project and to let the participants reflect on their "moral duties" as members of the research team. The framework therefore proved very useful for encouraging reflection in the ALwEN case. It prompted discussion and encouraged people to think about the fair distribution of responsibilities and to defend their opinions to their fellow team members. As a method of justification, the WRE framework was used to evaluate the agreed-upon division of moral labor in terms of the WRE of individuals. The framework also proved successful for justificatory purposes. All participants accepted the final distribution of responsibilities as fair.

Note: This case is drawn from Doorn (2012, 2014).

these tasks should be done within the context of their research project. This prompted normative questions that were outside the researchers' comfort zone but that could not be answered by an outsider without the input of the technical researchers themselves, as they knew best whether they had the required resources and capacities. This collectively situated knowledge was derived from the actual practice of jointly working in the project and could not have been assessed by an outsider. Some risks required a change to the originally scheduled research activities. During the WRE workshop, the researchers became aware that *for moral reasons*, they should shift the focus of their work from fundamental research to experimentation to assess the actual risks in a real-world setting, which also meant a shift from primarily individual research to more joint research activities.

This discussion of what could be expected from the technical researchers also questions the popular narrative of technical researchers, and natural scientists more generally, as only being experts in their own narrow field, while philosophers are the experts on moral issues. In his historical account of the role of science and scientists in our modern world, Shapin describes how modern science, on the one hand, has effectively been enfolded in institutions aimed at the production of wealth and the projection of power, but that the modern scientist, on the other hand, has limited authority in fields other than his or her particular domain. At least from the early twentieth century, scientists themselves declared that they did not possess particular moral authority (Shapin 2008, 442). However, application of the WRE approach in the ALwEN project suggests that the distinction between the "descriptive" and "normative" is not only difficult to make, but also that researchers may have knowledge that is decisive for establishing what is morally expected from them.[15]

Applied to the context of risk-related decision-making, the following lesson could be drawn. The criterion of inclusiveness of the WRE method calls for an open discourse (Doorn 2010a). It is important that all stakeholders have equal opportunities to participate in and contribute to the decision-making process. In a conversation involving both experts and laypeople, inclusiveness may, for example, require the vocabulary used by the experts to be understandable to all and for people to feel free to introduce unwelcome arguments (Doorn 2010b). If people are discouraged from doing so and remain silent, the process followed does not warrant a reliable outcome. Consequently, the outcome of the method, whether by consensus or not, cannot be deemed *just*. In the case study presented in Text Box 2, for example, anonymous meeting software was

used for discussion in the hope that people felt free to introduce unwelcome arguments. A deliberate choice was made not to exclude certain considered judgments from the discussion but instead to encourage participants in the study to focus on the reasoning process itself. This created credibility for particular judgments about who should take action and work on particular ethical issues. It was, for example, agreed that the project management should spend much more time on initiating activities that could assess how end users would experience the technology that the team was developing and that project management should ensure that this also happened. It could, for example, also have implied that the team should be complemented by new researchers and new expertise that appeared to be lacking. Credibility here does not mean that these judgments were interpreted as "true," but rather that they reflected a distribution of responsibilities that all participants in the WRE workshop considered fair and feasible.

How to Establish Whether an Equilibrium Has Been Reached

A further practical challenge for the utilization of the WRE approach concerns the question of who decides whether or not equilibrium has been reached. Rather than an empirical method for justification, Rawls primarily developed his method as a hypothetical framework for testing his own theory of justice. However, in a relevant, but relatively unnoticed, exchange between Habermas and Rawls in the *Journal of Philosophy*, Habermas emphasized that the WRE method would not provide the stabilizing force in society that Rawls aimed for, unless people's factual acceptance of the final WRE was also tested. In his political theory, Rawls elaborated a criterion of acceptability that was neutral with respect to different views of the good life. The underlying premise is that if the theory was neutral, it would be accepted by all reasonable people. As such, Rawls argued, it would secure social stability because people would adhere to it. Habermas then criticized Rawls for not making a sufficiently sharp distinction between what people actually accept as a theory of justice and the moral acceptability that Rawls developed in his theory (Habermas 1995, especially pp. 120-22). In Habermas's interpretation, Rawls mistakenly assumed that people are automatically convinced by a theory if the theory is consistent (i.e., if the theory is deemed acceptable on abstract theoretical grounds). Habermas emphasized that while it is one thing for a theory to

be consistent or acceptable, it requires a further cognitive step for people to accept it.

