Moral Responsibility in R&D Networks

Networks

A Procedural Approach to Distributing Responsibilities

Moral Responsibility in R&D Networks

A Procedural Approach to Distributing Responsibilities

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 10 mei 2011 om 15:00 uur

door Neelke DOORN
civiel ingenieur
doctorandus in de wijsbegeerte

geboren te 's-Gravenhage.

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. M.J. van den Hoven

Copromotor:
Dr. ir. I.R. van de Poel

Samenstelling promotiecommissie:
Rector Magnificus, Technische Universiteit Delft, voorzitter
Prof. dr. M.J. van den Hoven, Technische Universiteit Delft, promotor
Dr. ir. I.R. van de Poel, Technische Universiteit Delft, copromotor
Prof. dr. ir. S. van der Zwaag, Technische Universiteit Delft
Prof. dr. F.W.A. Brom, Rathenau Instituut
Prof. dr. G.J. van der Wilt, Universitair Medisch Centrum St. Radboud
Prof. dr. M. Davis, Illinois Institute of Technology
Prof. dr. S.O. Hansson, Kungliga Tekniska Högskolan
Prof. dr. ir. P.A. Kroes, Technische Universiteit Delft, reservelid

Contents

List of papers

Chapter 2

Doorn, N. (2011) 'Responsibility ascriptions in technology development and engineering: Three perspectives', in: *Science and Engineering Ethics*: DOI: 10.1007/s11948-009-9189-3.

Chapter 3

Doorn, N. (2010) 'Applying Rawlsian approaches to resolve ethical issues: Inventory and setting of a research agenda', in: *Journal of Business Ethics* 91 (1): 127-143.

Chapter 4

Doorn, N. (forthcoming) 'Exploring rationales for distributing moral responsibilities in R&D networks', accepted for publication in *Science, Technology, & Human Values*.

Chapter 5

Doorn, N. (2010) 'A procedural approach to distributing responsibilities in R&D Networks', in: *Poiesis & Praxis. International Journal of Technology Assessment and Ethics of Science* 7 (3): 169-188.

Chapter 6

Doorn, N. and Nihlén Fahlquist, J. A. (2010) 'Responsibility in engineering: Towards a new role for engineering ethicists', in: *Bulletin of Science, Technology & Society* 30 (3): 222-230.

Acknowledgments

There are many people I wish to thank for helping me make this PhD project possible, more than I will be able to mention here explicitly. First and foremost, I would like to thank my supervisor Ibo van de Poel at Delft University of Technology. Ibo, you are the best supervisor one could possibly wish: friendly, always there when needed, and constructive when searching for the right structure of some draft text. I feel honored having you as my supervisor and copromotor and I am very happy to continue working with you on the NanoNextNL project.

The next person I would like to thank is my promotor Jeroen van den Hoven. Jeroen, you are an admirable philosopher: critical in your work but at the same time thinking in terms of solutions rather than problems. I think you are the exemplar of the 21st century philosopher; away from the ivory tower and firmly grounded in the real world.

I would like to thank the project members of our responsibility project: Ibo, Jessica, Lambèr, Rosja, Sjoerd, and Tiago. I enjoyed working with you and I appreciate the attempts of "you formal guys" of making our concepts clear. Thanks also goes to Anthony Meijers, who offered me a working place at Eindhoven University of Technology in the first two years of the project. It resulted in a co-authored article with Lambèr Royakkers and Rob Raven, whom I would like to thank for the nice cooperation.

While I was working on my PhD project, I spent the spring 2010 term as a visiting researcher at the Philosophy Department of the Royal Institute of Technology (KTH) in Stockholm. I would like to thank Sven Ove Hansson for the hospitality and the cooperation. Tack till mina kolleger Dan, Jessica, Karin, Misse och Sara för att ni fick mig att känna mig hemma. Och tack till Jessica för att du hjälpte mig att förbättra min svenska text. I enjoyed my time in Stockholm and it provided me with the opportunity to present my work to a new audience. Thanks for all the social and professional support.

A big thanks goes to my colleagues at the Philosophy Department. I would like to thank Peter and Jeroen for providing me with this research position and for being very generous in supporting various visits abroad. Henneke and Diana, thank you for all kinds of practical and organizational support. My fellow PhD students – Behnam, Chris, Christine, Dingmar, Ilse, Luca, Maartje, Malik, and

Noëmi but also the fresh group of PhD students who recently started their PhD research while I was finalizing my manuscript – thanks for being such great colleagues. I enjoyed our meetings, both the formal and informal ones. Saskia, Nicole, Job, and Sofia, thanks for being such pleasant roommates. I appreciated our stimulating conversations on responsibility, gender and LGBT issues, and Brompton folding bikes. Daan and Ibo, thanks for the enjoyable cooperation in the responsible innovation project.

In the last year of my PhD, I started the trajectory to become a certified university teacher (BKO), which I hope to finish in June 2011. A special thanks goes to my BKO colleague Francien. I am very lucky to have you as a companion on the both inspiring and partially bumpy BKO road.

I would also like to thank the participants in the ALwEN project. I cannot name them all, but Frits, thanks for introducing me in the project and encouraging the others to lend me every possible support in doing my research. Thanks also to all the researchers who were so kind to participate in the workshop.

During my PhD trajectory, I participated in the graduate program of the Dutch Research School for Applied Philosophy. I very much enjoyed the seminars and the courses, which were not only instructive concerning philosophy but also a perfect opportunity for sharing experiences. André, Boukje, David, Eline, Frederike, Lonneke, Marieke, and Pieter, I enjoyed working with you in the PhD council.

Outside the professional context, there are many people who in some way contributed to my PhD project, either by asking philosophical questions or by not asking them. I cannot list them all individually, but I hope they know how much I appreciate having them with me. Thanks is due to my friends, including colleagues who have become friends as well. A special thanks goes to Ilse, for providing me with a sleeping place every Monday evening. Ilse, thanks for being such a great friend. I think I owe you liters of pumpkin soup by now.

Then my *paranymphs* Baukje and Raimond, our lives have become connected in ways we could not have foreseen when we first met long ago. Our friendship means a lot to me and I hope to enjoy it for many years to come.

Thanks goes to my parents, who supported me throughout my career: from engineering to philosophy. Petra, although living in Mozambique, you always feel close. I hope soon to compete in the favorite aunt race again and spoil Teun, Freek, Elisa and Koen with presents and activities they are not likely to do with other grown-ups.

Lastly Mathilde, thanks for travelling with me. A PhD thesis is probably the worst place to express what you mean to me. I cannot imagine having done this without you by my side. Let's get old and stay young together!

1 Introduction

In 1982, a Texan television station commissioned the construction of a new major TV tower with antenna.[1] The antenna was designed and manufactured by Antenna Engineering, Inc., a moderately-sized local firm. The small local firm Riggers, Inc. was contracted to raise and assemble the antenna. During the initial design, Antenna Engineering submitted plans to Riggers for their approval. The plans included the placement of special hoisting lugs, which provided an attachment point for lifting cables. These cables were intended for removing the antenna sections from the delivery truck, and for hoisting the sections into the air for assembly. The plans were approved by Riggers. A crew of seven experienced riggers was employed for assembling the antenna. The crew used a vertical climbing crane, mounted on the already-constructed portion of the tower to lift each section of the antenna into the air. Each section was lifted and strapped onto the tower, one piece at a time. The placement of the antenna sections proceeded as planned, until the last section arrived. This section was different from the other sections of the antenna, because it had microwave baskets attached to the side of the antenna. The baskets interfered with the lifting cables when the antenna was rotated to a vertical position for bolting onto the previous section. Riggers asked Antenna Engineering for redesign assistance to fix the problem, but Antenna Engineering declined. Riggers then requested permission to remove the microwave baskets to clear the lifting cables, which was refused. At a previous job, Antenna Engineering had allowed removal of the microwave baskets and they were not reinstalled properly, for which they were at that time held liable. As a result, Riggers devised its own solution, without consulting an engineer, although the firm did request a review of their plans by Antenna Engineering. This review request by Riggers was refused by Antenna Engineering due to their potential liability in the event of a mishap if they had

---

[1]   This case was prepared by Mike Rabins, Charles Harris, Michael Pritchard, Lee L. Lowery, Jr. and others on a NSF grant, for use in the ethics education of undergraduate engineering students at Texas A&M University. Because of the clear description by Rabins et al., most of the text is transcribed literally from the full case description rather than paraphrased. The full case description is available at http://ethics.tamu.edu/ethics/tvtower/tv3.htm (last accessed, December 7, 2010).

commented on the plans. A make-shift extension to the lifting lug had to be fashioned by the riggers to permit the last section's vertical hoisting. When the last section was hoisted, something went wrong. The bolts on the make-shift lifting lug extension failed, as a result of which the antenna collapsed, killing five people. After the accident, the bolts were tested and it was found that the shear strength of the bolt material was half of what it should have been. The bolt manufacturer was sued. Antenna Engineering and Riggers suffered no monetary loss for the accident. To avoid bad publicity, the bolt manufacturer settled out of court and the case was closed. Analysis of the calculations made by Riggers for designing the make-shift extension later showed that Riggers had underestimated the stress in the bolts by a factor of 7.

In this historical example, there are so many people involved that it is difficult to identify a person responsible for the accident. Even though some people may be held responsible for some actions or outcomes (the bolt manufacturer, for example, may be held responsible for producing weak bolts), it is unclear who is responsible for the accident as a whole, if there is any. In the literature this is known as the problem of many hands (Thompson: 1980). Thompson formulated the problem in the context of the moral responsibility of public officials. Because many different officials, at various levels and in various ways, contribute to policies and the decisions of the organization it is difficult to ascribe moral responsibility for the organization's conduct. For outsiders of an organization who want to hold someone responsible for a certain conduct, it is particularly difficult or even impossible to find any person who can be said to have independently formed and carried out a certain policy or taken some decision. In the example of the collapse of the TV tower, a similar situation occurred. Can we say that Riggers, Inc. is responsible, because it made a flawed design of the make-shift extension? Or should we say that the responsibility lies with the liable party, the bolt manufacturer? Or should we maybe attribute the responsibility to Antenna Engineering, Inc. because the firm refused to review the design despite the expertise of its people. These questions cannot be easily answered in organizationally complex situations like the one of the TV tower collapse.

The focus in this thesis is on moral responsibility in Research and Development (R&D) networks.[2] R&D is a particularly interesting context for studying responsibility related issues for several reasons. First, R&D includes a broad range of activities. It is mostly understood as that part of the innovation chain that builds upon the results of fundamental research and is followed by product development (see Figure 4.1).[3] Because of this broad scope, R&D involves many different actors and institutions, working together in complex cooperative ventures (network organizations). Similar to the example of the TV tower collapse, these network organizations often lack a strict hierarchy and a clear task division (Rogers and Bozeman: 2001; Saari and Miettinen: 2001). This increases the chance that certain aspects are not correctly looked after because people expect someone else to do or have done it.
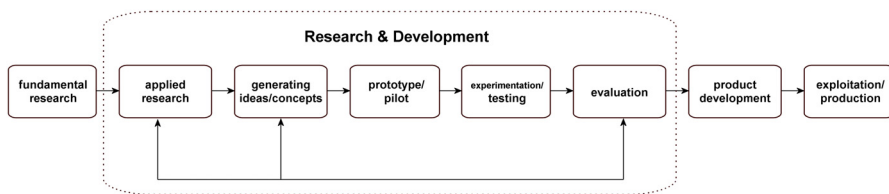
Figure 1.1: Innovation chain

Secondly, the whole trajectory from fundamental research to commercial exploitation has a long lead time. The negative consequences (risks and unforeseen side-effects) often only materialize during use (Swierstra and Jelsma: 2006), which makes it desirable to pay special attention to potential risks and

---

[2] In this thesis, I use the term R&D network to refer to professional teams working on a common project and not to the wider scientific community (which is sometimes also referred to as network).

[3] In a more verbal way, R&D can be described as: mostly building upon *existing knowledge*, aimed at finding *creative and innovative ideas* to design new *products* (or ways to improve existing ones) with the ultimate aim of *commercial exploitation*. R&D often includes an *experimentation phase* where prototypes are tested and pilot studies are carried out. Two notes regarding the schematic picture are in place. The R&D box is deliberately indicated with a dotted line because some take R&D to include fundamental research and product development as well (cf. OECD: 2010). Secondly, the different stages in the innovation chain do not necessarily follow one another in a linear way. Results at a later stage, for example, may bring about new input for fundamental research.

uncertainties already in an early phase of the trajectory and to look beyond the obvious and known consequences. This is further strengthened by the fact that the impact of technology, including its negative consequences, is often high (e.g., the use of asbestos, CFCs, DDT, nuclear waste and the greenhouse effect). Thirdly, the pace of technological development is increasing, often running ahead of adequate legislation (Schinzinger: 1998). In other words, while the impact of technology and the pace of its development increase, the contexts in which technology is developed increasingly blur the responsibility for societal consequences. These developments indicate that responsibility issues may become problematic in R&D networks.

The objective of the research presented in this thesis is to contribute to resolving responsibility issues in R&D, especially those issues stemming from different conceptions of moral responsibility. Two points in this objective are important to keep in mind. First, it is about *moral* responsibility and not responsibility in general. Secondly, it is aimed at solving problems related to *different conceptions* of moral responsibility. To explain the first point, let us go back to the example of the TV tower collapse. Responsibility is a notoriously complex term and it comes in several forms. Soon after the accident, the bolt manufacturer was held legally responsible (liable) for manufacturing flawed bolts. When it comes to the organizational responsibility (or the division of labor in the construction process), the primary responsibility lay with Riggers. However, does this mean that there was no responsibility at all for Antenna Engineering? Although Antenna Engineering did not act against the law or against the contract, they could have acted in a different, maybe more desirable, way. They could, for example, have helped with the (review of the) design of the make-shift extension or they could have allowed the removal of the microwave baskets. In both cases, the accident would probably not have happened. In addition to these more formalized types of responsibility, there is also individual moral responsibility; for example, the moral responsibility not to harm people. Apparently, there is more to say about responsibility than what is prescribed by law, organizational tasks, or job assignments. The same holds for the context of R&D. If we take the field of biotechnology, for example, the rapid unfolding of genetic codes poses serious problems related to patenting, which cannot be addressed by the current intellectual property laws (Koepsell: 2009). Does this mean that the researchers working in this field can pursue their own objectives without looking at the societal consequences? Or do they have a moral responsibility to prevent abuse (for example, preventing developments that

hinder therapeutic research)? And if they have such a responsibility, whose particular responsibility is it: the researcher in the laboratory, the person developing new ideas for applications, or is it maybe everyone's responsibility? For the moment, I capture the broader set of obligations, tasks or duties that are mandatory or wrong from a *moral* point of view under the heading of moral responsibility, to distinguish it from legal and organizational responsibility (Martin and Schinzinger: 2005).[4]

This brings us to the second point. As said above, R&D is almost always a joint undertaking so it not clear from the start who is responsible for doing what. This is often already the case when we look at organizational responsibilities and it holds even more for moral responsibilities because people hold different conceptions of moral responsibility. These different conceptions may lead to different – and sometimes even conflicting – opinions on who is responsible for addressing a particular moral issue. The objective of this thesis is therefore to alleviate the tension between these different conceptions of moral responsibility with the ultimate aim of reducing the problem of many hands.

The outline of this introductory chapter is as follows. In Section 1.1, I turn to the engineering ethics literature to get a clearer picture of the notion of moral responsibility in the context of R&D. In Section 1.2, I explain how procedural

---

[4]   In a professional setting, the term professional responsibility is sometimes introduced to refer to the special moral obligations of professionals that are above and beyond those applied to ordinary people. These professional obligations arise from voluntarily claiming membership in an organization (Davis: 2003). This definition presupposes a well-defined conception of profession. Although there are different definitions possible, I follow Michael Davis' definition of profession as "a number of individuals in the same occupation voluntarily organized to earn a living by openly serving a certain moral ideal in a morally-permissible way beyond what law, market, and morality would otherwise require" (Davis: 2002, 3). Following this definition, a professional is committed to a set of standards that belong to that particular profession. This definition implies that non-professionals doing the same work (that is, carrying the same occupation but not being part of the profession) are subject to lower moral standards than professionals. Contrary to the description of professional responsibility provided by Davis (2002), I leave it open what these higher moral standards exactly amount to. Since I expect professionals themselves not to distinguish between moral responsibility and professional responsibility, I take the term professional responsibility to be an aspect of the broader notion of moral responsibility, referring to moral obligations relevant for *professionals* carrying on their occupation. These profession-based moral obligations exist in addition to one's "general" moral obligations. This means that one's moral responsibility may also be dependent on the question whether or not one acts as a member of a profession. In other words, moral responsibility includes professional responsibility but is not limited to it.

political theory can be used to resolve issues in a situation of pluralism (in this context: a pluralism of responsibility conceptions). In Section 1.3, I discuss my research approach, followed by a detailed outline of this thesis in Section 1.4.

1.1.   Moral responsibility in engineering

Various authors discuss the topic of responsibility in engineering. Harris et al. (2005[1995]) dedicate a full chapter to responsibility in the third edition of their classic textbook on engineering ethics. The authors make a distinction between a positive and forward-looking conception of responsibility on the one hand and a negative and backward-looking conception of responsibility on the other. They refer to the former as obligation-responsibility and to the latter as blame-responsibility.[5]

    In their discussion of obligation-responsibility, Harris et al. take professional standards as the starting point for developing a notion of obligation-responsibility. They argue that a professional's responsibility goes further than simply following professional standards. They illustrate this on the basis of irresponsible action. Based on the notion of "reasonable care," they try to answer the question what a responsible engineer or technician should have done when an accident occurs. Even the people that are not "at fault" (that is, those people that actually follow standard operating procedures) may have a role to play in improving these procedures and enabling others to conform to them. This resembles the notion of professional responsibility developed by Michael Davis. He argues that engineers are primarily responsible for acting in accordance with their profession's code of ethics (and other technical standards). Although Davis does not see any special "obligations" or "duties" other than the ones specified by the code, he admits that the engineers' professional responsibility goes further than the acts specified by the code of ethics, in the sense of a requirement to support the code by "encouraging others to do as the code requires and by criticizing, ostracizing, or otherwise calling to account those who do not" (Davis: 1998, 59; see also Davis: 2002, 93-94 on interpretative obedience of codes of ethics).

---

[5]   Harris et al. use backward-looking responsibility and blame-oriented responsibility synonymously. However, they are not synonymous. Backward-looking responsibility does not necessarily have to coincide with blame. See Chapter 2 for a more elaborated taxonomy of different responsibility perspectives in engineering.

Harris et al. discuss blame-responsibility on the basis of an analogy between legal and moral liability. The authors illustrate how people can be held morally responsible when certain conditions are met. One of these conditions, the authors argue, is the obligation to act in accordance with the professional standards. These standards prescribe how a responsible professional should act. If an engineer fails to act in accordance to those standards (for example, when she fails to exercise due care), she can be a held responsible for the harm done. Hence, this condition ties blame-responsibility to obligation-responsibility.

So far, the discussion of the moral responsibility is in terms of the responsibility of individual engineers. However, engineering is most often a collective enterprise with different actors working in organizations or, in case of R&D, networks of organizations. How does this collective dimension of engineering affect the responsibility of individuals? It is in relation to this collective dimension of engineering that the problem of many hands is introduced in the engineering ethics literature.

The term "problem of many hands" was introduced by Dennis Thompson (1980) and he described it as the difficulty of identifying the person responsible for some outcome if a large number of people are involved in an activity. This description is rather broad and it leaves room for many different interpretations. In a recent paper, Michael Davis (forthcoming) points out that Thompson's formulation refers in fact to an epistemological problem and not a metaphysical one. Davis distinguishes between what he calls the problem of *many hands* and the problem of *many causes*. The first is the epistemic problem to identify the person responsible for harm because one does not know who actually made what contribution. This is mainly a problem for outsiders, Davis argues, because insiders generally know very well who made what contribution.