With the current pleas to use the WRE method to integrate empirical studies on social acceptance in normative analyses of moral acceptability, this theoretical exchange between Rawls and Habermas becomes relevant again. The way in which the WRE method is discussed in the applied ethics literature suggests that the empirically informed considered judgments feed into the WRE method and that it is then up to an ethicist or philosopher to establish whether or not consensus or coherence has been achieved. It is then silently assumed that the people whose considered judgments have been included will also accept the final outcome as fair. In a recent inventory of applications of the WRE approach in practical situations, only one of the twelve articles discussed in the review explicitly suggested that actors must be involved in assessment of the equilibrium (Doorn 2010a). In the other cases, it was either left to the ethicist/philosopher to assess whether a reflective equilibrium or overlapping consensus had been achieved or it was left open. Musschenga (2005) argues that reflective equilibrium is not an objective state of affairs that can be determined from a third person's point of view; it usually is a first-person judgment. We concur that the question of whether someone's considered judgments, moral principles, and background theories are in reflective equilibrium, or whether an outcome is merely a matter of modus vivendi, can best be answered by the actor who has "direct access" to these considerations. In the ALwEN case (Text Box 2), the final judgment of the distribution of responsibilities was left to the different participants in the study. They were asked whether they considered the agreed-upon distribution of responsibilities to be fair. Although the participants had received an explanation about the WRE model, including the explanation that an outcome can be considered fair if it coheres with the three layers of the WRE model, they were not explicitly asked whether the final outcome was in equilibrium with their background theories and moral principles. This was considered too artificial, as few people would deliberately ask themselves whether an outcome coheres with their background theories and moral principles in order to evaluate the outcome's fairness.

When using the WRE method for decision-making on technological risks, asking stakeholders for their considered judgments on these technologies seems insufficient. If the method is to fulfill its justificatory role, the stakeholders should also be asked at the end of the process whether they personally consider the eventual outcome to be fair. Depending on the specific situation, this may not be feasible, for example, when the group

of consulted stakeholders is too large to ask them all to reflect on the outcome or when the consultation takes place over a long time span, so that there is no clearly defined endpoint. In those situations, some alternative safeguards should be implemented to justify the jump from considered judgments to justified outcomes. An example of such a safeguard is an interview or survey on the fairness of the procedure followed. Additionally, it may not always be feasible to include all stakeholders in the decision-making process. In those situations, one could argue that not all stakeholders should necessarily be included in the decision-making process itself but, as a minimum requirement of inclusiveness, all stakeholders should at least be *represented* in the decision-making process and their interests should be on the agenda (Van de Poel and Doorn 2013).

That the assessment of an outcome in terms of individual WREs is a first-person judgment has implications for the role of an ethicist applying this method. On the one hand, this is a modest role. It is not up to the ethicist to say whether or not an outcome is fair in terms of an individual's WRE, precisely because the ethicist does not have direct access to these first-person considerations. Yet, a philosopher or ethicist may act as an impartial referee and bring potential inconsistencies to the fore. Even though this role is a modest one, it has proven fruitful as shown by the ALwEN case (Text Box 2). Additionally, an ethicist can put salient issues on the agenda. The ethicist has the expertise to say that certain issues need to be addressed before a technology can be safely developed any further.

Additionally, sometimes it is possible to construct several coherent WREs, in which case there are several acceptable solutions available. It is then not up to the ethicist to decide which solution to choose. One could see these different WREs as spanning an area with acceptable solutions as the negative of Goodin's idea of output filters. Whereas "output filters can be conceptualized as barriers erected at the back end of the social decision machinery, preventing policies predicated on perverse preferences from ever emerging as settled social choices" (Goodin 1986, 78), the area covered by one or several of the WREs represents acceptable solutions. In practice, though, it may be difficult enough to find one WRE and it is doubtful whether several WREs can be established in real-life situations, so this issue is probably a theoretical matter with no practical relevance for the applicability of the WRE.

Probably, more likely is the situation in which it is impossible to arrive at a consensus and construct a WRE. One way of looking at those situations is to see WRE as an ideal worth striving for, which can be achieved to a certain degree. Sometimes it is impossible to reach an outcome that is fully

acceptable for all people. In those situations, we may need to accept that moral rightness comes in degrees and that no solution is fully right but only right to some degree (Hillerbrand and Peterson 2014). To put this in terms of the WRE, sometimes the equilibrium cannot be achieved but one would accept the best approximation of that equilibrium as the best morally defensible course of action.

WRE and Normative Social Sciences: What Role for the Normative Researcher

In view of the recent calls for interdisciplinary collaborations with which we started this article, this discussion of the WRE approach prompts a comparison with other methods. In relation to other methods that have or are currently being developed within the normative social sciences, WRE seems to have most in common with MM, although both methods have different "roots" (MM seems to be more solidly based in the social sciences, while WRE has its roots in political philosophy). Comparing both methods, we can see some common challenges but also distinct features. As to the common challenges, both methods require the person employing it to have appropriate technological knowledge. For the MM researcher, this knowledge is primarily required to increase the deliberative capacity of the technical researchers; for the WRE researcher, this knowledge is especially important to understand how moral concepts (i.e., values, theories) are operationalized. As the WRE method is primarily project- or technology-focused, it seems easier to put particular ethical issues on the agenda of the research team. For the MM researchers, their involvement is primarily focused on the individual researcher and less on a specific project. Inclusiveness is more important in the WRE method than in the MM method, which suggests that questions about representativeness (Rowe, Marsh, and Frewer 2004; Harvey 2008) and power (Stirling 2007; Cook and Kesby 2013) are also more critical when applying the WRE method than when applying the MM method. People employing the WRE method could certainly benefit from the work that has been done in this regard within the normative social sciences.

With the current appeals to "responsiveness" and "inclusiveness" in interdisciplinary research, the normativity of the WRE method may be an advantage rather than something that should be avoided. However, if the WRE method is to be used for combining normative analyses with social science studies, it must be employed more widely and also by scholars who are—by training—more sensitive to issues of power and representativeness.

When a WRE or MM researcher has a formal role in a technical project, and as such formally becomes an "insider," a combination of personal skills and institutional safeguards may especially be required to deal with the power dynamics between the technical researchers and the normative researcher. Personal skills are required to maintain relationships of trust that allow critical appraisal; clear arrangements about reporting the outcomes of the ethical investigations should be part of the institutional safeguards that make the cooperation successful (Doorn and Nihlén Fahlquist 2010).

Concluding Remarks

In this article, we have explored the use of the WRE approach in applied ethics, with a special focus on decision-making on technological risks. We started from the observation that the WRE method is currently often proposed as a promising approach for combining ethical studies on the moral acceptability of new technologies with potential risks and social scientific studies on the social acceptance of these new technologies. We noticed that the discussions in applied ethics are still rather abstract and that few real applications of the method are presented in the literature. As a result, the discussion of the pros and cons of the method remains somewhat abstract and is less applicable to the specific context of technological risks.

We discussed three issues that are currently neglected in the applied ethics literature on the use of the WRE method. The first concerned the view on how to get from the more abstract level of background theories to moral principles. This operationalization requires much more empirical information than currently assumed. We illustrated this with an example from engineering design, but this applies equally to the context of health care, the domain in which the WRE method is most often seen as a promising method for resolving moral problems.

The second issue concerned the inclusiveness of the method and the demand for openness. We argued that the applied ethics literature focuses too much on the initial credibility of the arguments that people are allowed to introduce. If the method is to be used to resolve deeply rooted conflicts about technological risks, it may be more productive to focus on the inclusiveness of the method. As well as being more productive, it is also more in line with the moral justification that the method is supposed to provide.

The third issue concerned the establishment of whether a WRE has been reached. We wrote this article with the intention of providing tools that are helpful in *real-world* situations and not just thought experiments. In those situations, one cannot sidestep the question of how to conclude whether or

not equilibrium has been reached. The ethicist could play a salient role in putting important issues on the agenda but the decision whether a WRE has been reached is up to the stakeholders. It is also important to note that it is likely that the WRE cannot always be fully achieved in real-world situations. Sometimes the WRE is an *ideal* state and a best approximation of that ideal will suffice as the morally defensible course of action.