Other authors come with a metaphysical interpretation of the problem of many hands. Michael Davis prefers to call this the problem of many causes rather than the problem of many hands. The interpretation given by Harris et al., for example, is an instance of this problem of many causes.[6] They treat it primarily as an excuse, voiced by individuals, for evading personal responsibility for wrongdoing by pointing out that many individuals were involved in causing

---

[6] Although Davis is right in pointing out the categorical difference between both interpretations, I follow the common literature and refer to this metaphysical interpretation also as the problem of many hands.

the harm (Harris et al.: 2005[1995], 35).[7] Although Harris et al. do not discuss the arguments extensively, they seem to be somewhat skeptical about the use the problem of many hands as an excuse.

Whereas Harris et al. treat the problem mainly as an excuse, Helen Nissenbaum (1994; 1996) and Mark Bovens (1998) emphasize the conceptual dimension of this metaphysical interpretation of the problem of many hands. Our conditions for individual responsibility do not easily generalize to collective action so we need a different conception of responsibility, both authors argue. In the philosophy literature, this track is also followed by philosophers such as Peter French (1984; 1991) and Larry May (1992), who try to develop special principles that hold in situations where many actors are involved.

It is remarkable that all these interpretations of the problem of many hands refer to backward-looking or blame-responsibility. However, the fact that there are many actors (many hands) involved poses problems for ascribing forward-looking responsibility as well. Whereas the "backward-looking" interpretations of the problem of many hands are mainly about the question of identifying the "wrongdoer," a forward-looking interpretation could be the problem of agreeing on the question who is responsible for doing something.[8] In this thesis, I focus on forward-looking responsibilities and the corresponding forward-looking interpretation of the problem of many hands. The reason for this is that this forward-looking perspective is more in line with the prevailing attitude of professional engineers themselves. Davis (forthcoming) argues that professional engineers are keen on reporting errors in order to avoid mistakes that are made before (forward-looking), but that they are less interested in discussing blame (backward-looking). Although this does not exclude a focus on blame (for lawyers and legal scholars, for example, the question of blame and liability is an

---

[7] According to Michael Davis (forthcoming), ascribing responsibility in these situations is not problematic at all because the fact that many people causally contribute to some outcome does not exempt any actor from responsibility for her particular contribution to the harm.

[8] This interpretation is the most straightforward translation of the backward-looking interpretations of the problem of many hands into a forward-looking one. However, if we take the problem of many hands to refer to any problem that can be encountered when there are many people involved, the institutional setting of engineering work deserves attention as well. The organizational constraints that individuals encounter when they work in a collective setting are a notorious obstacle for people's ability to perform particular tasks. For an analysis of how organizational structure enables actors to discharge their responsibilities, the reader is referred to Doorn et al. (2011).

important one), my aim is to *improve* engineering in the sense of preventing as much as possible its negative outcomes. In order to do so, I try to stay as close as possible to engineering practice itself. I therefore use the following working definition of the problem of many hands: How to agree on the question of who is responsible for addressing the potentially negative implications of technology? This definition indicates both a forward-looking perspective and a focus on the *consequences* of the technology. That means that the "workplace ethics" (for example, the question whether or not the researchers drink organically produced coffee and tea) falls beyond the scope of my research.

Since there are many ways to conceive of responsibility, this question "Who is responsible?" is not a trivial one. Especially in situations where many people are involved, and R&D is such a context, people may answer this question differently. One of the reasons for these differences in answer is the fact that people have different *conceptions* of responsibility. These different conceptions reflect philosophical differences that go deeper than simply definitional misunderstandings and they can therefore not be solved by simply clarifying the terms (Thompson and Dean: 1996). To understand this, we have to distinguish between "concepts" and "conceptions," a distinction introduced by the political philosopher John Rawls in order to clarify philosophical disputes over the meaning of "justice" (Rawls: 1999 [1971]; see also Hart: 1961). Rawls explains that people may have different views on how to shape social cooperation; that is, they may defend a different set of principles that indicate what they find just. However, although people may disagree about exactly which principles should define the basic terms of cooperation, they "understand the need for, and are prepared to affirm, a basic set of characteristic principles for assigning basic rights and duties" (Rawls: 1999 [1971], 5).[9] This basic set of characteristic principles (the concept of justice) is the central element which all these different sets of principles (that is, the different conceptions of justice) have in common. This basic set of characteristic principles, for example, encompasses the idea that "like cases should be treated alike and different cases differently;" it leaves room, though, for different interpretations of what differences in cases are to be

---

[9] Rawls refers to the legal philosopher H.L.A. Hart who defends a similar distinction between what he calls a "leading precept" of justice (cf. the concept) and the relevant criteria of justice which "may often vary with the fundamental moral outlook of a given person or society" (cf. the conceptions; Hart: 1961, 155-159). I could not find any earlier text than the one by Rawls in which the terminology "concept" and "conceptions" was used in this particular way.

regarded relevant and what differences are not. Similarly, people may use stipulative definitions to indicate when the term responsibility applies and how it connects to other theoretical terms (such as duty or obligation), but these definitional clarifications cannot solve all disagreements on the question "Who is responsible" because the different conceptions of responsibility reflect different criteria when it is appropriate or fair to ascribe responsibility – let me call these the different "fairness requirements." These different fairness requirements, in turn, can be traced back to the different reasons people have for ascribing responsibility. One could think of the development of a new risky technology, where people need to distribute the moral responsibilities to prevent harm or to look after certain risks. For the general public, we want to avoid harm as much as possible and for this reason we want all risks to be looked after. Hence, if one has a conception of moral responsibility that is based on the no harm principle, for example, it may seem functional to make anyone in the innovation chain responsible for preventing harm. However, some researchers may judge it unfair that they have to look after (all) the risks, also highly unlikely ones and maybe risks they are themselves not familiar with. They may say that it is preferable to let a more specialized person deal with the risks. From a consequentialist perspective, this may be more effective.

This example shows that only two different conceptions may already lead to conflicting requirements. In practice, one could think of far more responsibility conceptions with consequently many different fairness requirements. The fact that these conceptions may sometimes conflict is not just a practical problem but also a moral problem: both conceptions are based on particular moral background theories, or – as Rawls calls it – "comprehensive doctrines." The term "comprehensive doctrine" is typically used to refer to a religious, philosophical, or other standard moral doctrine that "applies to all subjects and covers all values" (Rawls: 2001, 14). These doctrines have a particular "conception of what is of value in human life or, alternatively, of what is regarded as a fully worthwhile life" (Rawls: 2001, 19).[10]

In this thesis, I am not committed to any comprehensive moral theory. I defend the claim that all comprehensive doctrines that can be argued for on the basis of "reasonable arguments" are legitimate. Consequently, the corresponding

---

[10] Comprehensive doctrines contrast with political conceptions that are not related to any one particular comprehensive doctrines but that are compatible with (one or more) of these comprehensive doctrines (Rawls: 2001, 19).

responsibility conceptions are also – in principle – legitimate. Although one could think of conceptions of responsibility that do not one-to-one map to existing comprehensive doctrines, any conception of moral responsibility reflects some view on morality and of what is of value in human life and is therefore – if defended in terms of "reasonable arguments" – legitimate as well. Of course, this leaves me with the burden of providing a sound interpretation of what this reasonableness amounts to, but let me postpone that for a moment. If there are different conceptions of moral responsibility which are based on different comprehensive doctrines and if these different comprehensive doctrines are legitimate (a situation which we may refer to as "responsibility pluralism"), we need a framework for deciding on responsibility issues that is impartial towards any of the responsibility conceptions.

To recapitulate, the aim of the research presented in this thesis is to alleviate the problem of many hands in R&D networks, in particular as it stems from conflicting conceptions of responsibility. In the next section, I will argue how procedural political theories can be used to provide such a framework. In procedural theories, the focus is shifted from substantive conceptions of justice (or here: responsibility), which would unavoidably favor one comprehensive theory over another, to the justness or fairness of the decision making procedures, which allows for an impartial view of the different conceptions of justice (or responsibility).

The research presented in this thesis is original in its use of procedural political theory to assess the fairness of responsibility ascriptions. Contrary to, for example, medical ethics or business ethics, engineering ethics does not have a tradition of using concepts from procedural political theory. I will therefore use the next section to elaborate on the use of procedural political theory in engineering ethics and justify my choice for one theory in particular, namely Rawls' political liberalism.

1.2. Procedural justice

As explained in the previous section, the starting point for this research is the assumption that the problem of many hands can be conceived as a problem of conflicting responsibility conceptions. As long as we depart from one particular conception of responsibility, it is probably difficult to make the ensuing responsibility ascription fair in the eyes of adherents of other responsibility conceptions. Neither is it an option to leave fairness aside and to let everyone

pick only those particular (moral) tasks that he or she thinks is the appropriate responsibility to do. This may easily lead to gaps in the responsibility distribution because people may expect someone else to assume the remaining responsibilities or it may invite strategic, and unpredictable, behavior. This resembles the problem in political theory of accounting for different conceptions of justice.

In political theory, it is notoriously difficult to do justice to people's different conceptions of the good life, including visions of what justice amounts to. As long as we try to develop an idea of justice that starts from one particular comprehensive theory or moral framework, it will be very difficult to make it acceptable to all people. Some may argue that this is not problematic since "acceptance" is not part of justice. If one thinks that one's own comprehensive moral framework presents the "true" view on justice, acceptance is indeed not a matter of concern. However, history has shown that such views easily lead to totalitarian, or at least undemocratic, societies. The recognition that there is not one superior theory and that different comprehensive doctrines are all *legitimate* (although sometimes with qualifications) is seen as the cornerstone of democracy (Dryzek and Niemeyer: 2006). We could call this the ideal of pluralism, where pluralism is understood as the acknowledgment of diverse and competing values and visions of the good life. Some political philosophers have therefore introduced *procedural* approaches to justice (or more general: procedural theory) to avoid favoring some conceptions of justice over others. Rather than developing substantive views on justice, which are unavoidably biased towards some comprehensive doctrine, procedural theorists develop formal ideas on what procedures lead to fair outcomes. In its most general form, the main idea of procedural justice or fairness[11] is that the outcomes of those procedures that are supported by adherents of any of the comprehensive doctrines can be deemed fair. This often amounts to a procedure that affords those who are affected by a certain decision the opportunity to participate in the decision making.

---

[11] There does not seem to be one agreed upon distinction between procedural justice and procedural fairness. In the context of institutional arrangements, the term procedural justice is the more common term. However, some philosophers refer to fairness as the attribute of a certain outcome and justice as the attribute of the process of arriving at this outcome. Putting aside these quibbles, in this thesis, the terms procedural justice and procedural fairness can be read interchangeably.

The analogy between my research topic (responsibility) and political principles of justice is that the different responsibility conceptions are also related to different comprehensive doctrines (the responsibility pluralism, I referred to). Similar to political pluralism, the plurality of conceptions of moral responsibility reflects different visions of a just society. These different conceptions cannot be reduced to one overarching view on moral responsibility on the basis of which we can distribute the responsibilities among different actors in a research team and to assess the moral fairness of the distribution. Secondly, an important consideration when ascribing responsibilities is indeed the fairness of the ascription. Whereas other, equally important, moral problems could maybe be framed in more private terms (for example, the effect of human enhancement on one's private identity), ascribing and distributing responsibilities is pre-eminently a political issue because it is about actions and things we bring about "through our associations with other people" (Kutz: 2000, 1), which may have negative consequences for others.[12] Additionally, being responsible for something goes hand-in-hand with having power, which may in turn lead to a situation of conflicting stakes (Rueschemeyer: 1986). As such, the question of how to distribute responsibilities seems to resemble the political arena and, consequently, be related to issues of fairness and justice.

The idea followed in this thesis is therefore to use procedural theory as an alternative to substantive responsibility conceptions to assess the fairness of the responsibility ascriptions. In the remainder of this section, I discuss two highly developed and differentiated procedural political theories: deliberative democracy and political liberalism. Both attempt to propose the formal conditions under which decision making can be deemed fair. I do not want to suggest that these two theories cover the whole domain of procedural political theory. However, in terms of practical applications, these two theories are the most prominent ones. In what follows, I discuss these political theories,

---

[12] As explained by Strawson, this interpersonal dimension is also tacitly present in the debate on individual responsibility. In his famous essay *Freedom and Resentment*, Strawson (1974) distinguishes between a so-called "attitude of involvement" and an "objective attitude." Ascribing responsibility to a person involves an attitude of involvement, which reflects feelings of gratitude, forgiveness, affection, esteem, on the one hand, or resentment, contempt, malevolence on the other (ibid., 5). Hence, even in the private sphere, responsibility is primarily about interpersonal relations, according to this view.

including their application to real-world examples of decision making, and I explain my choice for Rawls' political liberalism.

1.2.1.   Deliberative Democracy

The term deliberative – or discursive – democracy was originally coined by Bessette (1980). After that, it was further developed by political theorists such as Elster (1986; 1998), Cohen (1989; 1997), Gutman and Thompson (1996; 2004), and Fishkin (1991; Fishkin and Laslett: 2003). The concept of deliberation can also be linked to the work of the German philosopher Jürgen Habermas (1990a). According to deliberative democracy theorists, the legitimacy of political arrangements has to be interpreted in terms of the right of those subject to collective decisions to participate in a political practice of argumentation and reason-giving among free and equal citizens. In this practice, all people deliberate according to some form of reason-based agreement (Forst: 2001). Deliberative democracy hinges on two ideas. First, the deliberation is an inclusive and non-coercive process that takes place between free and equal participants (i.e., no reference to power, authority or status). Second, these participants commit themselves to established rules of rational discourse, which includes the commitment to provide reasons for one's assertions (Elster: 1998). The "reasonableness" is crucial: since people deliberate in accordance with agreed upon reasons, the outcome (or solutions) can be deemed reasonable as well. Moreover, the deliberation produces a shared understanding of the problem at hand and potentially new – and innovative – solutions (Renn: 1999).

The "rules of reason" can be stated as follows (based on Alexy: 1990[1978]; Habermas: 1990b):

(1)   Every subject with the competence to speak and act is allowed to take part in a discourse.[13]

---

[13] The phrasing "competence to act" seems somewhat odd. In the original text, Alexy uses the phrasing "anyone who can speak." Habermas explicitly refers to the text by Alexy but he does not mention a reason for introducing the clause "competence to act." However, his writing does not suggest that he wants to exclude people with physical disabilities.

(2)   Everyone is allowed to

    a)   question any assertion whatsoever;

    b)   introduce any assertion whatsoever into the discourse;

    c)   express his attitudes, desires, and needs.

(3)   No speaker may be prevented, by internal or external coercion, from exercising his rights as laid down in (1) and (2).

Although some deliberative democracy theorists consider deliberation primarily as a prelude to voting and majority rules (e.g., Manin: 1987), others endorse the full consequences of deliberative theory and defend consensus, as long as it is established in an informed, un-coerced and reflective way, as the only legitimate alternative to voting and majority rule (e.g., Elster: 1986). The latter view is exemplified in Cohen's famous statement that "outcomes are democratically legitimate if and only if they could be the object of free and reasoned agreement among equals" (Cohen: 1989, 22). For Cohen, "ideal deliberation" leads to the following body of rights (or criteria):

(1)   It is a *free* discourse in two ways:

    a)   The participants consider themselves bound solely by the results and preconditions of the deliberation. They are free from any authority of prior norms or requirements.

    b)   The participants suppose that they can act on the decision made; the deliberative process is a sufficient reason to comply with the decision reached.

(2)   It is a *reasoned* discourse: Parties to deliberation are required to state reasons for their proposals, and proposals are accepted or rejected based on the reasons given, as the content of the very deliberation taking place.

(3)   Participants are *equal* in two ways:

    a)   Formal: anyone can put forth proposals, criticize, and support measures. There is no substantive hierarchy.

    b)   Substantive: The participants are not limited or bound by certain distributions of power, resources, or pre-existing norms.

(4)   Deliberation aims at a rationally-motivated *consensus*: It aims to find reasons acceptable to all who are committed to such a system of decision making. When consensus or something near enough is not possible, majoritarian decision making is used.

The foregoing discussion shows why deliberative democracy, at least the view defended by Elster and Cohen, can be deemed a "procedural" theory of justice. With its focus on communication rules, the deliberative democracy theorists do not give substantive criteria for what justice amounts to (apart from criterion 3b) but prescribe the procedural rules that legitimize or justify the outcome of the process. In lawmaking, this ideal is exemplified by Bohman as the "procedural ideal [that] ... the legitimacy of laws depends on the democratic character of the legislative process that makes possible a consensus of all citizens" (Bohman: 1998, 402).

Critics of deliberative democracy argue that the deliberative approach is not as democratic as proponents claim it is. Blattberg (2003) argues that deliberative democracy is biased in favor of liberalism because the condition of equal liberty of individuals is itself an affirmation of liberalism (160). He doubts whether equality, in any form, really is a requirement for genuine conversation. Whereas deliberative democracy theorists only allow the "force of the better argument" (Cohen and Rogers: 2003), Blattberg argues that this force can never be equally available to all for the simple reason that some people are smarter than others. To claim that everyone is equal in all respects is therefore neither realistic nor necessary.

Interestingly, one of Blattberg's other objections points in a different direction. He questions the sharp division between just and rational deliberation on the one hand and self-interested and coercive bargaining or negotiation on the other (162-163). Since deliberative theory presupposes a distinction between "valid" and "invalid" reasons – appeals to particular interests being an example of the latter – people are forced to make disengaged and generalized claims. According to Blattberg and political theorists such as Young (1996; 2000) and Mouffe (1999; 2000), this is potentially oppressive to minority groups. Since the goal of deliberation is consensus, the common good is all too easily equated with the interests of the powerful. Thus, under the sway of deliberation, the appeals to unity and the common good may sideline the legitimized concerns of the most marginalized (Dryzek and Niemeyer: 2006). The underlying critique is that the deliberative procedure seems to suggest that the distinction between reasonable

and unreasonable is an "objective" distinction, which does not require any substantive judgments. However, it is questionable whether this strict procedural approach can do without any substantive claims, so the opponents argue. To decide what counts as a legitimate reason introduces already some substantive content to the theory. Without any substantive content, the theory either fails to have normative force or suffers from an infinite regress in its determination of the "reasonable." The deliberative approach is therefore either implicitly substantive or it misses normative force.

Notwithstanding this critique, the ideal of deliberative democracy has found its way in applied contexts. It has been proposed in the context of technology management to decide on issues related to, for example, risk management (Renn: 1999; 2006; 2008), scientific and technological policy making (Laird: 1993), medical and participatory technology assessment (Einsiedel: 2002; Milewa: 2006), and stakeholder deliberation (Van de Kerkhof: 2006). In these cases, the framework is mostly used to distinguish valid from invalid reasons (point 2 in Cohen's body of rights) and not so much to compliance to the results and preconditions of the deliberation (points 1a and 1b). In most cases, the framework is also used to encourage learning among lay people. Additionally, the main objective is often policy influence or the discovery of values rather than full consensus (point 4). Although the formal equality of the participants is emphasized in all cases (point 3a), most analysts conclude that the knowledge divide makes it difficult to adhere to substantive equality (point 3b).

In the context of the democratization of science, this last point is very relevant indeed. Scientific expertise unavoidably gives some actors power over other actors. Blattberg seems to suggest that this asymmetry in power is not detrimental for genuine conversation. The only symmetry that is required according to Blattberg, arises from a "willingness to participate, [...] a tactful speaking and earnest listening" (Blattberg: 2003, 161). It is indeed questionable whether we should adhere to the ideal of full symmetry with respect to scientific expertise and knowledge. Although some authors argue for a democratization of the decision making involving science and technology (and rightly so, I would say. See, e.g., Sclove: 1995), it is something else to state that all assertions concerning the particular technology or science itself are of equal worth. In the context of highly specialized technological and scientific knowledge, this position is hard to defend. It *does* matter whether arguments are based on scientific research or not. However, the important point remains that scientific knowledge *alone* cannot provide definite answers to questions concerning the acceptability

of technology and its consequences. Questions related to professional responsibility fall into this latter category (for instance the question whether or not it is the researchers' responsibility to minimize potential risks). Hence, although the deliberative ideal of equality needs some qualification (in the sense that not all assertions about science itself are of equal worth), it seems valid in the context of responsibility distributions and it is therefore something worth striving for.[14] This suggests that the approach should be applied with due care for this criterion of equality.