Acknowledgments

We would like to thank Peter Kroes, Maarten Franssen, and two anonymous reviewers for their helpful feedback on earlier drafts of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work of Neelke Doorn is supported by the Netherlands Organization for Scientific Research (NWO) under grant number 275-20-041. The work of Behnam Taebi is supported by the NWO under grant number 275-20-040.

Notes

1. In this article, we have deliberately chosen not to provide a concrete definition of "risk," as it would position us too much in a particular risk discourse. We take technological risks to refer to the negative aspects of a particular technology, thereby departing from the more common engineering definition of risk as "probability times effect" (Doorn and Hansson 2015).
2. Compare Nukaga (2016) and various contributions in the special issues on empirical ethics in the journals *Health Care Analysis* (Holm 2003); *Medicine, Health Care and Philosophy* (Borry, Schotsmans, and Dierickx 2004a); and *Bioethics* (Molewijk and Frith 2009).
3. We do not take a position here on the desirability of these interdisciplinary collaborations. The aim of this article is to discuss the wide reflective equilibrium (WRE) as a method that fits this trend, which we consider, in the context of this article, as a given.
4. See Doorn and Nihlén Fahlquist (2010) for a more detailed description of ethical investigations of this type (cf. Fisher, Mahajan, and Mitcham 2006).
5. The description of Rawls's WRE approach in applied ethics is largely drawn from Doorn (2010a).

6. See Peterson (2017) for an overview of these approaches relevant to the ethics of technology, especially technologies that pose new risks.
7. For a defense of empirical ethics approaches against the naturalistic fallacy, the reader is referred to De Vries and Gordijn (2009) and Molewijk et al. (2004).
8. For a detailed discussion of intergenerational justice and safety in relation to nuclear energy production and nuclear waste management, see Taebi and Kadak (2010) and Taebi (2012).
9. The question becomes more intricate when we consider the fact that different people will be living in different periods of time in the future and that those different future interests could also potentially collide (Kermisch 2016). This question is also highly relevant to broader questions concerning which nuclear fuel cycle to choose from the perspective of ensuring the interest of future generations; the choice for a reactor is a determining factor for the fuel cycle (Taebi and Kadak 2010).
10. In his 1951 paper, "Outline of a Decision Procedure for Ethics," Rawls provided an extensive list of conditions under which people come to hold a valid considered moral judgment (Rawls 1951, 181-83). These conditions included criteria to warrant neutrality and correctness. Rawls refers to this personal correctness as "certitude," which he explicitly distinguishes from certainty. Certitude refers to a person-bound characteristic of a judgment indicating that it "[. . .] is felt to be certain by the person making it" (Rawls 1951, 182). In his later work, Rawls defined considered moral judgments as those judgments "given under conditions in which our capacity for judgment is most likely to have been fully exercised and not affected by distorting influences" (Rawls 2001, 29). In this work, Rawls no longer includes requirements for attaining some form of neutrality. The conditions for capacity of judgment together with a presumed desire to reach a correct decision are supposed to warrant a minimum level of credibility of the considered moral judgments.
11. It should be emphasized that this point of "initial credibility" may equally apply to principles and background theories. In the applied ethics literature, it is mistakenly assumed that only considered judgments can be biased or "irrational." Insights from science and technology studies (STS) show that more abstract theories and principles may also contain biases.
12. In their particular reflective equilibrium method, Van Thiel and Van Delden replace the notion of considered moral judgment with that of moral intuition. Based on the work of the moral psychologist Haidt (2001), they defend an interpretation of moral intuition as a response reflecting people's initial reactions when confronted with a moral case. These moral intuitions indicate which "direction" a judgment about a given case should take (Van Thiel and Van Delden 2010, 189).

13. For a more elaborate description of this strategy, see DePaul (1993) and Van Thiel and Van Delden (2009).
14. With its focus on inclusiveness, the empirical (STS) literature, in turn, has not fully explored how and under what conditions credibility is ascribed to the "diverse theories of justice that provide material principles" (Beauchamp 2003b, 26). For an exception, see Nukaga (2016).
15. This could also feed into discussions on engineering ethics generally. In their seminal article on engineering ethics, Lynch and Kline emphasized the need to improve the engineers' "ability to identify ethically problematic issues in a poorly structured problem field within an institutionally and culturally constrained set of tacit assumptions" (Lynch and Kline 2000, 209). The case study suggests that this ability also requires that engineers should recognize these constraints in the first place and understand what they can be done to enlarge their room for maneuver and address ethically problematic issues.

References

Achterhuis, H. ed. 2001. *American Philosophy of Technology. The Empirical Turn*. Bloomington: Indiana University Press.

Audi, R. 1993. "Ethical Reflectionism." *The Monist* 76 (3): 295-315.

Beauchamp, T. L. 2003a. "The Nature of Applied Ethics." In *A Companion to Applied Ethics*, edited by R. G. Frey and C. H. Wellman, 1-16. Malden, MA: Blackwell.

Beauchamp, T. L. 2003b. "The Origins, Goals, and Core Commitments of the Belmont Report and Principles of Biomedical Ethics." In *The Story of Bioethics: From Seminal Works to Contemporary Explorations*, edited by J. K. Walter and E. P. Klein, 17-46. Washington, DC: Georgetown University Press.

Beauchamp, T. L., and J. F. Childress. [1992] 2001. *Principles of Biomedical Ethics*, 5th ed. New York: Oxford University Press.

Beck, U. 1992. *Risk Society: Towards a New Modernity*. London, UK: Sage.

Bijker, W. E. 1995. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.

Borry, P., P. Schotsmans, and K. Dierickx. 2004a. "Editorial-Empirical Ethics: A Challenge to Bioethics." *Medicine, Health Care and Philosophy* 7 (1): 1-3.

Brey, P. 2010. "Philosophy of Technology after the Empirical Turn." *Techné* 14 (1): 36-48.

Callon, M., P. Lascoumes, and Y. Barthe. 2009. *Acting in an Uncertain World: An Essay on Technical Democracy*. Translated by G. Burchell. Cambridge, MA: MIT Press.

Carr, S. 1975. "Rawls, Contractarianism and Our Moral Intuitions." *The Personalist* 56 (1): 83-95.

Collingridge, D. 1980. *The Social Control of Technology*. New York: St. Martin's Press.

Cook, B. R., and M. Kesby. 2013. "The Persistence of 'Normal' Catchment Management Despite the Participatory Turn: Exploring the Power Effects of Competing Frames of Reference." *Social Studies of Science* 43 (5): 754-79.

Cotton, M. 2009. "Ethical Assessment in Radioactive Waste Management: A Proposed Reflective Equilibrium-based Deliberative Approach." *Journal of Risk Research* 12 (5): 603-18.

Daniels, N. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice, Cambridge Studies in Philosophy and Public Policy*. Cambridge, MA: Cambridge University Press.

De Vries, R., and B. Gordijn. 2009. "Empirical Ethics and Its Alleged Meta-ethical Fallacies." *Bioethics* 23 (4): 193-201.

DePaul, M. R. 1993. *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry*. London: Routledge.

Dignum, M., A. Correljé, E. Cuppen, U. Pesch, and B. Taebi. 2016. "Contested Technologies and Design for Values: The Case of Shale Gas." *Science and Engineering Ethics* 22 (4): 1171-91.

Doorn, N. 2010a. "Applying Rawlsian Approaches to Resolve Ethical Issues: Inventory and Setting of a Research Agenda." *Journal of Business Ethics* 91 (1): 127-43.

Doorn, N. 2010b. "A Procedural Approach to Distributing Responsibilities in R&D Networks." *Poiesis & Praxis. International Journal of Technology Assessment and Ethics of Science* 7 (3): 169-88.

Doorn, N. 2012. "Exploring Responsibility Rationales in Research and Development (R&D)." *Science, Technology, & Human Values* 37 (3): 180-209.

Doorn, N. 2013. "Wide Reflective Equilibrium as a Normative Model for Responsible Governance." *NanoEthics* 7 (1): 29-43.

Doorn, N. 2014. "Assessment of an Ambient-assisted-living Project: Between Technology and Application." In *Responsible Innovation Volume 1: Innovative Solutions for Global Issues*, edited by M. J. Van den Hoven, N. Doorn, T. Swierstra, B.-J. Koops, and H. Romijn, 301-14. Dordrecht, the Netherlands: Springer.

Doorn, N., and S. O. Hansson. 2011. "Should Probabilistic Design Replace Safety Factors?" *Philosophy & Technology* 24 (2): 151-68.