1.2.2. Political liberalism

The second procedural theory I would like to discuss is Rawls' political liberalism and the concepts of overlapping consensus and Wide Reflective Equilibrium (WRE).[15] Rawls introduced the notion of overlapping consensus to reconcile the ideal of pluralism with that of consensus. Rawls' aim was to develop a criterion of justice that would be agreed upon by all under conditions that are fair to all (JaF 15). Rawls developed a justification criterion to assess the fairness of the basic institutions of society. Individuals must be able to accept the fair terms of cooperation if it can be embedded in their individual comprehensive doctrine (RH 143); that is, if it coheres with their own firm convictions of justice or extends them in an acceptable way (ToJ 17, PL 28). Although Rawls at first wanted to develop a substantive theory of justice for a relatively homogeneous well-ordered society, in his later work he revised this idea of a well-ordered society. Recognizing the permanent plurality of incompatible and irreconcilable moral frameworks within a democratic society, he introduced the concept of overlapping consensus. People are able to live together despite conflicting moral values and ideals as long as they share a moral commitment to the society's basic structure.

---

[14] Note that in R&D networks, equality also seems a *realistic* presupposition. Contrary to hierarchical organizations, networks can indeed be characterized by more or less symmetrical relations between the actors constituting the network.

[15] In the remainder of the text, the references to Rawls' work will be denoted by the acronyms ToJ, PL, JaF, CP, RH and PR referring to *A Theory of Justice* (Rawls: 1999 [1971], 42-43), *Political Liberalism* (Rawls: 1993), *Justice as Fairness* (Rawls: 2001), *Collected Papers* (Rawls: 1999), and the papers 'Political Liberalism: Reply to Habermas' (Rawls: 1995) and 'The Idea of Public Reason revisited' (Rawls: 1997) respectively.

People with different comprehensive doctrines must be able to justify for themselves the acceptability of the claims of political justice (RH 143, ToJ 17, PL 28). Rawls introduced the idea of reflective equilibrium to refer to this individual justification. In this idea, a distinction is made between three levels of considerations: (1) considered moral judgments about particular cases or situations, (2) moral principles, and (3) descriptive and normative background theories. Assuming that all people want to arrive at a conception of justice that yields definite solutions and that is complete, in the sense that it is more than a mere collection of accidental convictions, people should aim at 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.

We speak of an *equilibrium* if the different layers cohere and are mutually supportive; it is called *reflective* if the equilibrium is arrived at by working back and forth between the different considerations and if all 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 (three-tiered view) and not only the considered judgments and moral principles (which is the case if we speak of a *narrow* reflective equilibrium; two-tiered view). In his later work, Rawls adopted this wide conception of reflection since this allows for taking into account moral conceptions advanced by others, thereby giving these the chance to influence one's own convictions (CP 289-290).

Rawls shares with the proponents of deliberative democracy and other liberalist theories of justice a main role for "public reason." However, in their interpretation of public reason Rawls and the deliberative democracy theorists (and Habermas in particular) depart. Habermas defends a conception of public reason which includes all unofficial arenas of public discourse; these unofficial arenas in fact ground democratic self-government and political autonomy (McCarthy: 1994, 49). Rawls has a more restricted notion of public reason. For him, public reason is limited to the official institutions. Since his procedural approach to justice aims at "uncovering a public basis of justification on questions of political justice given the fact of reasonable pluralism," it should proceed from "what is, or can be, held in common; and so [...] begin from shared fundamental ideas implicit in the public political culture in the hope of developing from them a political conception that can gain free and reasoned agreement in judgment" (PL 100-101). Hence, the function of public reason is not so much to be critical but rather to be constructive. Public reason, therefore,

needs to start from shared ideas and organize those into a political conception that can serve as the focus of an overlapping consensus, which in turn can enhance stability. Rawls connects his conception of reasonableness to T.M. Scanlon's principle of moral motivation, which is one of the basic principles of contractualism (Scanlon: 1982, 104, 115). The principle tells us that we have a "basic desire to be able to justify our actions to others on grounds they could not reasonably reject" (PL 49-50; fn 2).

Similar to deliberative theory, Rawls' concepts have also found their way to more applied contexts, although mainly as suggestions for solving particular problems rather than as fully elaborated procedures. Especially in the context of applied ethics, the tension between diverging moral frameworks is an urgent problem; for example, how to integrate the interests of multiple stakeholders? (Daboub and Calton: 2002; Jamali: 2008); or how to justly distribute risks stemming from new technologies? (Zandvoort: 2008). Rawlsian approaches seem promising for answering these kinds of questions since they offer a methodological alternative to the extreme positions of ethical theorism (or generalism) and (specified) principlism on the one hand and particularism (or casuism) on the other (Daniels: 1996; St. John: 2007; Van den Hoven: 1997).[16] Rawlsian justification avoids the drawbacks of both extremes because it aims at coherence between the abstract theoretical principles and the more particular considered judgments without giving priority to any of them. As such, Rawlsian approaches seem to offer a promising decision making procedure within applied ethics.[17]

---

[16] Proponents of the first category argue that applied ethics is essentially the application of general moral principles (Beauchamp and Childress: 1994; DeGrazia: 1992; Lustig: 1992) or theories (Gert et al.: 1997; Hare: 1988) to particular situations. This position is criticized for mistakenly assuming that valid principles can be formulated that govern all rational persons. Moreover, the critics argue, procedures for deducing answers to moral questions is impossible, unnecessary, and undesirable. These critics argue for situational adequacy, that is, an ideal of doing justice to persons in a particular historical context. The problem with particularism, on the other hand, is that it runs the risk of lacking moral justification. In most situations where ethical reflection is at stake, people should be able to justify their actions in terms of moral principles. However, if particularism is carried through to the extreme, it becomes difficult to provide public justification of moral judgments (Van den Hoven: 1997, 240-241).

[17] In addition to this justificatory application, Rawlsian approaches are sometimes used in a constructive way as well. In the latter case, they are used as a framework for structuring discussion and debate, with the aim of coming to a justified agreement. The method could then be used, for example, as a *means* to attain a coherent basis for decision making in ethical

Dependent on the field of application, Rawls' approach to justification has the advantage over deliberative approaches that it gives a more structured framework for evaluation. Although the deliberative rules are useful to provide the "rules of the game," the justification itself seems too thin to provide help in justifying a particular outcome. This strictly procedural approach cannot explain the outcome to people not taking part in the deliberative process other than saying that the decision procedure was fair.[18] Rawls' approach seems to provide a more elaborated justificatory framework. Especially the concept of WRE seems an attractive method for real-life justification. Even without reference to political concepts as overlapping consensus, the notion of WRE can help explain why people consider certain things fair or unfair. Moreover, since the Rawlsian approach takes the different layers of morality explicitly into account, the method seems to provide a better method to encourage reflection compared to the Habermassian approach. The criticism that the Rawlsian procedural approach to justice requires that people share the tradition of liberalism does not seem valid. It is sufficient that people acknowledge that reasonable pluralism is the permanent condition and that the concept of reasonableness replaces that of moral truth. This is not the same as sharing the comprehensive view of liberalism. In a professional setting where people are motivated to work towards a fair distribution of responsibilities, this demand of "reasonableness" is probably a realistic one. McCarthy (1994) argues that it is a strength of Rawls' theory that he allows for different levels of abstraction. The more difficult it becomes to agree on general interests and shared values, the higher the level of abstraction of the overlapping consensus. However, it could also work the other way around. In case of responsibility distributions, people can disagree on the abstract levels of responsibility conceptions and principles, but agree on particular responsibility ascriptions.

---

committees or to gain support for particular decisions in the context of public policy (Holmgren: 1987). This second way of applying Rawlsian approaches is comparable to the constructive application of deliberative approaches, albeit the Rawlsian ones take the moral background theories and principles more explicitly into account.

[18] In that sense, the criticism phrased by Mouffe is not a harmful one. The deliberative process is too little rather than too much focused on some outcome.

For the reasons mentioned above, I have chosen to use Rawls' procedural approach for further elaboration in the context of responsibility distributions and their fairness.[19]

1.3.   Research approach

1.3.1.   Research objective and research questions

The topic of the research presented in this thesis is moral responsibility in R&D networks. Because of the involvement of many actors in R&D and the long causal chain from research to the actual market introduction, R&D is vulnerable to the problem of many hands. The research objective is to develop a procedural approach based on Rawls' idea of Wide Reflective Equilibrium (WRE) for resolving responsibility issues in engineering in a situation of pluralism. The research question is how a procedural approach, based on Rawls' idea of WRE, can be (further) developed and used to resolve the tension between competing conceptions of responsibility, which lies at the core of the problem of many hands.

This question is answered in a series of steps. As stated, the main idea underlying this research is that the problem of many hands can be partly traced back to different requirements for a desirable distribution of responsibilities that result from different responsibility conceptions. These different requirements are sometimes conflicting, which makes it difficult to come with one distribution of responsibilities that is acceptable to all people involved. It should therefore first be investigated what these different responsibility conceptions are and how they relate to particular moral background theories (e.g., deontology, consequentialism, and rights-based ethics).

The first subquestion is therefore:

(Q1)   What are the different responsibility conceptions and what are the accompanying criteria or requirements for the responsibility ascriptions?

---

[19]   I do not want to suggest that we need to reject the whole deliberative approach. The "conversation rules" exemplified by Habermas can be useful guidelines to keep in mind in the process of decision making. However, for justificatory purposes the method seems too thin.

In order to answer this question, I analyze which different responsibility conceptions are distinguished in the philosophical literature and how the different conceptions for ascribing responsibility are grounded on particular moral background theories. If there is a relation between the responsibility conceptions and moral background theories, we can say that there is a pluralism of legitimate responsibility notions which cannot be reduced to one overarching view on responsibility.

The next step is to look for and further develop a procedure or method that does justice to this pluralism of responsibility conceptions. In this thesis, I explore how Rawls' Political Liberalism can be used to develop such a procedure.

The second subquestion is therefore:

(Q2) How can Rawls' procedural approach be applied to resolve the tension between competing responsibility conceptions?

Since Rawls' procedural theory has been applied or at least suggested before, I make an inventory of different applications of Rawls' approach in the field of applied ethics in order to identify the main obstacles and lessons learnt. The results of this inventory feed into the Rawlsian approach for distributing responsibilities that I further develop in this thesis.

The aim of this research is to contribute to a solution of the problem of many hands. This aim involves both a constructive and a justificatory element. The constructive element relates to the question to what extent the method developed in this research does actually contribute to a solution of the problem of many hands. This is an empirical question. By developing a method based on Rawls' procedural approach, the judgments and opinions of the relevant stakeholders are indispensable for assessing the fairness of the method's outcomes because the justificatory force (that is, the fairness) is derived from the coherence between these judgments and opinions. Hence, both the constructive and the justificatory aim indicate that empirical data are needed for testing the method.

For this justificatory element, we need to know real people's considerations when distributing responsibilities in an R&D network. These empirical data can then be analyzed in terms of the Rawlsian justificatory framework, which may eventually provide an answer to the question whether the approach is indeed constructive for resolving the tension between the competing responsibility conceptions. This yields the following empirical subquestions:

(Q3)   What are the rationales for distributing responsibilities as displayed by engineers and other technological researchers?

(Q4)   Does a procedural approach to fairness resolve the tension between competing conceptions of responsibility?

For answering these two empirical questions, I decided to focus on one specific R&D project rather than doing several comparative studies. The motivation for this methodological choice is that it allows me to study the data more in-depth, which I find more informative at this explorative stage. After all, Rawls' political theory has not been used before in the context of responsibility distributions and this research can therefore not build upon previous comparable empirical studies. Realizing that this methodological choice allows me to answer subquestions Q3 and Q4 only as far as they concern this specific project, I will pay attention to the generalizability of the empirical outcomes in the concluding chapter.

The specific project that was selected for studying is the ALwEN project, an technological project aimed at the development of an in-house monitoring system for elderly people. This project was selected because it is a prototypical R&D project, containing the whole R&D chain. More details of the ALwEN project are provided in Section 1.3.2.

Although not the final chapter in this thesis, one additional subquestion is taken up as an epilogue. This question reads as follows: What role could engineering ethicists play in distributing responsibilities during R&D? Since this research is an instance of empirical ethics, the research contains both social scientific and conceptual elements. Given this multidisciplinary character, the question arises how the empirical data support the normative conclusions (that is, the justification) and what the role of the ethicist is in this justificatory process. This question is especially relevant because the procedural justification seems primarily a first-person judgment and the role of the ethicist is therefore not obvious.

The thesis is organized as follows. Following this introductory chapter, Chapters 2-6 are dedicated to the five research subquestions respectively. In Chapter 7, I return to the research objective and questions and reflect on the research methodology. Conclusions concerning the outcomes are formulated, as well as recommendations for the role of engineering ethicists in R&D practice and points for further research.

1.3.2.   Research methodology

As the previous section indicates, the empirical data in my research is indispensible for my research, which is therefore a form of empirical ethics. Since this term is not unambiguously defined, I use this section to explain my interpretation of empirical ethics. The same holds for the idea of Rawls' specific type of coherentist justification, which leaves room for different interpretations. I therefore elaborate a bit more on my particular approach. But before doing so, I first discuss the methodology for retrieving the empirical data.

Case study methodology
The methodology followed to retrieve the empirical data is case study research. The empirical data used in this research mainly concern answers to so-called "how" and "why" questions: How do researchers distribute responsibilities? Why do they judge some responsibility ascriptions as fair and others not? Case studies are particularly useful for answering these kinds of questions (Yin: 2003, 5-6). The case that plays a central part in this thesis is the ALwEN project, where ALwEN is an acronym for Ambient Living with Embedded Networks. This project concerns the development of an in-house monitoring system based on Ambient Intelligence technology.[20] (In the remainder of this thesis, I will refer to the technological research project itself as the ALwEN project, such as to distinguish it from the ethical research.) The empirical part of this thesis – most notably Chapter 4 and 5 – draws heavily on the material collected through studying this case.

The ALwEN project was selected because it matched the following criteria:

- the project covered the whole trajectory from fundamental research to application;
- the project team resembled a network structure: Several partners were involved and their mutual relationships were non-hierarchical;
- the project members were willing to allow for the involvement of an outsider;

---

[20] Ambient Intelligence reflects a vision of the future of ICT in which intelligence is embedded in virtually everything around us, such as clothes, furniture, etc. The technology consists of Wireless Sensor Networks (WSN), the combination of body sensors, ambient sensors and wireless networks.

- the timing of the technological project: The ALwEN project started in December 2007 and is expected to be finished by the end of 2011. The case study research period lasted from September 2008 to August 2010.

Using case studies as a research strategy requires due care with respect to the data collection and interpretation in order to prevent problems establishing the construct validity and reliability of the case study evidence, where validity refers to the question to what extent the operationalization of the concept actually measure what it says it does and reliability to the consistency of the data collection (that is, does the measurement lead to the same outcomes when done at different times and by different people?). The most important principle to follow is *triangulation*; that is, the principle that the advocacy of the argument should come from multiple sources of evidence (Yin: 2003, 97-99). Since each source has its strengths and weaknesses, combining the findings from several sources helps building evidence. In the case study presented in this thesis, the following data sources are included:

- direct observation at project meetings;
- project reports and meeting minutes;
- interviews with researchers;
- workshop;
- external documents (official policy documents at the national and European level, scientific articles).

The interviews and personal workshop data were approved by the participants. Regarding the interpretation of the data, the relevant participants were asked whether or not they agreed with the specific interpretation (most notably concerning Chapter 4 and 5). Together, this warrants the validity of the results.

Empirical ethics

The bonds between applied ethics and the social sciences have become increasingly tight in the last two decades. Many ethicists are currently involved one way or another in social scientific research (Nijsingh and Düwell: 2009). This fits the, already ongoing, shift in focus in ethics. From a predominantly meta-ethical enterprise in the beginning of the twentieth century, the focus in the sixties gradually shifted to more applied forms of ethics, of which medical ethics is probably the most prominent one. In these early years of applied ethics,

the focus was still mainly on the application of ethical theories to practical problems, for instance deontology and utilitarianism (Manders-Huits and Van den Hoven: 2009). In the last two decades, applied ethics has taken a step further by incorporating insights from the social sciences. In the field of engineering ethics, this so-called empirical turn has taken shape in the form of a close cooperation between scholars from STS and engineering ethicists.

Especially in the field of medical ethics, there is a fierce debate on the appropriateness of using empirical data in ethics.[21] Whereas some traditional philosophers take the extreme position of rejecting any link between empirical data (that is, context) and ethics, most critics take a more nuanced position (Musschenga: 2005). They recognize that context should be taken into account when making an ethical analysis; however, they reject the view that context could serve as a *source* of morality. This is the distinction between, what Musschenga calls, *context-sensitive* and *contextualized ethics*. Context-sensitive ethics means that recommendations and prescriptions need to be fine-tuned to the particularities of the context and active cooperation should be sought with the actors in the "target group" (be it the doctors and patients in medical ethics or engineers and the general public in engineering ethics). In other words, ethics should take seriously the context-specific findings from social sciences and attempt to incorporate such findings while developing ethical principles and theories (Musschenga: 2005, 473). Proponents of contextualized ethics give the empirical data a more prominent role by asserting that the context provides a *source* of morality. Van de Scheer and Widdershoven (2004), for example, argue that the intuitions of the people in the "target group" themselves are considered of paramount importance as those people have direct access to the specific circumstances of a practice. In my research, I take a step further. The moral intuitions and judgments of the actors are not only a source (in the sense of being informative about the specific characteristics of the practice) but they are an indispensible ingredient for the justification of the outcome. In other words, without the empirical data the whole approach would be pointless. Hence, my research fits the category of contextualized ethics and it is therefore also potentially vulnerable to the criticism on contextualized ethics. In what follows, I explain why this criticism is unwarranted.

---

[21] Cf. the recent special issues on this topic in the journals *Health Care Analysis* (Holm: 2003), *Medicine, Health Care and Philosophy* (Borry et al.: 2004a), *Bioethics* (Molewijk and Frith: 2009), and *Ethical Perspectives* (Draulans: 2010).

The idea underlying the criticism on contextualized ethics is that a theory loses normative force when empirical data is used. This criticism goes back to the eighteenth-century philosopher David Hume, who pointed out that a normative statement (an "ought") could never be derived from a factual statement (an "is") (Hume: 1739). In the remainder of the text, I refer to this point as "Hume's Law." Critics of empirical ethics argue that, as an implication of Hume's Law, empirical data could never provide the answer to normative questions.[22] Although empirical investigations can provide insight in people's opinions on morality (descriptive ethics), they could never provide us with reasons why we *ought* to do something.[23] My response to this criticism is that Hume's Law says nothing more than that empirical data *alone* could not give the answer to normative questions. To explain this point, let us take a closer look at Hume's Law.

Hume's Law allows for two interpretations: a metaphysical and a logical one.[24] The first states that no *moral* conclusions can be derived from non-moral premises alone. Frankena, for example, argues that Hume's point is that moral conclusions cannot be drawn validly from non-moral premises (Frankena: 1939, 467).[25] Hence, the problem is not so much the non-*normative* nature but the non-*moral* nature of these premises.[26] It is equally problematic to derive moral

---

[22] Note that Hume's point was mainly about the logical validity of normative claims. Some people mistakenly call this the "naturalistic fallacy." The latter term was originally coined by Moore (1903) in his *Principia Ethica*, where he argued that we should not try to capture moral terms (such as "good" or "right") in natural properties (such as "happiness" or "pleasure"), thereby especially going against Mill's utilitarianism.

[23] See, for example, Borry et al. (2004b; 2005), and De Vries and Gordijn (2009) for a more elaborate formulation or Pigden (1989), Weaver and Trevino (1994), Pellegrino (1995), and Düwell (2009) for a defense of this criticism.