Doorn, N., and S. O. Hansson. 2015. "Design for the Value of Safety." In *Handbook of Ethics and Values in Technological Design*, edited by M. J. Van den Hoven, P. Vermaas, and I. R. Van de Poel. Dordrecht, 491-511. the Netherlands: Springer.

Doorn, N., and J. A. Nihlén Fahlquist. 2010. "Responsibility in Engineering: Towards a New Role for Engineering Ethicists." *Bulletin of Science, Technology & Society* 30 (3): 222-30.

Doorn, N., D. Schuurbiers, I. R. van de Poel, and M. E. Gorman. Eds. 2013a. *Early Engagement and New Technologies: Opening Up the Laboratory*. Dordrecht, the Netherlands: Springer.

Doorn, N., D. Schuurbiers, I. R. van de Poel, and M. E. Gorman. 2013b. "Early Engagement and New Technologies: Towards Comprehensive Technology Engagement?" In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by N. Doorn, D. Schuurbiers, I. R. Van de Poel, and M. E. Gorman, 231-51. Dordrecht, the Netherlands: Springer.

Ebertz, R. P. 1993. "Is Reflective Equilibrium a Coherentist Model?" *Canadian Journal of Philosophy* 23 (2): 193-214.

Fisher, E., R. L. Mahajan, and C. Mitcham. 2006. "Midstream Modulation of Technology: Governance from Within." *Bulletin of Science, Technology & Society* 26 (6): 485-96.

Fisher, E., and D. Schuurbiers. 2013. "Socio-technical Integration Research: Collaborative Inquiry at the Midstream of Research and Development." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by N. Doorn, D. Schuurbiers, I. R. Van de Poel, and M. E. Gorman, 97-110. Dordrecht, the Netherlands: Springer.

Giddens, A. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.

Goodin, R. E. 1986. "Laundering Preferences." In *Foundations of Social Choice Theory*, edited by J. Elster and A. Hylland, 75-102. Cambridge, MA: Cambridge University Press.

Habermas, J. 1990a. *Moral Consciousness and Communicative Action*. Cambridge, MA: MIT Press.

Habermas, J. 1990b. "Discourse Ethics: Notes on a Program of Philosophical Justification." In *The Communicative Ethics Controversy*, edited by S. Benhabib and F. Dallmayr, 60-110. Cambridge, MA: MIT Press.

Habermas, J. 1995. "Reconciliation through the Public Use of Reason: Remarks on John Rawls's Political Liberalism." *The Journal of Philosophy* 92 (3): 109-31.

Hart, H. L. A. 1961. *The Concept of Law*. Oxford, UK: The Clarendon Press.

Haidt, J. 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment." *Psychological Review* 108 (4): 814-34.

Harvey, M. 2008. "Drama, Talk, and Emotion Omitted Aspects of Public Participation." *Science, Technology, & Human Values* 34 (2): 139-61.

Hillerbrand, R., and M. Peterson. 2014. "Nuclear Power Is Neither Right Nor Wrong: The Case for a Tertium Datur in the Ethics of Technology." *Science and Engineering Ethics* 20 (2): 583-95.

Holm, S. 2003. "Introduction." *Health Care Analysis* 11 (1): 1-2.

Huijts, N., E. Molin, and L. Steg. 2012. "Psychological Factors Influencing Sustainable Energy Technology Acceptance: A Review-based Comprehensive Framework." *Renewable and Sustainable Energy Reviews* 16 (1): 525-31.

Jaeger, C. C., T. Webler, E. A. Rosa, and O. Renn. 2013. *Risk, Uncertainty and Rational Action*. London, UK: Routledge.

Kekes, J. 1986. "Moral Intuition." *American Philosophical Quarterly* 23 (1): 83-93.

Kermisch, C. 2016. "Specifying the Concept of Future Generations for Addressing Issues Related to High-level Radioactive Waste." *Science and Engineering Ethics* 22 (6): 1797-811.

Knorr-Cetina, K. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford New York: Pergamon Press.

Kroes, P. and A. Meijers. eds. 2000. *The Empirical Turn in the Philosophy of Technology*. Amsterdam, the Netherlands: JAI.

Latour, B. 2005. *Reassembling the Social: An Introduction to Actor–network Theory*. Oxford, UK: Oxford University Press.

Lynch, W. T., and R. Kline. 2000. "Engineering Practice and Engineering Ethics." *Science, Technology, & Human Values* 25 (2): 195-225.

McCormick, S. 2007. "Democratizing Science Movements: A New Framework for Mobilization and Contestation." *Social Studies of Science* 37 (4): 609-23.

Molewijk, B., and L. Frith. 2009. "Empirical Ethics: Who is the Don Quixote?" *Bioethics* 23 (4): II-IV. doi:10.1111/j.1467-8519.2009.01707.x.

Molewijk, B., A. M. Stiggelbout, W. Otten, H. M. Dupuis, and J. Kievit. 2004. "Empirical Data and Moral Theory: A Plea for Integrated Empirical Ethics." *Medicine, Health Care and Philosophy* 7 (1): 55-69.

Musschenga, A. W. 2005. "Empirical Ethics, Context-sensitivity, and Contextualism." *Journal of Medicine and Philosophy* 30 (5): 467-90.

Nielsen, K. 1982a. "Grounding Rights and a Method of Reflective Equilibrium." *Inquiry* 25 (3): 277-306.

Nielsen, K. 1982b. "Considered Judgements Again." *Human Studies* 5 (2): 117-29.

Nukaga, Y. 2016. "Ethics Expertise and Public Credibility: A Case Study of the Ethical Principle of Justice." *Science, Technology, & Human Values* 41 (4): 709-31.

Orlikowski, W. J. 2007. "Sociomaterial Practices: Exploring Technology at Work." *Organization Studies* 28 (9): 1435-48.

Owen, R., J. Bessant, and M. Heintz. 2013. *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. Chichester, UK: Wiley.

Peterson, M. 2017. *The Ethics of Technology: A Geometric Analysis of Five Moral Principles*. Oxford, UK: Oxford University Press.

Rawls, J. 1951. "Outline of a Decision Procedure for Ethics." *Philosophical Review* 60 (2): 177-97.

Rawls, J. 1974-1975. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association* 48:5-22.

Rawls, J. 1999. *Collected Papers*, Rev. ed. Cambridge, MA: Harvard University Press.

Rawls, J. [1971] 1999. *A Theory of Justice*, Rev. ed. Cambridge, MA: The Belknap Press of Harvard University Press.

Rawls, J. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: The Belknap Press of Harvard University Press.

Rip, A., and D. K. R. Robinson. 2013. "Constructive Technology Assessment and the Methodology of Insertion." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by N. Doorn, D. Schuurbiers, I. R. Van de Poel, and M. E. Gorman, 37-53. Dordrecht, the Netherlands: Springer.

Rowe, G., R. Marsh, and L. J. Frewer. 2004. "Evaluation of a Deliberative Conference." *Science, Technology, & Human Values* 29 (1): 88-121.

Schot, J. W., and A. Rip. 1997. "The Past and Future of Constructive Technology Assessment." *Technological Forecasting and Social Change* 54 (2/3): 251-68.

Schuurbiers, D., N. Doorn, I. R. van de Poel, and M. E. Gorman. 2013. "Mandates and Methods for Early Engagement." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by N. Doorn, D. Schuurbiers, I. R. Van de Poel, and M. E. Gorman, 3-14. Dordrecht, the Netherlands: Springer.

Shapin, S. 1995. "Cordelia's Love: Credibility and the Social Studies of Science." *Perspectives on Science* 3 (3): 255-75.

Shapin, S. 2008. "Science and the Modern World." In *The Handbook of Science and Technology Studies*, edited by E. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman, 429-48. Cambridge, MA: MIT Press.

Sismondo, S. 2008. "Science and Technology Studies and an Engaged Program." In *The Handbook of Science and Technology Studies*, edited by E. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman, 13-30. Cambridge, MA: MIT Press.

Stirling, A. 2007. ""Opening Up" and "Closing Down": Power, Participation, and Pluralism in the Social Appraisal of Technology." *Science, Technology, & Human Values* 33 (2): 262-94.

Stirling, A. 2011. "Pluralising Progress: From Integrative Transitions to Transformative Diversity." *Environmental Innovation and Societal Transitions* 1 (1): 82-99.