[24] De Vries and Gordijn rightly point that the second interpretation, although sometimes labeled the "standard interpretation" is not uncontroversial (De Vries and Gordijn: 2009, fn 24). Adherents of the first interpretation argue that the context of the Hume's *Treatise* points to the first interpretation.

[25] Although Frankena uses the adjectives "ethical" and "moral" interchangeably, I use the word "moral" throughout the text for referring to the right or good thing and I reserve the adjective "ethical" for referring to ethics, the reflection on morality.

[26] Frankena does not seem to distinguish between normative and prescriptive statements. However, the class of normative statement is sometimes also taken to include evaluative statements which evaluate a certain state-of-affairs or option in terms of value but that do not prescribe what one ought to do (for example, aesthetic statements); see Williams (1985) and Dancy (2000; 2006) for this distinction.

conclusions from normative but non-moral premises according to this interpretation. The second – logical – interpretation states that no *normative* conclusions can be derived from descriptive premises alone (Hare: 1952; Nowell-Smith: 1954). In other words, whereas the first interpretation emphasizes the specific substantive nature of moral statements, the second interpretation emphasizes the normative form. I think the first interpretation of Hume's Law is unnecessarily speculative on the nature of moral judgments. Not counting obvious normative statements like "Thou shall not kill," there is no agreed-upon distinction between moral and non-moral content of normative judgments. Take for example the statement "You may not drive while drunk." This is obviously a *legal* statement (at least, in most countries) and therefore a normative one; for some people this is also a *moral* statement but certainly not for everyone. Hence, I think that the only generally valid interpretation of Hume's Law is the second one.

Although this strict logical interpretation does indeed state that no normative conclusions can be derived from descriptive premises alone, it does not rule out the possibility that descriptive and normative statements together could lead to a normative conclusion. In other words, in order to derive normative conclusions, descriptive and normative statements should be combined (McMillan and Hope: 2008). In a discussion of the different goals that studies in empirical ethics could pursue, De Vries and Gordijn show that most studies do not aim at arriving at a conclusion that is itself a moral judgment or principle. They state that "only studies that try to correct deficiencies in existing moral theories [...] and ones that describe the actual moral opinions of those involved in a practice [...] seem to result in a moral conclusion" (De Vries and Gordijn: 2009, 200). In such studies, though, the conclusion is not exclusively based on empirical findings, thereby avoiding violation of Hume's Law.

The application of Rawls' procedural theory to practical problems as described in this thesis (viz. to assess the fairness of responsibility distributions) is an instance of this last objective mentioned: to describe the actual moral opinions of those involved in a practice. In my empirical research, the justification consists of bringing coherence in a set of different layers of morality, containing both opinions on responsibility and moral rationales (the empirical results), and moral principles and background theories (normative statements). Thus, the moral conclusion is derived from both empirical *and* normative (moral) findings, and as such it is not a violation of Hume's Law. Even stronger, fulfilling the justificatory objective would not be possible without

the empirical data. In other words, the fact that I use empirical data in my study does not undermine but strengthens the normative conclusions of my study and it is for that reason that I take them into account.

Coherentist justification

A second characteristic of the normative research presented in this thesis, which particularly pertains to Rawls' reflective equilibrium methodology, is the nature of justification. Reflective equilibrium methodology is a coherentist model of justification, which means that its justificatory force is derived from the coherence between the different elements (here, the different layers of morality: considered moral judgments, moral principles, and moral background theories) and not from a noninferential foundation (which mostly is one particular moral background theory or principle). Since pure coherentism implies that all elements have 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. Some people therefore question the validity of including considered moral judgments; they argue that considered moral judgments do not have the stability that is 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,[27] this capacity is still considered a vague and ambiguous term (cf. Audi: 1993; Carr: 1975; Ebertz: 1993; Kekes: 1986; Nielsen: 1982b). Because of this ambiguity, different interpretations of the reflective equilibrium methodology exist. Although this does not necessarily frustrate the method, it has implications for its justificatory power. Especially when it is used as a

---

[27] In his 1951 paper "Outline of a decision procedure for ethics," Rawls provided an extensive list of the conditions under which people come to hold a valid considered moral judgment (Rawls: 1951, 181-183). 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 does no longer include 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.

practical method, choices as to which types of beliefs and arguments to include need to be made, which is inevitably selective (Van der Burg and Van Willigenburg: 1998). When we focus too much on the initial credibility of the considered judgments, the model becomes exclusive, in the sense that some judgments are not taken into account because of their alleged incredibility. However, not putting any demand on the initial credibility of the considered judgments makes the model unreliable as a justificatory method. Hence, on the one hand we need a mechanism that provides some minimum level of credibility to the considered judgments, on the other hand the ideal of inclusiveness calls for incorporating the "broadest evidence" available (Daniels: 1996, 2-3).

Whereas most philosophers give priority to reliability of the method by putting extra demands on the credibility of the initial input (cf. Beauchamp and Childress: 2001[1994]; Nielsen: 1982a), I follow the solution proposed by Van Thiel and Van Delden (2010) and DePaul (1993), who give priority to the inclusiveness of the method. Rather than trying to warrant credibility in the level of belief at the start of reasoning (labeled as the "credible input-justified output strategy"), Van Thiel and Van Delden seek credibility in the *process* of reasoning. Hence, instead of putting entry demands to the considered moral judgments, they formulate criteria indicating what "good reasoning" amounts to and, in doing so, aim at providing sufficient warrant for a reliable and inclusive outcome ("good reasoning-justified outcome strategy").[28] This good reasoning-justified outcome strategy has the advantage over the credible input-justified output strategy that it allows for the inclusion of a broader set of elements, which is the cornerstone of the method.[29] Moreover, it is also more in line with real life deliberation. When discussing moral issues, I think people are more likely to correct one another for not providing sufficient ground for their statements than for uttering a specific statement *per se* (cf. the difference between "I do not agree with you because you haven't clearly explained why you think we should do X,"

---

[28]   In their particular reflective equilibrium model, 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 and providing the people that hold them with a sense of direction in which a judgment about a given case should go (Van Thiel and Van Delden: 2010, 189).

[29]   For a more elaborate description of this strategy, see (DePaul: 1993; Van Thiel and Van Delden: 2009).

and "I do not agree with you because your utterance that we should do X is unreliable"). Except for extremely delicate topics (such as death penalty), I think people do not dismiss particular statements for lacking insufficient *a priori* credibility but for lack of consistency with other statements, which is assessed during the deliberative process. For these two reasons, I follow the good reasoning-justified outcome strategy suggested by Van Thiel and Van Delden and DePaul.

1.4.   Detailed outline of the thesis

The body of the research presented in this thesis is made up of five chapters, which originally appeared or were submitted as separate and independent papers. Except for some explanatory additions on the basis of the comments made by the Doctorate Committee, the chapters are for the greater part identical to the original papers. This implies that there is unavoidably some repetition. The general discussion of individual moral responsibility conditions, for instance, is discussed in two chapters (Chapter 2 and Chapter 6). Rawls' political liberalism is discussed in both the Introduction and Chapters 3 and 5. The only substantial redundancy I have removed is the discussion of the empirical results in Chapter 5, which to a large extent overlapped with the presentation of the empirical results in Chapter 4. Some minor discrepancies between the various papers were unavoidable due to changing insights. In Chapter 2, for example, I started with the idea of three different perspectives for ascribing responsibility with fairness being the rationale for the blame-oriented perspective. The empirical results, however, pointed to a somewhat different interpretation of fairness. Another shift that I made during the whole process concerns the notion of "conception." In the first article (Chapter 2), I use the term "perspective" to refer to the different ways to conceive of responsibility. In the later chapter, I replaced the term "perspective" with the term "conception," because of its better theoretical underpinning. These apparent differences in terminology do not undermine the main argument put forward in this thesis that there is a plurality of responsibility conceptions which leads to different, sometimes conflicting, requirements for responsibility distributions.

In order to clarify the main point in each chapter and its contribution to the main line of argumentation, the consecutive chapters are summarized in the remainder of this section.

Chapter 2: Responsibility ascriptions in engineering: Three perspectives

Forthcoming in: *Science and Engineering Ethics*: DOI: 10.1007/s11948-009-9189-3 (original title: Responsibility ascriptions in technology development and engineering: Three perspectives).

In this chapter, I introduce three main approaches or perspectives to responsibility, namely a merit-based, a rights-based, and a consequentialist one. Based on two "appropriateness criteria," I discuss the pros and cons of the three approaches. The blame-oriented or merit-based perspective is based on the traditional responsibility conditions, as found in the philosophical literature on responsibility. This perspective translates into the criterion of fairness towards potential wrongdoers. The rights-based perspective is based on the individual right of people to be safeguarded from the consequences of another person's actions (the so-called no harm principle). This perspective translates into the criterion of informed consent. The consequentialist perspective, lastly, is instrumental in nature. It is based on states of affairs or outcomes: it does not prescribe what action ought to be done but rather what should be achieved. The criterion of this third perspective is efficacy.

Using the example of the development of a new sewage water treatment plant, I show how the different approaches for ascribing responsibilities have different implications for engineering practice in general, and R&D or technological design in particular. The chapter concludes with the observation that there may be a tension between the demands that follow from the different approaches, most notably between fairness towards potential wrongdoers and the efficacy of the responsibility ascription. Since it is impossible to reduce the different demands into one overarching criterion, it is proposed to look for alternative, more procedural ways to distribute the responsibilities.

Chapter 3: Applying Rawlsian approaches to resolve ethical issues: Inventory and setting of a research agenda

Published in: *Journal of Business Ethics* 91 (1): 127-143.

In Chapter 3, I do not explicitly address the topic of responsibility in engineering but discuss the use of Rawls' procedural theory in the field of applied ethics. This chapter serves to relate the empirical part and the theoretical part of this thesis. The research presented in this thesis is an instance of empirical ethics, in the sense that it is based on empirical data. Since the methodological

underpinning of empirical ethics is sometimes problematic, I added this chapter to inventory recent applications of Rawls to learn how Rawlsian approaches are applied and to identify (practical) obstacles for applying Rawlsian approaches. To this end, a taxonomy of different kinds of applications is presented, classified according to their purpose, content, and type of justification. On the basis of this taxonomy, an inventory of 12 recent applications is made. From the inventory it followed that methodological questions, such as the question how to retrieve the relevant empirical data, are scarcely addressed in literature. To advance the use of empirical approaches in general, and that of Rawlsian approaches in particular, I argue for the use of methodological insights from social sciences to further advance the field of empirical ethics. In this chapter, I also recommend to give stakeholders a more active role in the assessment and justification of these methods.

Chapter 4: Exploring rationales for distributing moral responsibilities in R&D networks

Accepted for publication in *Science, Technology, & Human Values*

In Chapter 4, I explore the opinions of scientists and engineers on social and moral issues related to technology. On the basis of a series of interviews and a workshop with the researchers working on the ALwEN project, I investigate how actors within a research network distribute responsibilities for these issues. This chapter partly serves as an empirical check of the theoretical analysis in Chapter 2. Some of the rationales that were distinguished in Chapter 2 are confirmed by the empirical data. However, it was also found that the different rationales are not predictive for the actual responsibility distributions.

The empirical study indicates that the actors, when discussing how to address ethical issues or how to distribute the responsibility for addressing them, recurrently refer to normative arguments related to actor-specific moral background theories. It was found that these deliberative processes could best be interpreted in terms of an interplay between different layers of morality, rather than a in terms of the actors' roles. I tentatively introduce the Rawlsian WRE model to describe the outcome of these processes. The central theoretical contribution of this case relates to the different levels in moral deliberation, in particular regarding the distribution of responsibilities. Distinguishing between rationales for distributing responsibilities and the actual distributions allows for a more procedural way of justification. The benefit from acknowledging these

different rationales is that it enables actors to recognize the legitimacy of other people's opinions, ultimately contributing to an overlapping consensus on how to distribute the responsibilities and which can therefore be accepted by all as justified.

Chapter 5: A procedural approach to distributing responsibilities in R&D Networks

Published in: *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science* 7 (3): 169-188.

The descriptive analysis presented in Chapter 4 is further elaborated in Chapter 5, where I apply Rawls' procedural approach to the results of the case study. The empirical results suggest that in professional settings, people often have diverse and competing conceptions of responsibility. This concerns not only the exact *meaning* of responsibility (for example, responsibility as accountability, responsibility as virtue or responsibility as liability), but also the *conditions* that determine when it is appropriate to be or make someone responsible. When people disagree on these conditions, they will probably have different expectations of whose responsibility it is to do something. This may lead to undesirable gaps in the distribution of responsibilities because people expect someone else to do something. In this chapter, a procedural model is developed for alleviating the tension between diverging responsibility conceptions. The model is based on the Rawlsian concepts of WRE and overlapping consensus, complemented with two procedural norms based on literature on policy and innovation networks (reflective learning and inclusiveness). Analysis of the case shows that, in a pluralist setting, a procedural approach can be useful for encouraging discussion on the legitimacy of different responsibility conceptions and the question what a fair responsibility distribution amounts to.

Chapter 6: Responsibility in engineering: Towards a new role for engineering ethicists

Published in: *Bulletin of Science, Technology & Society* 30 (3): 222-230.
This is a joint publication with Jessica Nihlén Fahlquist

This chapter serves as an epilogue to the discussions on responsibility in R&D. Given the discussions in the previous chapters, what role could engineering ethicist take upon themselves in distributing responsibilities? On the basis of a

brief history of technology management, where the focus was traditionally put on the management of technology before or after the phase of development, I advocate the idea of technology management as "midstream modulation" or "real-time Technology Assessment." This requires technology managers and engineering ethicists to become more involved in the technological research rather than assessing it from an outsider perspective. Instead of focusing on the question whether or not to authorize, approve, or adopt a certain technology or on the question of who is to blame for potential mistakes, the guiding question in this new approach is how research is to be carried out. Engineering ethicists could help distributing responsibilities and ensure that no important responsibilities are overlooked.

In this chapter, a tentative list of relevant considerations is presented that could serve as a tool for assigning and distributing responsibilities. In this list, the different demands for a desirable distribution of responsibilities are accounted for. In their new role, engineering ethicists can facilitate a discussion of potential risks and other ethically relevant issues at a time the changes in the technological design can still be made. Subsequently, they can guide the discussion on how to assign and distribute the responsibilities for addressing these issues.

2 Responsibility ascriptions in engineering: three perspectives[30]

2.1. Introduction

In the last decades increasing attention is paid to the topic of responsibility in technology development and engineering.[31] The topic is often raised in the context of disasters due to technological failure, such as the Bhopal disaster (2005; Castleman and Purkavastha: 1985), the explosion of the Challenger (Davis: 1998; Harris et al.: 2005[1995]; Vaughan: 1996), and the sinking of the Herald of Free Enterprise (Berry: 2006; Richardson and Curwen: 1995). The discussion of responsibility then typically focuses on questions related to liability and blameworthiness.[32] Asking these questions might suggest that there is one, unambiguous definition of responsibility. This is far from true, however. In moral philosophy, few concepts are more slippery than that of responsibility (Miller: 2001, 455). What the questions of liability and blameworthiness share, is that the question of responsibility is asked after some undesirable event has occurred. However, the ascription of responsibility can also refer to something that ought to happen in the future: being responsible then means that an agent has been assigned, taken on, or accepted a certain task or set of obligations to see

---

[30] This chapter is forthcoming as an article in *Science and Engineering Ethics* under the title "Responsibility ascriptions in technology development and engineering: Three perspectives." I would like to thank my colleagues at the philosophy department, and Ibo van de Poel and Jessica Nihlén Fahlquist in particular, for the valuable comments they provided on an earlier draft of this article. The article has also profited from the useful comments of the reviewers.

[31] In this thesis, I only discuss the ethical aspects of moral responsibility. The metaphysics of moral responsibility, which is closely related to the free will debate, is outside the scope of the present research. The reader is referred to the vast array of literature on this topic (Berofsky: 1966; Dennett: 1984; Frankfurt: 1971; Kane: 2002; O'Connor: 1995; Pink: 2004; Watson: 1982; Widerker and McKenna: 2002; Wolf: 1990).

[32] A good example is the discussion of the case study "The West Gate Bridge: Who was Responsible?" in the engineering section of the recent anthology on professional ethics (Allhoff and Vaidya: 2008). Also Swierstra and Jelsma (2006) ask the question "to what extent engineers can be held responsible in normal practice" (309).

to it that a certain state of affairs is brought about (or prevented). In that latter case, responsibility is often ascribed from a consequentialist perspective.[33] As a third approach one could also distinguish the question of responsibility from the perspective of the rights of potential victims, which often focuses on the question who should put a situation right (e.g., by compensating for certain damage).

Recent discussions in engineering ethics call for a reconsideration of the traditional quest for responsibility. Rather than on alleged wrongdoing and blaming, the focus should shift to more socially responsible engineering, in which "to maximize the service to the larger society" should become the moral norm (Durbin: 2008, 230). Responsibility as blameworthiness should therefore be replaced by, or complemented with the notion of engineering as a responsible practice (Pritchard: 2001). Until the late 1990s, scholarly literature on engineering ethics, however, seemed to be biased towards the blame-oriented or merit-based perspective on responsibility rather than this more forward-looking perspective (Durbin: 1997; Pritchard: 2001, 391).

Similarly, both in the general field of moral philosophy, and more specifically in the field of engineering ethics, there has also been a call to shift the focus of ethics from an abstract outsider's perspective towards the practice in which moral deliberation takes place. For example, in the general field of moral philosophy Alasdair MacIntyre and Michael Walzer argue for an insider's perspective when trying to improve a practice.[34] In the field of engineering ethics, philosophers such as Michael Pritchard, Mike Martin, Vivian Weil and Michael Davis are firm proponents of taking an insider's perspective on engineering and its ethical issues. Michael Davis, for instance, argues that the

---

[33] One could also distinguish a virtue ethics approach to responsibility, which is forward-looking as well. In the remainder of the text I focus on the consequentialist perspective in general, since the virtue ethics approach is primarily aimed at relations between people. This does not imply that there are no leads to apply this approach to the field of technology development and engineering, but until now this has hardly been done. The elaboration of this relatively novel approach falls outside the scope of the present thesis.

[34] Michael Walzer argues that only a passionately committed "connected critic" can effectively challenge a prevailing culture. Such a critic can only be effective because he is committed and involved (Walzer: 1987; Walzer: 2002). Alasdair MacIntyre questions the distinction between theory and practice. These are thoroughly intertwined, and as such, the search for the good life always develops on the basis of the embeddedness in a particular practice. By participating in a practice, people form their opinions of what the good life amounts to. Moral deliberation should therefore not be separated from the practice itself (MacIntyre: 1984[1981]).

discussion of responsibility is too much about "holding others responsible" instead of "assuming responsibility" (Davis: forthcoming).

This shift from an outsider's perspective towards an insider's perspective might have implications for the topic of responsibility as well. The present chapter aims at exploring three main approaches to responsibility in order to see which one is most appropriate to apply in engineering and technology development, where I take appropriateness to mean two things:

(1) The approach should reflect people's basic (and preferably enduring) intuitions of when it is justified to ascribe responsibility to someone. An approach that contravenes these basic intuitions will probably be deemed unfair. Whether such an approach should depart from abstract principles and work top-down to considered judgments about particular cases, or depart from these considered judgments and work bottom-up to more general principles is still open for discussion. It is important, though, that people recognize that the responsibility ascription is justified.

(2) The approach should inform the direction of technology development and therewith improve technological design. In order for this to be so, it should be possible to apply the approach to specific contextualized moral issues that are raised by specific technological and scientific developments rather than to more general abstract issues. This second requirement follows from recent discussions within engineering ethics, and ethics concerning NEST (New and Emerging Science and Technology) in particular, in which it is argued that the ethical and social aspects of new technologies should be addressed at an early stage of technology development in order to adapt technology to society's needs (Swierstra and Rip: 2007; Van de Poel: 2008).[35]

The outline of this chapter is as follows. I first discuss three different perspectives for ascribing responsibility: a merit-based perspective, a rights-based perspective and a consequentialist perspective. After a brief intermezzo on

---

[35] Cf. the contributions in the special issue on Ethics and Engineering Design in *Science, Technology, & Human Values* (May 2006 issue), edited by Van de Poel & Verbeek (2006).

forward-looking and backward-looking responsibilities, I apply the three perspectives to the example of the development of a new sewage water treatment plant. A comparison of the three approaches shows that the consequentialist perspective is especially suited for distributing responsibilities since it is most akin to the engineering work and it (therefore) offers the best opportunities for improvement of technological design. The chapter ends with recommendations for further developing the field of engineering ethics by incorporating insights from political philosophy.