Taebi, B. 2012. "Multinational Nuclear Waste Repositories and Their Complex Issues of Justice." *Ethics, Policy & Environment* 15 (1): 57-62.

Taebi, B., and A. C. Kadak. 2010. "Intergenerational Considerations Affecting the Future of Nuclear Power: Equity as a Framework for Assessing Fuel Cycles." *Risk Analysis* 30 (9): 1341-62.

Taebi, B., and J. L. Kloosterman. 2015. "Design for Values in Nuclear Technology." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory,*

Values and Application Domains, edited by M. J. Van den Hoven, P. Vermaas, and I. R. Van de Poel. Dordrecht, 805-29. the Netherlands: Springer.

Taebi, B. Forthcoming. "Bridging the Gap between Social Acceptance and Ethical Acceptability." *Risk Analysis*. doi:10.1111/risa.12734.

Van de Poel, I. R. 2016. "A Coherentist View on Social Acceptance and Moral Acceptability of Technology." In *Philosophy of Technology after the Empirical Turn*, edited by M. Franssen, P. Vermaas, and P. A. Kroes. Dordrecht, 177-93. the Netherlands: Springer.

Van de Poel, I. R., and N. Doorn. 2013. "Ethical Parallel Research: A Network Approach for Moral Evaluation (NAME)." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by N. Doorn, D. Schuurbiers, I. R. Van de Poel, and M. E. Gorman, 111-36. Dordrecht, the Netherlands: Springer.

Van de Poel, I. R., and S. D. Zwart. 2010. "Reflective Equilibrium in R&D networks." *Science, Technology, & Human Values* 35 (2): 174-99.

Van den Hoven, M. J., N. Doorn, T. Swierstra, B.-J. Koops, and H. Romijn. 2014. *Responsible Innovation: Innovative Solutions for Global Issues*. Dordrecht, the Netherlands: Springer.

Van der Burg, W., and T. Van Willigenburg. 1998. "Introduction." In *Reflective Equilibrium: Essays in the Honour of Robert Heeger*, edited by W. Van der Burg and T. Van Willigenburg, 1-25. Dordrecht, the Netherlands: Kluwer Academic.

Van Thiel, G. J. M. W., and J. J. M. Van Delden. 2009. "The justificatory power of moral experience." *Journal of Medical Ethics* 35 (4): 234-37.

Van Thiel, G. J. M. W., and J. J. M. Van Delden. 2010. "Reflective Equilibrium as a normative empirical model." *Ethical Perspectives* 17 (2): 183-202.

Wildson, J., and R. Willis. 2004. *See-through Science: Why Public Engagement Needs to Move Upstream*. London, UK: Demos.

Wustenhagen, R., M. Wolsink, and M. J. Burer. 2007. "Social Acceptance of Renewable Energy Innovation: An Introduction to the Concept." *Energy Policy* 35 (5): 2683-91.

Wynne, B. 1992. "Misunderstood Misunderstanding: Social Identities and Public Uptake of Science." *Public Understanding of Science* 1 (3): 281-304.

Author Biographies

Neelke Doorn is full professor of "ethics of water engineering" at Delft University of Technology. She holds master's degrees in civil engineering, philosophy, and law and a PhD degree in philosophy of technology. Her current research concentrates on moral issues in technological risk and water governance, with a special focus on how the current resilience paradigm in water management and climate adaptation

involves a transition of responsibilities from government to citizens. She is the coeditor in chief of *Techné: Research in Philosophy and Technology* (*Journal of the Society for Philosophy and Technology*) and coeditor of two volumes on responsible innovation and early engagement methods in technology development.

Behnam Taebi is associate professor in ethics of technology at Delft University of Technology and an associate with the Harvard Kennedy School's Belfer Center for Science and International Affairs. He studied material science and engineering and received his PhD in philosophy of technology. He is the coordinating editor of a volume on *The Ethics of Nuclear Energy* (Cambridge University Press, 2015) and a special issue of *Journal of Risk Research* (2015) on "Socio-Technical Challenges of Nuclear Power Production." He is currently editing a special issue with the journal *Sustainability* on "Sustainability and Ethics: Reflections on the UN Sustainable Development Goals" and writing a monograph on "Ethics and Engineering" (under contract with Cambridge University Press).