2.2. Three perspectives for ascribing responsibility

In this section, I discuss three approaches or perspectives for ascribing responsibility: a merit-based perspective, a rights-based perspective and a consequentialist perspective.[36] Although the latter is common in non-philosophical discussions (for example in organizational and management literature), the philosophical literature is mainly focused on responsibility as blameworthiness (i.e., the merit-based perspective).[37] Being the most common approach in philosophical literature, I start the present overview with this merit-based perspective.

2.2.1. A merit-based perspective on responsibility

In the philosophical literature on moral responsibility, the aim for ascribing responsibility is mostly retributivist. In the traditional view, being morally responsible means that the person is an appropriate candidate for reactive attitudes, such as blame or praise (Fischer and Ravizza: 1993; Miller: 2004; Strawson: 1974). Being morally responsible (i.e., being eligible for reactions of praise and blame) is not the same as being causally responsible. One can imagine a situation where a person did indeed causally contribute to certain outcome but is not eligible for moral evaluation, and hence not for reactive

---

[36] The way the three approaches are presented here might suggest that, when talking about responsibility, people apply either one of the three approaches. In reality hybrid approaches exist as well. However, as an analytical concept it seems useful to separate the three approaches since they each serve a different purpose and as such they are distinct.

[37] A few exceptions are Goodin (1995), Van den Hoven (1998), Young (2006), Nihlén Fahlquist (2006b; 2009).

attitudes of praise or blame (e.g., in case of positive outcomes due to sheer luck, or negative outcomes which one could not reasonably avoid). In both cases, it is not warranted to praise or blame the person for the outcome. Hence, since moral responsibility in this above elaborated view is related to reactive attitudes, which may have consequences for the well-being of an agent, the ascription of moral responsibility is only warranted if these reactive attitudes and their consequences are merited or deserved (see Eshleman: 2008; Magill: 2000; Wallace: 1994; Watson: 1996; Zimmerman: 1988). This is usually translated into certain conditions that have to be met before it is fair to ascribe responsibility to someone. In the remainder, I call this the fairness criterion of responsibility ascriptions. Although academics disagree on the precise formulation, the following conditions together capture the general notion of when it is fair to hold an agent morally responsible for (the consequences of) their actions (see Bovens: 1998; Corlett: 2006; Feinberg: 1970; Fischer and Ravizza: 1998; Hart and Honoré: 1985):

(1) Moral agency: the responsible actor is an intentional agent concerning the action. This means that the agent must have adequate possession of his or her mental faculties at the moment of engaging in the action. Young children and people whose mental faculties are permanently or temporarily disturbed will not be (fully) held responsible for their behavior. However, to put oneself knowingly and voluntarily into a situation of limited mental capacity (by drinking alcohol or taking drugs for example) does not, in general, exempt one from being responsible for the consequences of one's behavior. Some people phrase this condition in terms of intention, meaning that the action was guided by certain desires or beliefs.

(2) Voluntariness or freedom: the action resulting in the outcome was voluntary, which means that the actor is not responsible for actions done under compulsion, external pressure or hindered by other circumstances outside the actor's control. The person must be in the position to determine his own course of action (cf. condition 1), and to act according to that.

(3) Knowledge of the consequences: the actor knew, or could have known, the outcome. Ignorance due to negligence, however, does not exempt one from responsibility.

(4) Causality: the action of the actor contributed causally to the outcome; in other words, there has to be a causal connection between the agent's action or inaction and the damage done.

(5) Transgression of a norm: the causally contributory action was faulty, which means that the actor in some way contravened a norm.

Note that especially the first two conditions are closely interrelated. Being an intentional agent means that one has the opportunity of putting the will into effect and that one is free from external pressure or compulsion (Lewis: 1948/1991; May and Hoffman: 1991; Thompson: 1980). With regard to the fifth condition, extensive debate has been going on as to what counts as a norm. In daily life the norm can be much vaguer than in criminal law where the norm must be explicitly formulated beforehand (the *nullum crimen, nulla poena sine praevia lege poenali* principle).[38]

2.2.2. A rights-based perspective on responsibility: The no harm principle

A second approach for ascribing responsibilities within the field of science and technology is based on the individual right of people to be safeguarded from the consequences of another person's actions (the so-called no harm principle). This implies that "actions are right if and only if: either there are no (possible) consequences for others; or those who will experience the (possible) consequences have consented to the actions after having been fully informed of the possible consequences" (Zandvoort: 2005b, 46). The aim of this approach is remedial: it refers to the duty or obligation to put a situation right (Miller: 2004). In practice this rights-based approach translates into two requirements for decision making regarding the development, production and use of technology (Zandvoort: 2008). The first is the (legal) requirement of strict liability, which holds that actors are unconditionally required to repair or fully compensate for any damage to others that may result from their actions, regardless of culpability or fault (Honoré: 1999; Van Velsen: 2000; Vedder: 2001; Zandvoort: 2005a). Hence, the question of responsibility is reduced to the question "Who caused the particular outcome?" (causal responsibility). As such, blame is not the guiding

---

[38] The literal translation of this principle reads "no crime, no punishment without a previous penal law."

concept in ascribing responsibility.[39] The second requirement relates to the principle of informed consent, which holds that "for all activities that create risks for others, all who are subjected to the risks must have given their informed consent to the activities and the conditions under which the activities are performed" (Zandvoort: 2008, 4).

Instead of fairness towards potential wrongdoers, this approach focuses on fairness towards potential victims. Given the importance of informed consent, the engineering ethics literature on this approach to responsibility therefore focuses on the conditions under which consent can be gained and its implications for, for example, risk communication and risk assessment.

2.2.3.   A consequentialist perspective on responsibility

The third perspective for ascribing responsibility is the consequentialist perspective. In the consequentialist perspective, responsibility is ascribed for instrumental reasons rather than retributivist (merit-based) or remedial (rights-based) reasons. In the consequentialist perspective, the most important question when ascribing responsibility is not whether the reactive response triggered by the responsibility ascription is warranted but whether the reactive response would likely lead to a desired outcome, such as improved behavior by the agent (Eshleman: 2008).[40] Where fairness is the main criterion for the merit-based perspective and informed consent the basis for the rights-based approach, efficacy is the criterion for consequentialist responsibility ascriptions, which means that they should contribute to the solution of the problem at hand (Nihlén Fahlquist: 2006a; 2009). According to a strict consequentialist view, the responsibility ascription that yields the best consequences is the morally optimal

---

[39] This does not necessarily hold for all versions of liability. The principle of fault liability holds that an offender can only be held liable in case of culpably careless or faulty behavior (Zweigert and Kötz: 1998[1977]). Note also that some authors defend the claim that establishing someone as the cause of some undesirable event is already a way of blaming – in the sense of criticizing – the particular person; hence, also strict liability includes an element of blame, these authors argue (see, e.g., Davis forthcoming).

[40] Note that this consequentialist perspective does not imply that one necessarily promotes some material utility function. Kutz, for instance, defends an instrumental – or functionalist, as he calls it – conception of responsibility without claiming that practices of accountability are aimed at optimizing aggregates states of social welfare. Accountability, in Kutz' view, serves to sustain relationships among discrete individuals (Kutz: 2000, 54).

responsibility ascription. Responsibilities, in this view, do not take specific actions of persons as their object but they rather have the character of obligations to see to it that a certain state of affairs is brought about (or prevented). As such responsibilities are outcome and result oriented (Van den Hoven: 1998, 107).

In the case of engineering and technology development, this consequentialist perspective could be taken to imply that for a technology to be "right," in the sense that it is from a societal point of view desirable or at least acceptable that the technology is being developed, potential implications for society (e.g., human health and the environment) should be taken into account during the design phase. In other words, for every potential implication, whether this is a risk or some other problematic issue, someone should be ascribed the responsibility to address this issue. This does not mean that all risks should be completely excluded – a requirement which is impossible to live up to – but that at least everything that can reasonably be known should be considered during design and development phase. Sometimes this might imply that, after deliberation, a potential risk will be accepted as is since the (societal) costs of preventing it do outweigh the (societal) costs of accepting it.[41]

2.2.4. The three perspectives compared

In the overview presented above, a distinction was made between the goals that were aimed at in the different perspectives. In addition to a different aim, we could also say that the three approaches each depart from a particular moral background theory and that they each try to answer a different moral question.[42] The merit-based approach fits into a deontological framework, which is

---

[41] For the moment, I leave it open how to determine "what can reasonably be known." For the remainder of the argument it is not required to exactly define it. A good starting point might be the state of knowledge of peers in one's field.

[42] Similar to what was said in footnote 36, this classification is meant for analytical clarification and as such it shows a somewhat simplified picture of the "ethical landscape." The use of a merit-based perspective is not applied exclusively by deontologists, neither is it impossible to think of consequences in a deontological or rights-based discourse. It should also be noted, though, that moral philosophers who take rights as their starting point should also have something to say about duties, as a right to protection or compensation cannot exist without someone else's duty to protect or compensate. However, the primary focus of the three approaches is sufficiently different to distinguish between the three. The classification does describe the moral theory *most akin* to a certain responsibility perspective.

primarily a theory of "right actions." The rights-based approach fits into an ethics of rights and freedoms (see, e.g., Mackie: 1978; Nozick: 1974). This theory shares with deontological ethics that it takes "action" as the primary object of evaluation. Where deontological ethics departs from duties, a right-based discourse departs from people's individual rights and freedom and uses these to determine which actions are permissible and which are not. In both cases the content of the responsibility ascription is action that ought to be abstained from (merit-based) or that ought to be done (rights-based): to breach a duty is to perform a blameworthy action (merit-based) or to be liable for compensation (rights-based). The consequentialist approach, which (unsurprisingly) fits best into some form of consequentialism, has a different focus. Rather than on particular action, the consequentialist approach is focused on states of affairs. It does not prescribe what action ought to be done but rather what should be achieved.

A summary of the three approaches is listed in Table 2.1.

Table 2.1: Perspectives for ascribing responsibility

Perspective	Ethical Theory	Aim	Criterion	Content
Merit-based	Deontological ethics	Retributivist	Fairness	Actions
Rights-based	Ethics of rights and freedoms	Remedial	Informed consent	Actions
Consequentialist	Consequentialism	Instrumental	Efficacy	States of affairs

2.3.   Forward-looking versus backward-looking responsibility

Before continuing the application of the three perspectives on a real engineering case, some clarifications regarding responsibility need to be made.

One could argue that the merit-based and the consequentialist perspective responsibility are not comparable in the sense that they refer to different time horizons. We therefore cannot speak of two perspectives on the same concept but we rather should speak of two different types of responsibility, each with a different criterion. For example, the merit-based perspective is often applied after-the-fact and it is therefore backward-looking or retrospective.[43] The

---

[43]   It should be noted that merit could play a role in the ascription of forward-looking responsibilities as well. One could think of responsibility ascriptions on the basis of one's past contribution, seniority, or ability.

consequentialist perspective is often applied in a forward-looking or prospective sense (i.e., before something has happened). However, despite the difference in focus, the two perspectives are closely related. Imagine an engineer E who designs some artifact A. Unfortunately, there is a serious flaw in the design and the artifact causes the death of some innocent person P. Imagine further that E could have easily designed an artifact A* with similar (functional) characteristics but without the property leading to the death of P. In fact, E knew that the design was flawed and he intentionally did not improve the design, even though he had the freedom to do so. From this we would probably conclude that E is morally responsible for the death of P. But why is that so? As explained in Section 2.2.1, the merit-based perspective involves a moral assessment of the agent in terms of the conditions discussed above. Except for the condition of causation, which determines whether someone did causally contribute to a certain outcome, the other four conditions bridge the gap between causal and moral responsibility. In the example, four conditions are obviously met: the engineer is a moral agent (condition 1), he was free (2), he knew of the consequences (3) and he causally contributed to the death (4). But what about the fifth condition: the transgression of a norm? Most people would probably say that E is blameworthy because he did not pay enough attention to the lethal consequences of the artifact. Apparently, the fifth condition in the merit-based perspective implies a forward-looking responsibility to be careful or to pay attention. Both in law and professional ethics this forward-looking responsibility is operationalized in the duty of (reasonable) care to avoid (foreseeable) harm to others. At the minimal level, this duty of care implies that E should consider how to redirect foreseeable harm to people who are affected by his artifact. However, it could also be argued that he has the (broader) responsibility to look after potentially dangerous but as yet unforeseen risks. The duty of care implies that there are certain acts or omissions that should be avoided. In this simplified case, the duty of care requires that the engineer should not develop artifact A but rather A*. Hence, also in a merit-based perspective, people have forward-looking responsibilities.

If we depart from the consequentialist view, we also see that the forward-looking and backward-looking responsibilities are closely related. It is because blame and praise can have a motivational force to take up one's forward-looking responsibility that backward-looking responsibilities are being ascribed. Hence, forward-looking responsibilities translate into backward-looking responsibilities and vice versa.

2.4.   Development of a new sewage treatment technology

Now we have clarified the different approaches to ascribing responsibility, we can apply these to the field of technology development. I do so on the basis of an ethical research that was carried out parallel to the technical development of a new sewage treatment technology (Van de Poel and Zwart: 2010; Zwart et al.: 2006). The idea behind this so-called embedded ethical research or ethical parallel research is that ethical investigations are carried out parallel to, and in close cooperation with, a specific technological R&D project.[44] The ethicists interact with the technological researchers, allowing the ethicists to co-shape new technological developments. By applying the three responsibility perspectives (merit-based, rights-based, and consequentialist) to technology development, I explore the appropriateness of the different perspectives in engineering practice in terms of the two criteria formulated in the introduction of this chapter.

2.4.1.   Ethical parallel research into the upscaling of the GSBR technology

The ethical parallel research concerned the development of a new sewage treatment technology, the so-called granular sludge sequencing batch reactor (GSBR) (see Text box 2.1 for a description of the technology). In the technological project, different parties contributed, classified by the ethical parallel researchers according to their role in the project team. These were the role of researcher, technology producer (including activities like design and consultancy), user of the technology, and financer of the technology. The ethical parallel research consisted of a qualitative research, based on interviews, document analysis, attendance of technical meetings, and the organization of an interactive session in the Group Decision Room (GDR; an electronic brainstorming facility) with the different stakeholders, where questions related to risks and responsibilities were addressed.

---

[44] The names "ethical parallel research" and "embedded ethical research" both refer to ethical research done on an technological project. Strictly speaking, the two approaches are different in nature. In case of ethical parallel research, the ethicists do not become part of the team but rather do their investigations parallel to the technological project. In case of embedded ethical research, the ethicists become part of the technological research team. However, notwithstanding these differences, the names are often used interchangeably. The ethical research on the GSBR technology concerned parallel investigations.

One of the crucial elements in the development of the technology was the upscaling of the three-liter laboratory reactor to an outdoor pilot plant of 1.5 m³. This upscaling was partly based on several unproven assumptions about which microbiological mechanisms are at work. The ethical parallel research, therefore, focused on the question of how this incompleteness of knowledge was dealt with in the choice of scaling-up steps. Incomplete knowledge can lead to the introduction of certain risks, which may become manifest in the research done during the development of the technology, but also later in the eventual use of this technology. The aim of the ethical parallel research was to find out how risks and uncertainties are handled and how this is open to improvement.

During the ethical parallel research, it was observed that the risks due to so-called secondary emissions (i.e., unwanted but not yet regulated substances in the effluent) were not addressed by any of the engineers and researchers involved. The users of the technology delegated the risk of secondary emissions to the research phase, for which they were not primarily responsible, and most of the researchers allocated the risk to a phase for which they in turn bore no responsibility. Nobody therefore assumed responsibility for dealing with this risk. The argument put forward by the researchers and users was that the impact of the risks due to these secondary emissions was negligible and that problems were expected to be solvable in the next phase of the research. This was based on the presumed similarity between biological processes in traditional sewage plants and the biological processes in the GSBR technology. As a result, the issue who is responsible for checking or preventing secondary emissions never became an object of discussion. The ethical parallel researchers state that it cannot be concluded that "such emissions are a serious cause of concern; the situation is rather one of insufficient knowledge. Thus the question arises which of the actors in the network are responsible for reducing this knowledge deficiency, and which actors are responsible for reducing potential secondary emissions in case they turn out to be a serious concern" (Van de Poel and Zwart: 2010). As a result of the ethical parallel research, the consultancy firm together with the university applied for additional funding to carry out research into the secondary emissions.

In the remainder of this section, I try to show how the different responsibility approaches can be applied to the development of this new technology and how these affect engineering practice, focusing on the issue of secondary emissions.

Text box 2.1: Development of a granular sludge sequencing batch reactor

One drawback of traditional biological wastewater treatment plants is their large space demand or footprint, which is caused by the use of separate settling tanks and the slow settling velocity of the sludge. In the aerobic GSBR technology both size increasing factors are addressed. By using high-density granules, the time needed for the sludge to sink to the bottom at the end of each cycle is substantially reduced. Subsequently, the shorter deposit time increases the throughput of the installation and reduces the footprint. Second, it is hoped that different ecological zones inside the granules will be able to take care for the entire treatment process in one reactor instead of several separate tanks.

The GSBR technology has been developed at the Department of Biotechnology, Delft University of Technology, the Netherlands. After successful laboratory experiments, the Dutch Foundation for Applied Water Research (STOWA) was found willing to invest in the scaling-up of the three-liter laboratory reactor to an outdoor pilot plant of 1.5 m^3. In parallel to the upscaling of the pilot plant, funds were acquired for a PhD-project (funding organization: Technology Foundation STW). Finally, an international engineering and consulting firm, with water management technology as one of its main domains, showed interest in the commercial exploration of the GSBR technology. This firm was in charge of the research at the pilot plant, operated by a local water board. The results of the pilot plant have been positive and the firm anticipates a large demand for GSBRs (Van de Poel and Zwart: 2010; Zwart et al.: 2006).

2.4.2.   A merit-based perspective on harm caused by the GSBR technology

The first approach I discuss is the merit-based perspective on responsibility. In Section 2.3 it was shown that, although focused on blame, the merit-based perspective implies the ascription of forward-looking responsibilities as well. It was argued that these forward-looking responsibilities are primarily derived from the duty of reasonable care. This means that people should take measures against foreseeable harm and possibly also look after as yet unforeseen harms. It is notoriously difficult to assess what "reasonable care" exactly amounts to in technology development, especially in the case of new and emerging technologies where the consequences are even harder to predict. A possible starting point for the evaluation of due care is the test of independent peers. If

peers think that some negative consequences were foreseeable, we could probably conclude that the engineers did not exercise due care.

Let us assume that the GSBR technology is being further developed and commercially exploited. Now suppose that secondary emissions, contrary to expectations, cause some problems for farmers who have their surface water treated with the GSBR technology. Can we point to some person or institution as being morally responsible for these problems? The ethical parallel researchers asked the developers of the technology whom they would ascribe moral responsibility for the secondary emissions to (in the sense of preventing or investigating the harmful effects). They did not get a unanimous answer: some ascribed the responsibility to the researchers at laboratory scale, some to the operators of the pilot plant and some to the users of the technology. Some even argued that no-one carries moral responsibility for these harmful consequences because "introduction of new technology introduces risks and we have to learn to live with that" (Van de Poel and Zwart: 2010). The latter answer suggests that the principle of due care was not breached at all. However, the fact that some researchers from adjacent scientific fields did express their concerns about the technology (Van de Poel and Zwart: 2010, 193-194) suggests the opposite. Apparently, before involvement of the ethical parallel researchers there was not enough incentive to take up the forward-looking responsibility to further investigate the potential risks of these secondary emissions, even though the researchers were aware of the lack of knowledge regarding these emissions. As such we could say that the duty of reasonable care was undefined.

If we discuss moral responsibility in terms of the traditional criteria, probably no-one can be held morally responsible. Although the different actors all contributed to the development of the technology, we cannot single out one particular actor or institution that fulfilled all the criteria. Whereas the criteria, if applied to the complete research group, were fulfilled, probably none of the actors or institutions within the research group fulfilled all the responsibility criteria individually. Especially the knowledge condition, requiring that one can only be held responsible if one knew or could have known the negative consequences, is a problematic condition in this case. Since none of the actors took up the responsibility to reduce the knowledge deficiency regarding the secondary emissions, which would show that the secondary emissions are not as harmless as the technology developers thought they were, no further preventive measures were taken to reduce the risks. This suggests that this is in instance of "willful ignorance," something deserving of blame. However, in case something

goes wrong, it is not clear who should have taken up the responsibility for preventive measures. The responsibility for the knowledge deficiency probably lies with the researchers, whereas the causal responsibility lies with the technology producers and users. Hence, if we apply the five conditions of the merit-based approach, nobody can be held responsible for the negative consequences of the technology (i.e., the secondary emissions), even though the research team as a whole breached the duty of (reasonable) care.[45] In the literature this is called the problem of many hands, which is first defined as such by Thompson (1980).[46] It refers to the difficulty to identify, even in principle, the person responsible for some outcome, if a large number of people is involved in an activity. But sometimes it is the joined acting of individuals within a collective that brings about negative consequences, precisely because collectives can create potentially greater harms than individuals working independently. Acting on an individual basis, neither the water board nor the researchers could have built a treatment plant with the innovative technology, but as a collective they were able to do so.

Some people therefore propose to hold the collective as a whole morally responsible. All individuals within the collective are held equally responsible (May and Hoffman 1991). This ascription of responsibility to the whole collective is criticized for being morally unsatisfactory. People are then being held responsible for the conduct of others, which is unfair (Lewis 1991). This raises a fundamental problem for individual responsibility: either no-one can be fairly held responsible and hence the problem of many hands occurs, or moral responsibility is ascribed to the whole collective of people who in some way contributed to the outcome, leaving aside an individual assessment in terms of the responsibility conditions, which is unfair. The latter holds especially if sanctions are coupled to the ascription of responsibility. After all, being part of a collective that caused some negative event does not imply that one's individual

---

[45] It should be emphasized that in reality the research team *did* further investigate the secondary emissions and so the duty of care was adequately exercised.

[46] Although the problem of many hands is mostly discussed in retrospective terms, it is strictly speaking not limited to backward-looking responsibilities. One could also think of a situation where people need to distribute a number of tasks or subtasks to bring about a certain goal. In case this distribution of responsibilities is not complete, for example because certain necessary tasks or subtasks are overlooked, the problem of many hands manifests itself as well.

actions were immoral or illegitimate and hence that one is eligible for blame.[47] Some philosophers have therefore introduced notions to distinguish between individuals who are responsible for the conduct of the organization and individuals who are not. Kutz (2000) gives a minimalist criterion for individuals to be responsible for the group's outcome. If individuals act on overlapping participatory intentions, they can be said to be promoting a collective act and be responsible for the outcome. Similarly, May (1992) argues that individuals are responsible for the organization's actions when they voluntarily joined the group. At the very least we could say that people who voluntarily join a group and derive benefits from the membership of that group are to some extent blamable. In a professional context, this criterion is often fulfilled.

We could see this "collective problem" also as a tension between what we owe to potential wrongdoers (not being blamed unless it is fair to do so) and what we owe to potential victims (to make someone responsible for preventing disasters). Although I think that individual responsibility should not too easily be dismissed on the grounds that individuals are powerless cogs in the machinery of their organization, the point remains that this traditional individualistic approach seems to put much more emphasis on what we owe to potential wrongdoers than on what we owe to potential victims.[48] Consequently, the problem of many hands is a serious threat to this approach.

This more conceptual problem of individual responsibility raises an important practical problem as well. Due to the inability to ascribe moral responsibility, an important opportunity for improvement is missed. Ascribing moral responsibility may lead to learning processes, which may ultimately prevent similar disasters from happening again in the future. If no-one can be

---

[47] Some philosophers have therefore introduced notions to distinguish between individuals who are responsible for the conduct of the organization and individuals who are not. Kutz (2000) gives a minimalist criterion for individuals to be responsible for the group's outcome. If individuals act on overlapping participatory intentions, they can be said to be promoting a collective act and be responsible for the outcome. Similarly, May (1992) argues that individuals are responsible for the organization's actions when they voluntarily joined the group. At the very least we could say that people who voluntarily join a group and derive benefits from the membership of that group are to some extent blamable. In a professional context, this criterion is often fulfilled.

[48] Note that alternative ethical outlooks, notably a virtue ethics approach, may put more emphasis on what we owe to potential victims since these approaches depart from questions about the good life and virtuous behavior rather than a rights-based or duty-based discourse (Ladd: 1982).

held responsible, this opportunity for learning will not be fully exploited (Nihlén Fahlquist: 2006a).

Summarizing, if we adopt the merit-based perspective on responsibility it is difficult to ascribe responsibilities. In the light of engineering practice, this approach seems rather powerless. In the extreme case, no-one learns from the mistakes being made and the development of the technology continues as if nothing happened. As a consequence, there is little incentive to take up the forward-looking responsibility to prevent negative consequences.

2.4.3.   A rights-based perspective on harm caused by the GSBR technology

As said in Section 2.2, the rights-based perspective focuses on the task or obligation to set a situation right. With regard to the question of liability, all people involved in the project (including the end users) unanimously agreed that water boards using the new technology are legally liable when incidents (such as problems related to the secondary emissions) would occur.

If we apply the principle of strict liability, it is questionable whether institutions, such as the water board in the present example, will ever participate in innovative research projects. They will most probably be very reluctant in participating in the development of innovative and radically new technologies. Some scholars even argue that unrestricted liability would hamper any large-scale investment, also desirable ones (Perrott: 1982). After all, existing problems sometimes require radical technological innovations (think of technological innovations relating to green energy). Technologies are primarily developed to "change positively the quality of life" (Berloznik and Van Langenhove: 1998, 24), in the sense that they try to solve or reduce existing problems. In the development of new technologies trade-offs have to be made between competing values; in the GSBR case between sustainability and safety. In engineering, more generally, overdesign (to enhance safety) comes at the cost of excess usage of energy and other natural resources. The categorical rejection of a technology because it does not satisfy one of the demands is not a viable option, since this creates risks of its own (Sunstein: 2005). With its focus on avoiding new risks, the strict liability approach may ignore existing ones.

As explained in Section 2.2.2, the procedure of "informed consent" is introduced as a possible response to this problem: in case of risk for irreversible harm the principle of strict liability requires that consent of all people who are subjected to this risk be obtained. If this consent cannot be obtained, the risk

should simply not be posed (Zandvoort: 2008, 8). The fact that this approach takes seriously the perspective of potential victims of (high-risk) technologies is unmistakably a strength. The risks of these technologies cannot be imposed on anyone without his or her informed consent. Hence, an unfair distribution of risks by majority decision making is not allowed according to this approach. However, despite its democratic aim, this approach runs the risk of paralyzing the debate on potentially risky technologies. After all, the principle of actual consent implies that anyone has the right to veto activities that impose risks, which ultimately creates a society of stalemates where nothing can be done, as Hansson (2006; 2009) argues. Informed consent is problematic if applied to affected individuals collectively. Zandvoort (2008) therefore discusses procedures to increase the willingness to consent. These are all based on monetary compensation (either directly or indirectly, such as the building of a new city theatre in case the city consents to the building of nuclear plant in the neighborhood) or improvement of the credibility of risk assessment. It is striking that both approaches do not give any incentive to improve the technology itself. The focus is on ready-made technologies rather than participation in decision making process along the way of development (Hansson: 2006, 150).

Summarizing, the rights-based approach emphasizes the right of people to be safeguarded from harm caused by others. However, the operationalization of this right by way of the principle of informed consent is problematic in the context of collective decision making. Moreover, the approach in itself seems problematic because of its focus on monetary compensation instead of improvement of the technology.

2.4.4. A consequentialist perspective on potential harm caused by the GSBR technology

The third approach is the consequentialist perspective, which is in fact the approach that was taken by the ethical parallel researchers. In Section 2.4.1, I discussed how the ethical parallel research influenced the development of the GSBR technology. The ethical parallel research led to the identification of gaps in the distribution of responsibilities, in particular the responsibility for secondary emissions. As a result, funds were acquired to carry out additional research into the secondary emissions. As such the analysis of the responsibilities by the ethicists led to an improvement of the division of labor

amongst the technological researchers and engineers, which in turn led to an improved technological design. The responsibilities were not distributed on the basis of fairness criteria but on the basis of efficacy and the question what is needed to discharge one's responsibility (that is, the required capacity, power, resources).[49] By making the technological research team aware of the responsibility issues, some of the technological researchers took the initiative to incorporate the secondary emissions in the research project. As such the effect of the ethicists' involvement on the engineering practice was not blaming or sanctioning but rather that of co-shaping. The ethical parallel research did not so much pose limits to the technology development but guided it.

Summarizing, since responsibilities are ascribed according to the criterion of efficacy, the problem of many hands does not manifest itself – or at least, not as severely – as would be the case in a strictly merit-based perspective. By taking a consequentialist stance, the ethicists encouraged the engineers and researchers to improve the technological design.[50]

2.5.   The three perspectives compared

If we compare the different approaches all three have their merits. The merit-based perspective emphasizes the fairness of a responsibility ascription. It takes seriously the moral question: who, from a moral point of view, is or should be responsible? This moral notion of responsibility is in line with common morality, and especially in case of victims of irreversible harm, people will be interested to hear the answer.[51] We sometimes "want to ascribe responsibility to the person who is responsible – for example, someone who intentionally and

---

[49]   Indirectly, one's capacity, power, and resources play a role in the merit-based perspective as well (without the necessary capacity, power, and resources, it is not fair to be ascribed a certain responsibility). However, whereas the merit-based perspective seems to take these for granted, the consequentialist perspective is more actively focused on making sure that the required capacity, power, and resources become available, such as to be able to adequately discharge one's responsibility.

[50]   The encouragement to take up the forward-looking responsibility to improve technological design seems in line with the virtue ethics aim of "responsible engineering."

[51]   Although the term "common morality" is a slippery term, most people agree that there are certain values that most "thoughtful people implicitly use in arriving at moral judgments" (Gert: 2004). I think that the fairness of responsibility ascriptions is part of this shared system of morality, which is also reflected in penal law.

culpably brought about an unwanted event – irrespective of the impact on future events of our responsibility ascriptions" (Nihlén Fahlquist: 2006a, 17). The merit-based perspective does make a serious attempt to try to answer this question of "Who is responsible?" However, this classical view on responsibility is based on an individualistic assessment of responsibility, as we saw, which makes it problematic in the context of collective action. Kutz (2000) argues that, as long as individuals are only assessed in terms of the actions they produce, the disparity between collective harm and individual effect results in the disappearance of individual responsibility. And with the disappearance of responsibility, so goes an incentive for individuals to improve their behavior, he argues (113).

The question of "Who is responsible?" was found to be less problematic in the rights-based approach, since it uses only the causal condition rather than the full range of responsibility conditions. With its focus on compensation and consent, this approach put most emphasis on the interests of potential victims. However, it was also shown that this approach gave no or only little incentive to actually improve technological design.[52] Moreover, this approach seemed to have a hampering effect on the exploitation of innovative new technologies.

The consequentialist approach, as a third approach, appeared to be most powerful in terms of the second point identified at the start of the chapter: the ability to shape the direction of technology development. It should be noted first that engineers themselves are often driven by a consequentialist heuristic of "problem solving" (Davis: forthcoming). More than discussing who is to blame, they are guided by questions of how to prevent the (re-)occurrence of harmful events. This attitude of "problem solving" is necessarily context-specific. When engineers design a new technology they want that technology to work under real-world circumstances and not only in a laboratory. They therefore engage in extensive studies of errors and mistakes. As Davis puts it,

> Whatever is true of other professionals, engineers consider it their responsibility to
> study any disaster that seems to arise from what they did – and to report what they
> find. To commit a certain mistake once, even a serious one, is something engineers

---

[52] Although it could be added that, as a secondary effect, legal liability for bad design leads to higher insurance rates for bad design, hence encouraging liable organizations to take improve their design. However, in case of radical new design, this approach may still be overly risk-averse towards new technologies and therewith too much inclined to accept the status quo.

> tolerate as part of advancing technology [...]. What engineers do not tolerate is that an engineer, any engineer, should make the same mistake. Once a mistake has been identified, the state of the art advances and what was once tolerable becomes intolerable (a kind of incompetence). [...] Engineering is unusual among professions in recognizing an obligation to "acknowledge their errors." (Davis: forthcoming)

We could say that the consequentialist perspective is most typical of the engineering practice itself. The background question is always "Does it solve the problem at hand?" By focusing on real issues rather than abstract duties or principles, this perspective is also more sensitive to the context in which technology development takes place.[53] If a certain responsibility ascription does not lead to the desired solution to a real problem, this responsibility should not be imposed or should be imposed differently. Compare this with the rights-based perspective that focuses solely on the question whether or not ready-made technologies are harmful. The rights-based perspective seems to influence not so much the direction but rather the pace of technology development.

Secondly, the consequentialist approach allows for more fine-grained responsibility ascriptions. Since the merit-based perspective is often applied after the fact (i.e., after something undesirable has happened), the question of responsibility becomes a matter of all-or-nothing: one is either responsible for the undesirable outcome or not (Bovens: 1998; Goodin: 1985; Lynch and Kline: 2000). Some therefore argue that this merit-based perspective is about non-responsibility: it defines excusing conditions that exempt people from responsibility (Ladd: 1989). However, recent insights from Science and Technology Studies (STS) show that before dramatic cases occur, often incremental small decisions have to be made that ultimately lead to undesired outcomes. Instead of focusing on blame for these – sometimes catastrophic – events, engineering ethics should pay more attention to the "complexities of engineering practice that shape decisions on a daily basis", STS scholars argue (Lynch and Kline: 2000), in order to modulate technology into the desired direction (Bovens: 1998; Swierstra and Jelsma: 2006; Van de Poel and Van Gorp: 2006). The consequentialist responsibility ascription is based on the

---

[53] It should be noted that this does not hold for consequentialism in general. A common critique of consequentialism, as an ethical ideology, is that it is too narrowly focused on the promotion of one single value. However, in the distribution of responsibilities within a particular practice, the consequentialist perspective resonates with the engineer's heuristic of "problem solving."

capacity of each agent to contribute to the shaping of technology. After all, within the consequentialist perspective, with its criterion of efficacy, responsibilities ought to be ascribed according to the capacity of each agent to discharge them. This is in line with the common intuition that having the capacity, power, and resources to contribute to the solution of a social problem entails a forward-looking responsibility to do so (Nihlén Fahlquist: 2009). For example in case of risky technologies, engineers, more than any stakeholder, have the knowledge of the risks and possible ways to reduce them. From the consequentialist perspective, this entails the responsibility to address these risks. This responsibility ascription, then, is not derived from a merit-based view in which particular actions are deemed faulty, but rather from the set of obligations to see to it that a certain state of affairs is brought about (i.e., a situation in which risks are prevented or at least addressed properly). This approach to ascribing responsibility fits nicely with the insights from more sociologically oriented literature on the dynamics of engineering and technology development.

However, efficacious as the consequentialist perspective may be, the fairness of the responsibility ascription cannot be ignored all together. This brings us to the other requirement of appropriateness: the question whether or not the responsibility ascription reflects people's reliable intuitions of when it is justified to ascribe a certain responsibility. It is unlikely that a purely consequentialist approach is psychologically feasible. The motivational force of responsibility ascriptions that are inconsistent with basic intuitions of fairness will therefore be undermined (Kutz: 2000, 129). This is in line with the point made in Section 2.3, where the relation between forward-looking and backward-looking responsibility was discussed. The motivational force to take up one's forward-looking responsibility is partly derived from expressions of praise and blame. The researcher in the GSBR project who judges his or her own responsibility within the project as fair will be motivated to act according to it, whereas the researcher that is assigned a responsibility unfairly may be inclined not to act according to it or to do it less carefully.[54] Moreover, from a moral point of view it

---

[54] One could think of the simple task of writing the minutes of a meeting. If it is decided by majority rule (but not consensus) that the same person should always take minutes of the meetings, this distribution of responsibilities is efficacious in the sense that for all meetings someone is ascribed the responsibility of writing the minutes. However, after some time, this person might become less motivated to accurately write down the minutes because he does not consider it fair that it is always him who should do the writing. However, if the person

is also undesirable to ascribe responsibilities in ways that contravene our basic feelings of fairness. Even if fairness is not the overriding criterion, we do not want a responsibility ascription that is morally unfair – both for the victims and for the people who are potentially blamed. Hence, even though fairness is not the ultimate criterion in the consequentialist perspective, it should still somehow be taken into account. Especially in case different people are involved there can be a tension between the requirement of efficacy and that of fairness. Whereas the fairness requirement is somewhat restrictive in ascribing responsibility, the efficacy requirement seems to have the opposite effect. It broadens rather than narrows the scope of responsibility ascriptions. If we focus on the fairness criterion, we probably end up with an ascription of responsibilities which is undesirable from a consequentialist perspective. If we only stress the efficacy of the responsibility ascription, we probably end up with an unfair distribution of responsibilities. Hence, we somehow have to incorporate both perspectives if we ascribe responsibilities.

A possible way to reduce the tension between the requirements of fairness and efficacy is to focus on alternative fairness criteria (i.e., criteria that are not related to the traditional substantive fairness criteria for individual responsibility). Insights from political philosophy show that fairness could also be achieved in a more procedural way. According to a procedural approach to fairness, a responsibility distribution can be rendered fair if it is established in a fair way (i.e., if it is the result of a fair procedure). Further research is needed to explore this procedural approach to fairness. A possible starting point may be the Rawlsian approach of Wide Reflective Equilibrium (WRE), according to which a procedure can be justified as fair if it fits within the individual set of background theories and moral principles of each relevant actor involved. The establishment of this procedural fairness could be part of an ethical parallel research (Doorn: 2010b; Van de Poel and Zwart: 2010). Questions as to which actors are relevant to include and how to assess such a WRE need to be further explored (Doorn: 2010a).

---

realizes that it is fair that he is given this task and that he will be blamed in case of sloppy minutes, he will most probably be motivated to come up with accurate minutes. More related to technology development, one could think of the responsibility related to the social impact or acceptance of the technology. If this is not recognized by the researchers as fairly being part of their work, it is questionable whether it will be addressed adequately, even if someone is explicitly given the task to look after the social impact.

The discussion above indicates an important role for the ethicist in the process of distributing responsibilities and identifying potential (negative) side-effects and consequences. The obvious question is then how this approach would work in case the technical work is not paralleled by an ethicist. I think we have to make a distinction between two situations. The first is one where a group of researchers have currently no ethicists embedded in or working parallel to their project but who have some experience with ethical research in previous projects. In this case the researchers have experienced how ethical research could be carried out. It is a challenge to sustain this "ethical attitude" in future projects. This is a challenge that somehow should be considered already during the ethical parallel research itself. The future will tell to what extent the impact of the past ethical parallel research will indeed lead to more permanent ethical reflection by the engineers themselves during their work. It goes without saying that the ethicists desire that their involvement is not just a passing phase and that they want an enduring impact on engineering practice. Further research into the different methods for doing ethical parallel research and possible ways to sustain its impact is therefore required.

The most common situation, however, is one where the research team has never been paralleled by a team of ethicists. How to make sure that ethical reflection is also incorporated in the work of these teams? Let me start by saying that there is a positive trend in requirements by funding organizations. It is nowadays often required to have a paragraph on ELSA (ethical, legal, and social aspects) in funding proposals. Although this attention for ELSA still runs the risk of being nothing more than "checkbox ethics," it points to a direction of more awareness for the social implications of technology. In addition to this requirement from funding organizations, (prospective) engineers should be trained in recognizing moral issues during their professional work. Engineering ethics should therefore be part of every engineering curriculum. Whether this will make the role of the ethicists completely replaceable is doubtful, but it will probably make engineers more prone to inviting ethicists in their project if they need their advice.

2.6. Conclusions

In this chapter, I discussed three "responsibility approaches" in the light of the development of a new technology. It was found that the merit-based perspective was rather powerless with respect to engineering practice because of the

problem of many hands. As a result, opportunities for learning and improvement were not optimally used. The rights-based perspective appeared to be most pessimistic about technology development. Due to its focus on monetary compensation, the effect of this approach on technology development was rather restrictive. Funding organizations and commercial partners would probably become reluctant to sponsoring innovative research. Moreover, it did not provide a strong incentive to improve the technology itself. The effect of the consequentialist perspective on engineering practice was most profound. This approach allowed for more fine-grained responsibility ascriptions and was found to fit nicely with insights from STS literature.

Although the consequentialist approach was found most powerful in co-shaping the direction of technology development, it was argued that the fairness requirement could not be ignored all together. It was shown that, for both moral and a consequentialist reasons, responsibility ascriptions should reflect our basic intuitions of when a particular responsibility ascription is justified. Since there is a potential tension between the traditional fairness criteria and the criterion of efficacy, it was proposed to conceive of fairness in a more procedural rather than substantive way, in order to reconcile the two demands of responsibility ascriptions.

3 Applying Rawlsian approaches to resolve ethical issues

Inventory and setting of a research agenda[55]

3.1. Introduction

Insights from social science are increasingly used in the field of applied ethics. Questions concerning the just distribution of scarce resources (medical ethics, business ethics), risks (ethics of technology), or multiculturalism (political philosophy) are a few examples of questions applied ethicists are confronted with nowadays and which are difficult to answer on the basis of traditional ethical theories alone. Accordingly, the 1980s onwards showed a "crossing-over" between philosophy and social sciences. Philosophers in general and practical ethicists in particular increasingly used insights from the social sciences and vice versa. This led to a more descriptive methodology by philosophers, reconstructing, for example, the social meanings of justice in a variety of social contexts instead of adhering to strict conceptual analysis, theory construction, and critical evaluation (Birnbacher: 1999). This so-called empirical turn in applied ethics contributed to a rise in development of "context-sensitive" methodologies. In a review of empirical ethical decision making literature in business ethics, it was found that the empirical turn did indeed contribute to knowledge on ethical decision making but that the field lacked theoretical grounding (O'Fallon and Butterfield: 2005). Recent insights show that ethical dilemmas often require a search for individual justification within a context of conflicting moral frameworks (Coughlan: 2005; d'Astous and Legendre: 2009; Van de Poel and Royakkers: 2007).

Philosophical approaches that have received ample attention in the field of applied ethics are the method of Wide Reflective Equilibrium (WRE), initially developed by Rawls (ToJ) and further elaborated by Daniels (1979; 1996), and

---

[55] This chapter originally appeared as an article in the *Journal of Business Ethics* 91 (1) 2010.

the related method of overlapping consensus (PL). In the remainder of this chapter, I will refer to either of these as "Rawlsian approaches." Rawls developed his method as a thought process with a justificatory purpose in theory development. He wanted to develop a theory that could account for the different moral background theories people hold while maintaining that they could morally agree on certain issues. Given the diversity of moral perspectives in our pluralist society and the need to take these somehow into account, it is not surprising that the attractiveness of Rawls' method has not been limited to merely (theoretical) political philosophy. Especially in the field of applied ethics, the struggle between diverging moral frameworks is an urgent problem – for example, how to integrate the interests of multiple stakeholders (Daboub and Calton: 2002; Jamali: 2008); how to do justice to different cultural and moral traditions in international business (Ho: 2003); or how to justly distribute risks stemming from new technologies (Zandvoort: 2008)? Rawlsian approaches are attractive for answering these kinds of questions because they do not take one of the extreme positions of giving authority to either moral theory or the empirical data. Instead, moral theory and empirical data are integrated in order to reach a normative conclusion with respect to moral practice (Molewijk et al.: 2004).[56] Moreover, whereas many empirical approaches seem to lack theoretical grounding, Rawlsian approaches are well supported by theory. In the field of business ethics, the approach is promising because it allows for decision making in a pluralist context with different stakeholders, without giving a priori priority to any of them.[57] The latter is important to gain support for the decision made.

Although Rawlsian approaches have attracted ample attention – they are often recommended in scholarly literature as a proposed way out of an ethical impasse or as a fruitful approach to do further research on (cf. Brand-Ballard: 2003; Gracia: 1995; Van de Poel and Royakkers: 2007) – the actual application of

---

[56] For a description of other approaches see (Molewijk et al.: 2004). For elaborate description of the main points of criticism on the use of empirical data in ethics the reader is referred to Elster (1992), Schmidt (1994), Birnbacher (1999).

[57] However, as one of the reviewers of the journal version of this chapter correctly pointed out, ultimately a decision proposed by one of the stakeholders may be chosen as the most favorable. In that sense, one could argue that in fact priority is given to that particular stakeholder. However, this is justified only if the decision fits within each individual's WRE. The priority is then *a posteriori*, after the deliberative process of constructing WRE.

these approaches is still relatively rare. Consequently, insight in the actual "performance" and the potential obstacles for application is lacking.

The objective of this chapter is twofold. First, by making an inventory of recent applications of Rawlsian approaches in applied ethics this chapter aims at learning how the Rawlsian approaches are applied. To this end, a taxonomy of different kinds of applications will be given, classified according to their purpose, content, and type of justification. Secondly, these applications will be studied in more detail to investigate what kind of (practical) obstacles are encountered when applying Rawlsian approaches. On the basis of the results, recommendations are given together with an identification of potential lacuna in the research.

This chapter proceeds as follows. First the theoretical concepts of WRE and overlapping consensus are discussed, followed by a taxonomy of different type of applications. On the basis of the outlined taxonomy different applications are discussed. To conclude, recommendations for further application of Rawlsian approaches are given.

3.2. From wide reflective equilibrium to overlapping consensus

Rawls developed the method of wide reflective equilibrium for explicating and defending his theory of justice. Rawls tried to develop a criterion of justice that would be agreed upon by all under conditions that are fair to all (JaF 15). For that he needed a point of view that is removed from and not distorted by the particular features and circumstances of the existing basic structure of society. Rawls therefore introduced the so-called "original position." Starting from this hypothetical situation, representatives of citizens are placed behind a veil of ignorance, depriving them of information about the individuating characteristics of the citizens they represent, in order to let them reflect upon and after deliberation agree upon a principle of justice that would be acceptable to all, regardless of those individuating characteristics of the citizens. This leads to the conception of justice as fairness, specifying the fair terms of social cooperation between free and equal citizens (ToJ 15; PL 22–23; JaF 18).

In addition to this "pure procedural justice" (PL 72–73), Rawls developed a justification criterion to assess whether the hypothetical contract situation articulates the considered convictions of political justice of individual citizens. Individuals must be able to accept the agreement reached in the original position if it can be embedded in their individual comprehensive doctrine (RH 143), that

is, if it coheres with their own firm convictions of justice or extend them in an acceptable way (ToJ 17, PL 28). People with different comprehensive doctrines must be able to justify for themselves the acceptability of the claims of political justice. Rawls introduced the idea of reflective equilibrium to refer to this individual justification. In this idea, a distinction is made between three levels of considerations:

(1)   considered moral judgments about particular cases or situations;
(2)   moral principles; and
(3)   descriptive and normative background theories.

Assuming that all people want to arrive at a conception of justice that yields definite solutions and that is complete, in the sense that it is more than a mere collection of accidental convictions, people should aim at coherence between the considerations at the different levels. By moving back and forth between different levels of considerations and revising the considerations, principles, and theories that do not fit well, people arrive at a reflective equilibrium. We speak of an *equilibrium* if the different types of considerations cohere and are mutually supportive; it is called *reflective* if the equilibrium is arrived at by working back and forth between the different considerations and all 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 (three-tiered view) and not only the considered judgments and moral principles (which is the case if we speak of a *narrow* reflective equilibrium; twotiered view). Rawls stated that "the original position serves as a mediating idea by which all are considered convictions, whatever their level of generality...can be brought to bear on one another" (PL 26).

As already said, Rawls developed his idea of reflective equilibrium in the domain of political philosophy, as a method for deriving a theory of justice. At first, Rawls defined the reflective equilibrium in the narrow sense; a reflective equilibrium that arises through reflection on merely one's own prior convictions. However, Norman Daniels has argued that any narrow reflective equilibrium is difficult to accept, because it leaves us with the traditional two-tiered view of moral theories and is therefore particularly ill-suited to providing a basis for justification (Daniels: 1979). After all, coherence between our considered judgments and principles that provide generalization does not rule out the possibility of an arbitrarily prejudiced view point (Blackburn: 1993; Singer:

1974). The problem lies in the lack of evidence for the reliability of considered judgments independent of the principles and other theoretical elements they are supposed to support and manifest (Sencerz: 1986; Van der Burg and Van Willigenburg: 1998). By focussing solely on particular cases and moral principles, the reflective equilibrium that is arrived at is based on fixed (moral) background theories. A narrow reflective equilibrium might therefore be characterized as typically utilitarian or Kantian. Such a narrow reflective equilibrium may seem to be a descriptive rather than a justificatory method. To give the method justificatory credibility also, Daniels proposed seeking coherence between the widest set of moral and non-moral beliefs and also including the level of background theories in the process of reflection. In order to provide evidence for the credibility of a set of moral principles and the moral conception they embody, all three levels should be scrutinized and open for revision. The result of this three-tiered criterion of justice is then wide reflective equilibrium, which also incorporates the level of background theories. In his later work, Rawls adopted this wide conception of reflection since this allows for taking into account moral conceptions advanced by others, thereby giving these the chance to influence one's own convictions (CP 289–290). Given the objective of Rawls' reflective equilibrium – the development of a concept of justice – the incorporation of moral convictions endorsed by others is part of its justification.[58]

The original idea of justice as fairness was based on a well-ordered society, which is relatively homogeneous in its basic moral beliefs and conceptions of what constitutes the good life. In his later work, *Political Liberalism*, Rawls revised this idea of a well-ordered society. Recognizing the permanent plurality of incompatible and irreconcilable moral frameworks within a democratic society, he introduced the concept of overlapping consensus. People are able to

---

[58] Rawls seems somewhat ambiguous at this point. On the one hand, WRE serves as an individual justification criterion. As such the inclusion of convictions held by others is part of systematizing one's own relevant convictions into "one unified, plausible conception of justice" (Pogge: 2007, 166-167). Part of this coherence is to pay attention to the thoughts of others, "whose intelligence, integrity and life experience I have learned to respect" (167). This aim of WRE fits well with Rawls' view on citizens as reasonable persons. On the other hand, confrontation with others forms an important motive for seeking WRE in the first place. As such, WRE is sought "not merely as a guide for conduct but also to show others that I am genuinely concerned with matters of justice and hence willing to restrain my conduct in accordance with firmly held and enduring principles" (ibid.).

live together despite conflicting moral values and ideals as long as people share a moral commitment to society's basic structure. The complete idea of justice as fairness will most probably not be part of a wide reflective equilibrium but in a plural society it can still be endorsed by adherents of reasonable comprehensive doctrines as a political conception of justice, that is, as a basis of social unity in a constitutional democracy with a plurality of reasonable but incompatible – religious, philosophical and moral – doctrines. People with divergent comprehensive doctrines can overlap in their acceptance of a conception of justice. They do not have to agree on everything but they do agree on "principles of fairness" related to the political realm. Being the focus of an overlapping consensus, these principles specify the fair terms of cooperation among citizens and the conditions under which a society's basic institutions can be deemed just (PL 133). When all citizens recognize that they affirm the same public conception of political justice (i.e. the conception of political justice is affirmed in their own considered judgments), reflective equilibrium is not only wide but also general, by Rawls referred to as full reflective equilibrium (JaF 31).[59] Because the principles are part of an overlapping consensus and not the result of a negotiation leading to some compromise, the result is more stable. With the shift from reflective equilibrium towards overlapping consensus the emphasis is also shifted away from the original position towards the distinction between public and non-public reason.[60] Public reason asks us "to conduct our fundamental discussions in terms of what we regard as a political conception" (PL 241), providing as such a *pro tanto* justification of political values as laid down in the political conception of justice. Rawls argued that a reasonable overlapping consensus with respect to the political conception of justice (i.e. justice as fairness) is complete in the sense that "the political values specified by

---

[59]   Note that this is not to say that this general reflective equilibrium is *shared* by all. It is only with regard to the political conception of justice as fairness that the general reflective equilibrium is a *shared* wide reflective equilibrium.

[60]   Rawls provides a threefold meaning for public reason: "it is the reason of the public; its subject is the good of the public and matters of fundamental justice; and its nature and content is public, being given by the ideals and principles expressed by society's conception of political justice" (PL 213). He explicitly rejects the distinction public–private in the context of political liberalism, because there is no such thing as private reason involved. Besides public reason there is social reason, concerned with associations in society, and there is domestic reason, concerned with families as small groups in society. Citizens participate in all these kinds of reason (PL 220; fn 7).

it … give a reasonable answer by public reason to all, or nearly all, questions concerning constitutional essentials and basic justice" (JaF 142). This *pro tanto* justification of the political conception of justice is done "without looking to, or trying to fit, or even knowing what are, the existing comprehensive doctrines" (RH 145). An individual citizen can then try to fit this political conception of justice into his own comprehensive doctrine. This is what Rawls calls full justification, which is carried out by an individual citizen as a member of civil society and in which the citizen accepts a political conception and fills out its justification by embedding it in his own comprehensive doctrine. Public justification by political society, as a third kind of justification, happens "when all the reasonable members of political society carry out justification of the shared political conception by embedding it in their several reasonable comprehensive views." In the latter case, "reasonable citizens take one another into account as having reasonable comprehensive doctrines that endorse that political conception" (RH 142-143).

3.3. A taxonomy of Rawlsian approaches

Partly because of the significant revision in Rawls' theory the term "Rawlsian approaches" does allow for a rather diverse range of applications. They can vary in the purpose they serve, their content, and the kind of justification, if any, they provide. In this section, I present a taxonomy of the different type of applications.[61]

Purpose of the method
Rawls' method, although derived with a justificatory purpose, is sometimes used in a descriptive or even constructive way. In the latter case the method is used as a framework for structuring discussion and debate, with the aim of coming to a justified agreement. The method could then be used, for example, as a means to

---

[61] A word of caution is due here. With the introduction of Rawlsian approaches to the field of applied ethics, we extend the application domain beyond the political. Moreover, for Rawls the concepts were used mainly hypothetically with the purpose of developing a theory of justice. In applied ethics actual decision problems are at stake, which calls for other than merely hypothetical considerations. This necessarily implies that different versions of the original method are being used, depending on the issue at stake. A certain pragmatic stance is therefore required in using Rawlsian approaches (Van der Burg and Van Willigenburg: 1998).

attain a coherent basis for decision making by ethical committees or to gain support for particular decisions in the context of public policy (Holmgren: 1987). In the case of descriptive use, the explanatory power of the method is dependent on epistemic and cognitive considerations, for example on how people form and revise their considered judgments, moral principles, and background theories. These epistemic and cognitive considerations are less important if the model is applied with a constructive purpose. In that case the method serves a practical purpose and its effectiveness can be judged on practical grounds.

Content of the method

Besides the threefold purpose (justificatory, descriptive, and constructive), a second important distinction lies in the content of what is to be derived by the method. With his idea of the original position, Rawls developed a universal procedure for reaching moral agreement. Whatever the outcome of this procedure is, the result is fair because it is the result of a just procedure specified by (representatives of) the citizens themselves (TJ 104, PL 72). This is called pure procedural justice: what is just is specified by the outcome of the procedure, whatever that is. As Rawls explains, "there is no prior and already given criterion against which the outcome is to be checked" (PL 73).

However, in his *Theory of Justice* Rawls goes further. He does not only derive procedural principles of fairness, he also gives substantive content to the principles of justice (equal basic rights and liberties, fair equality of opportunities, difference principle).[62] In the application of Rawlsian approaches, this distinction is often overlooked. The concepts derived from the conception of procedural justice on the one hand, and those derived from the justice principles on the other, are used almost interchangeably. However, the two are fundamentally different. Whereas the former (Rawls' concept of procedural justice), represents a (thin) procedural or political view describing under what conditions a procedure is fair, the latter (Rawls' principles of justice) represents a thick comprehensive view referring to the fairness of the outcome of this

---

[62] Note that this is exactly the point where Rawls has attracted most criticism. Opponents argue that it is intrinsically impossible to fairly derive procedurally neutral principles for the political domain and at the same time endorsing a comprehensive view of how those principles are to be substantiated. There seems to be an unbridgeable gap between a political doctrine of liberalism (i.e. to be tolerant towards non-liberal comprehensive doctrines) and endorsing political liberalism as a comprehensive doctrine at the same time (cf. Tan: 1998).

procedure.[63] Hence, the former provides a procedure for deriving a fair or just outcome; whatever the outcome, it will be deemed just because it is derived from a fair procedure. In case of the latter, an independent criterion is given marking what outcome of the procedure can be judged fair or just (ToJ 75).

Kinds of justification[64]

As explained in Section 3.2, Rawls makes a distinction between three kinds of justification. The first is political justification, which is a *pro tanto* justification of the political conception, taking into account political values only. This is the kind of justification done in the original position by representatives of the citizens.

The second type of justification, full justification, deals explicitly with citizens within their own life and with their own comprehensive doctrine. The question here is whether they can accept the political conception and embed it in their own comprehensive doctrine. The type of equilibrium that is at stake here is individual WRE.

Public justification, as a third type, is justification by political society. Crucial here is that citizens "do not look into the content of others' doctrines, and so remain within the bounds of the political. Rather, they take into account and give some weight to only the fact – the existence – of the reasonable overlapping consensus itself" (RH 144). In this case, the shared political conception is the ground of the justification. Since all people are in WRE, there is a general reflective equilibrium also. However, each person's WRE is an individual one: the general reflective equilibrium does not coincide with the individual WREs. It is only insofar as the political conception of justice is concerned that the general reflective equilibrium and the individual WREs overlap.

In the next section, several applications of Rawlsian approaches are discussed and classified according to the above-mentioned content and purpose distinctions. In case the purpose was justificatory, the type of justification will also be mentioned.

---

[63] The thin–thick distinction is often made with regard to liberalism. A thin conception of liberalism is primarily aimed at tolerance towards other comprehensive doctrines. No single comprehensive doctrine can have the ultimate authority in political debate. Thick liberalism is itself a comprehensive doctrine with a particular view on the good life in which the values of autonomy and individuality extend into most areas of life.

[64] For the sake of analytical clarity the way the three kinds of justification are approached is slightly different from Rawls' approach.

3.4.   Inventory of recent applications of Rawlsian approaches

Recent applications of Rawlsian approaches to actual problems have been identified by searching the ISI Web of Knowledge databases (Web of Science, ISI Proceedings, including the Science, the Social Science, and the Arts and Humanities citation indices) and the SilverPlatter Philosopher's Index. A search on the basis of the criteria TS = ("wide reflective equilibrium" OR "overlapping consensus") and PY = (1998–2007) was carried out, where TS refers to the topic search (either of these entries are found in the title, the keywords, or the abstract of an article) and PY refers to the time span within which these articles appeared.[65]

The search resulted in a list of 68 different articles, of which 12 covered an empirical or hypothetical application of the method of WRE or overlapping consensus, by which I mean that Rawlsian approaches were explicitly used to elucidate a certain moral problem or justify certain solutions for a particular moral problem. In 31 articles Rawls' theory was discussed or criticized without applying it and 16 articles referred to something other than Rawls' method of WRE or overlapping consensus.[66] Nine articles were left out because they were not written in English.

Table 3.1 shows the list of 12 applications including the purpose, the content, and the type of justification. A short description of each of the applications is given in the appendix at the end of this chapter. The fifth column of the table shows whether the method of WRE or overlapping consensus (OC) was used in the paper. Of the 12 cases listed, only two applications did not cover the political domain. These were also the only applications where the method of WRE was used instead of overlapping consensus. In one case, the connection was made between WRE and overlapping consensus (March: 2006).[67] The remaining nine articles (more or less) covered the political domain and applied the method of

---

[65] Other sources for papers on WRE/overlapping consensus applications are available but the objective of this research is not give a complete inventory of all applications of WRE/overlapping consensus. Rather, the objective is to obtain a random sample of applications of Rawlsian approaches.

[66] The term "overlapping consensus" seemed to be a microbiological term as well, referring to a particular genetic phenomenon.

[67] March did in fact not refer to wide reflective equilibrium but to reflective equilibrium, leaving the question whether this is a wide or narrow equilibrium open.

Table 3.1: Classification of cases in terms of purpose, content, and type of justification

Case	Purpose	Content	Type of justification[a]	WRE or OC
Same-sex marriage (McClain: 1998)	Justificatory	Procedural	Public Justification	OC
The status of the EU and international law (La Torre: 1999)	Justificatory	Procedural	Public Justification	OC
Curriculum reform (Halliday: 1999)	Constructive	Substantive	-	OC
Liberal education (Paris and Kimball: 2000)	Descriptive	Substantive	-	OC
European citizenship (Lehning: 2001)	Constructive	Procedural	-	OC
Pediatric cochlear implants (Reuzel et al.: 2001)	Justificatory/ Constructive	Procedural	Public justification	WRE
International Business vs. Globalization (Ho: 2003)	Constructive / Descriptive	Procedural	-	OC
Environmental values (Preston: 2004)	Descriptive	Procedural	-	OC
Human rights (Tobin: 2005)	Justificatory/ Constructive	Substantive	Public justification	OC
Muslim minorities (March: 2006)	Justificatory/ Constructive	Substantive	Public justification	OC/ WRE
Abortion (Moran: 2006)	Justificatory	Procedural	Public justification	OC
Technical and technological innovations in sport – seeding rules in tennis (Sheridan: 2007)	Justificatory/ Descriptive	Substantive	Political justification	WRE

[a] only relevant when applied with justificatory purpose

overlapping consensus. With respect to the dimensions "purpose" and "content," all possible combinations were covered at least once. The most prevalent combination was that of justificatory purpose and procedural content, which is similar to the way the "late" Rawls (i.e., Rawls in *Political Liberalism*) applied his method.

From the inventory it follows that applied ethicists are rather pragmatic in applying Rawlsian methods. In none of the papers was an attempt made to simulate something like the original position in order to make the procedure "more fair" and increase the justificatory power. Instead the justificatory purpose was sought by taking into account the right people (viz., Ho: 2003; La Torre: 1999; Reuzel et al.: 2001) or by stressing the kind of arguments that people are allowed to use (viz. McClain: 1998; Moran: 2006), or a combination of the two (March: 2006; Twiss: 2004). This is in line with the fact that most scholars refer to public justification instead of political justification. The former does, strictly speaking, not require that people are ignorant of their personal life and interests.

In seven cases, the Rawlsian method had (at least) a justificatory purpose. In four cases the justificatory purpose was combined with a constructive or descriptive purpose. In two cases the purpose was purely descriptive (liberal education and environmental values) and in two cases it was purely constructive (curriculum reform and European citizenship). In none of these cases was full justification at stake. Political justification was at stake in only one case (seeding rules in tennis).

In his elaboration of WRE, Rawls makes a distinction between background theories, moral principles, and considered judgments. In the two papers that covered WRE, the agreement that was sought for was a method for evaluation. Although neither of the authors explicitly mentioned on what level this agreement was to be found, in their papers they indirectly addressed the point by defining agreement as "a temporary consensus on a shared construction of the technology under scrutiny ..." (Reuzel et al.: 2001, 252) and looking for agreement "about what they [i.e., the decision makers in tennis, ND] consider to be the best interpretation [...] of what a fair seeding structure looks like" (Sheridan: 2007, 186). The similarity between Reuzel et al.'s construction and Sheridan's interpretation is striking and it suggests that the agreement cannot be framed solely in terms of one of the three levels. It is rather a certain meaning attributed to the thing being evaluated, in these cases the cochlear implant and the seeding rules. The question of how to include background theories was not discussed in detail. In the article by Reuzel et al., this issue was not dealt with at

all; in the article by Sheridan the background theories were described only on a very abstract level ("a theory of the person, a theory of procedural justice, a general social theory, and a theory of the role of morality in society;" Sheridan: 2007, 184). In case of overlapping consensus, the agreement was generally on the level of principles or values: liberal and democratic principles (Lehning: 2001; March: 2006; McClain: 1998; Moran: 2006), principles of Community Law (La Torre: 1999), values in education (Halliday: 1999; Paris and Kimball: 2000), values in economics (Ho: 2003), and environmental values (Preston: 2004). In the paper on human rights (Tobin: 2005), the possibility of distinguishing between the three levels was questioned. Tobin argues that when these moral principles (viz. human rights) are made substantive, "premises or assumptions that attach to the moral-social world" (viz. considered judgments) and the "comprehensive doctrines that shape these world" (viz. background theories) inevitably are built in (Tobin: 2005, 38), thereby rejecting the Rawlsian analytical distinction. It remains to be seen to what extent this really conflicts with Rawls' own view. More important than the impossibility of distinguishing between the three levels is possibly the strong focus on moral principles. Rawls explicitly argues that his idea of wide reflective equilibrium supposes that neither background theories, nor moral principles, or considered judgments have priority over one another. Considered judgments can revise moral principles and background theories as well as vice versa. It is therefore questionable whether agreement on the level of principles is sufficient to settle moral disputes. As long as considered judgments and/or background theories significantly differ, the agreement may not be as stable as the authors suggest.

3.5.   Discussion

In six out of 12 papers, the application of Rawlsian approaches was judged positively (Ho: 2003; La Torre: 1999; March: 2006; McClain: 1998; Moran: 2006; Reuzel et al.: 2001). According to those authors, Rawlsian approaches are helpful in explicating conditions for justification (La Torre: 1999; March: 2006; McClain: 1998; Reuzel et al.: 2001), and can serve as a goal of political change (Ho: 2003) or cultural transformation (Moran: 2006). Two authors took a more or less neutral stance (Lehning: 2001; Preston: 2004) and in four articles the authors took a more critical stance towards the application of Rawlsian approaches (Halliday: 1999; Paris and Kimball: 2000; Sheridan: 2007; Tobin:

2005). In what follows, the most important points of the different articles are briefly discussed, followed by an elaboration of the main points of criticism.

McClain (1998) discusses Rawls' overlapping consensus in the light of the question what type of arguments citizens may appeal to or what kind of reasons they may proffer to support their positions. A similar application is found in La Torre (1999). Both authors use the notion of overlapping consensus to define a criterion by which to settle disputes. McClain stresses the importance of critical examination of a supposed overlapping consensus in terms of the arguments used. La Torre emphasizes the "open ended" discursive process that should lead to the overlapping consensus. The use of Rawls' method of WRE in the article by Reuzel et al. (2001) is also similar to that of McClain. Reuzel et al. discuss the issue of feasibility and justification of interactive technology assessment (iTA) within the context of decision making on medical technologies. Reuzel et al. do not apply the method of WRE but propose it as a way to come to a justified agreement on decision making in a medical context. Similar to McClain, the authors give criteria for a reflective equilibrium. Not any agreement or equilibrium will do: a WRE is only morally justified if the considerations of all relevant actors are involved, which we can call an "inclusiveness criterion."

This question of inclusiveness is also addressed in March (2006). Where the three previous papers all described a procedural content, March uses Rawls' method to justify substantive arguments. The author argues that an Islamic doctrine of citizenship in non-Muslim liberal democracies can be said to reflect equilibrium when it is as inclusive as possible of believing Muslims without violating any essential features or aims of a well-ordered society. Besides the justification of substantive arguments, the author applies Rawls' method with a constructive purpose to stimulate discussion between different people. This constructive purpose is also found in the papers by Ho (2003) and Moran (2006). Both authors use concepts of Rawls' theory to guide political and cultural change.

Lehning (2001) and Preston (2004), who both take a more or less neutral stance towards the use of Rawlsian approaches, use the Rawlsian notions rather loosely without explicating why and how these notions were helpful. Lehning claims that an overlapping consensus that results in a political conception of justice, shared throughout a political community, does generate a shared identity that will supersede rival identities based on, for instance, ethnicity. Preston tries to extrapolate principles of Rawls' theory of justice to the field of environmental ethics. Although Rawls himself does not consider environmental issues to be

part of the public good – and therefore not part of fundamental political justice – it is desirable that environmentalists appeal to public reason to support their claim for environmentally friendly policies, Preston argues. In his article, environmental issues are the stepping stone to exploring the limits of fundamental political justice. The application of Rawlsian notions sheds some light on the topic of environmental ethics but is does not so in depth.

In the remaining four articles, the authors take a more critical stance towards the application of Rawlsian approaches. Tobin (2005) rejects the claim, held by many contemporary human rights theorists, that the universality of human rights can be established by appeal to and grounded on the notion of overlapping consensus because the various belief systems on which justification of human rights are based are too divergent to provide "the kind of meaningful, substantive agreement required … as action-guiding norms for cross-cultural moral evaluation and critique" (38). Instead of grounding universality on overlapping consensus she proposes to "construct universality through actual dialogue both within and between communities" (33).

The papers by Paris and Kimball (2000) and Halliday (1999) also question the feasibility of Rawls' consensus. Paris and Kimball (2000) try to give a description of the course of liberal education in the US in the twentieth century. Based on previous work by Kimball (1995a; 1995b), in which this course was described and explained in terms of Rawls' notion of overlapping consensus, the authors revise Rawls' concept of overlapping consensus into a more pragmatic consensus which is less "thick" than Rawls' version. They thereby refer to Rawls' two principles of justice. The authors question whether these principles are as broadly supported as Rawls suggest. They therefore try to limit themselves to as thin a consensus as possible. This also allows for substantive changes over time if these are required. A similar course is followed by Halliday (1999). He discusses recent proposals to introduce citizenship and values education in the UK. Although widely welcomed, the values underlying this curriculum reform are open to debate. Halliday argues for a Wittgensteinian reinterpretation of overlapping consensus with regard to the political values in order to arrive at "a conception of citizenship to which all can happily assent" (47). The consensus consists not so much in doctrines as in beliefs about what ought to be done in particular circumstances. The curriculum reform should not be a State affair but rather something which is done inside the locus of community.

This emphasis on community links up to the criticism raised by Sheridan (2007), who explores the use of WRE as a possible decision making method for

the rational evaluation of technical and technological innovations in sporting practices. Her paper describes the implementation of a new seeding system in male tennis at the Wimbledon championships. The WRE method is not used to justify the new seeding rules but only to hypothetically describe how the process of coming to these new rules might have developed. She argues that it is precisely in the requirement that people have to detach themselves from their own personal interests, that the method becomes too far removed from the persons' particular practices.[68] Sheridan therefore rejects the method for lacking both descriptive and justificatory power.

Three main obstacles that are exemplified by the papers deserve closer attention. The first is the issue of inclusiveness (March: 2006; Reuzel et al.: 2001), the second is the requirement that people have to detach themselves from their personal practices and interests (Sheridan: 2007), and the third is the issue of feasibility (Halliday: 1999; Paris and Kimball: 2000; Tobin: 2005). The remainder of this section focuses on these three issues in more detail. Discussion of the feasibility issue will show that more empirical research is needed, in particular into the question how the equilibrium or consensus is established. In the present discussions on Rawlsian approaches this issue does not seem to get the attention it deserves, as I will explain below. This section therefore concludes with recommendations for further exploration of the Rawlsian justification in practice.

Inclusiveness

The issue of inclusiveness is related to the justness of the method. In the literature on Rawls' theory it is recognized that the method has some excluding mechanism in it because it neglects the arguments of unreasonable people who

---

[68] Although practice is not part of Rawls' own vocabulary, it is a notion often used by communitarians to refer to the MacIntyrian idea of socially-established cooperative human activity. Because these practices are constitutive for the good life, it would be wrong to demand that people detach themselves from a practice and the motives, considerations, and interests that follow from being part of it. The full definition of MacIntyrian practice reads as follows: a practice is "any coherent and complex form of socially established co-operative human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence which are appropriate to, and partially definitive of, that form of activity, with the result that human powers to achieve excellence, and human conceptions of the ends and goods involved, are systematically extended" (MacIntyre: 1984[1981], 187).

refuse to accept a sufficient critical mass of common considered judgments, that is, people who do not already have some inclination towards or belief in morality (Norman: 1998). The people Norman seems to refer to are the people on the extreme end of the reasonable–unreasonable spectrum, where the problem seems to be more of a hypothetical problem. The issue becomes more relevant as we approach the middle area of this spectrum, where it is less obvious whether or not something can count as "reasonable." This is even more true if we use it as a practical method. Methodological considerations regarding the choice about which types of beliefs and arguments to include are inevitably selective. As such the method cannot fulfill its promise of strict absence of arbitrariness (Van der Burg and Van Willigenburg: 1998).[69]

The criterion of inclusiveness has a second facet, which is openness. When using the method in practice the criterion of inclusiveness calls for an open discourse. In Reuzel et al. (2001), for example, it is not only important that all relevant actors are included, but that they also have equal opportunities to participate in and contribute to the decision making process. If a group of researchers is engaged in a conversation with doctors and patients it is important that the vocabulary used by the experts is understandable to all. The criterion of inclusiveness also requires that people feel free to introduce unwelcome arguments. If patients are discouraged from doing so and remain silent, the consensus that is arrived at cannot be deemed just.

Inclusiveness, understood this way, might conflict with the requirement that people have to deliberate in terms of public reason. After all, public reason requires people not to deliberate in terms of their comprehensive doctrines. Inclusiveness, on the contrary, requires that people are allowed to bring in unwelcome arguments that are important for them. The latter might be based on their private, comprehensive doctrines. In this regard it is important to distinguish between the different kinds of justification. In case of *pro tanto* justification, the requirement of public reason is an important one because it guarantees the "impartiality" of the justification. However, in cases of public or political justification, people's comprehensive doctrines come into play because

---

[69] Van der Burg and Van Willigenburg use the word bias instead of arbitrariness. I prefer the latter since bias indicates a systematic selectivity, which does not necessarily have to be the case. The point is that every selection implies a line to be drawn between what to include and what not. The drawing of such a line is intrinsically selective. This selectivity may, but does not necessarily have to be, biased also.

justification is done by embedding it in the individual people's comprehensive doctrines. Individual WRE is always about political or public justification and, consequently, when applying the method of WRE the criterion of inclusiveness "overrules" that of public reason or detachment. When applying WRE or overlapping consensus as a practical method, the different kinds of justification should therefore be carefully distinguished. One should bear in mind that the type of justification determines whether priority should be given to either the criterion of public reason/detachment or to that of inclusiveness.

Detachment

The point of criticism raised by Sheridan is related to the communitarian criticism which says that the method is mistakenly aimed at people being detached from their personal life (Walzer: 1983). To justly derive principles of fairness, people are put behind a veil of ignorance, such that they are deprived of information about their position in society. However, disregarding people's context results in a "shallow" method with which people are not able to decide on important matters concerning their own life because the method becomes too far removed from the individuals' particular practices and interests, the communitarian criticism goes. People are embedded in their social environments and as such their rational morality is constituted in their community. These practices define what the moral goods are people should strive at (MacIntyre: 1984[1981]).

In the example of the seeding rules in tennis (Sheridan: 2007), this point of criticism was especially relevant. This can be explained by the fact that the problem at stake is a very "specialized" issue within a certain community with its own particular norms and values. The disagreement among the different decision makers is probably not on the level of the background theories and the (moral) principles. Probably everyone agrees on the fact that the winner should preferably be the best player or the best team, hence luck should not be a decisive factor. Questions regarding how to exactly determine the winner and how to exclude luck are lower-level considerations. Hence, a narrow reflective equilibrium will probably be sufficient in those cases. However, in situations where the normative and descriptive background theories do indeed differ significantly, people will have to distance themselves from their practices, at least hypothetically. Otherwise it will be impossible to agree upon decisions. Within a certain sport practice, this will generally not be the case.

The example of the seeding rules in tennis shows that the communitarian criticism is not a knockdown argument against the application of Rawlsian approaches. Rather it shows where the application can be fruitful and where it probably cannot. Rawls developed his theory for decision making in a situation with a plurality of moral frameworks. The types of questions that are at stake in such a situation differ from those in a situation that can be characterized as relatively homogeneous. Compare, for example, a discussion in an ethics committee on the allowance of abstaining from medical treatment and a discussion within a particular sports club on the selection procedure for the best team. In the former case, the interests of the patient, his relatives, and the medical doctors can significantly diverge. In the latter example a Rawlsian approach is less powerful indeed, precisely because the people of the sports club are part of the same community and to a large extent share their relevant (moral) background theories. But one could consider this also a lack of demand for Rawlsian approaches: in cases where people to a large extent share their background theories they will probably not need a systematic approach for reaching moral agreement.

Feasibility

Given the fact that all applications deal with hypothetical cases, the issue of feasibility is difficult to discuss. The authors using WRE or overlapping consensus in a constructive way emphasize that overlapping consensus is a process rather than a state of affairs (cf. the plea for dialogue Tobin: 2005) and discussion March: 2006). The way this process is shaped to a large extent determines the success of its outcome. Moreover, the type of questions that are addressed also determines the success of the application. Reuzel et al. (2001) argue that iTA (viz. WRE) is most suitable as a pre-assessment, generating two kinds of result:

(1) insight into the conditions under which a certain technology, or particular decisions with respect to the technology, is acceptable to all persons involved; and

(2) a set of research questions, considered relevant and feasible by all persons involved, having to be answered in order to know to what extent these conditions are, or can be, met.

This suggests that WRE and overlapping consensus, in order to be successful, requires a focus. To reach agreement about a technology in general may be too ambitious. However, if the focus is on specific changes, consequences, or implications, the approach might be more fruitful (Reuzel et al.: 2001, 252). This focus on conditions is in line with Paris and Kimball's argument that people are more likely to agree on thin (procedural) than on thick (substantive) notions of justice.

On the basis of hypothetical cases alone, it is premature to reject or embrace overlapping consensus or reflective equilibrium in terms of feasibility. Neither the objections raised by critics, nor the recommendations given by Reuzel et al. give a decisive answer to the question whether WRE or overlapping consensus are attainable. To answer this question, empirical research is needed.

Justification in practice

One of the striking things in the selection of articles surveyed is the lack of actor involvement in the judgment whether a reflective equilibrium or an overlapping consensus has actually been achieved. Only in the proposal by Reuzel et al. (2001) was it explicitly suggested the actors be involved in assessment of the equilibrium. In the other cases, it was either left to the researcher to assess whether a reflective equilibrium or overlapping consensus had been achieved or it was left open. This can partly be explained by the fact that all cases were more or less hypothetical. However, if the method is to be used with a justificatory or constructive purpose, it will significantly gain cogency if the actual actors are directly involved in this judgment. After all, the question whether an agreement is part of an overlapping consensus or merely a matter of *modus vivendi* or whether someone's considered judgments, moral principles, and background theories are in reflective equilibrium can best be answered by the actor who has "direct access" to these considerations. This more empirical consideration seems to have attracted scarce attention in the literature. Musschenga (2005) indirectly addresses this point when he criticizes the feasibility of Rawlsian methods. He argues that reflective equilibrium is not an objective state of affairs that can be determined from a third-person point of view; it usually is a first-person judgment. For Musschenga this is a reason to doubt the possibility of finding a standpoint that is in equilibrium with the possibly diverse beliefs, principles, and background theories of all the members. He seems to worry more about the seeming difficulty of arriving at equilibrium than about the question of first-person versus third-person assessment of the equilibrium.

In real applications, as opposed to hypothetical cases, the possibility of uncovering the relevant data (i.e., the background theories, moral principles, and considered judgments) deserves close attention. It is important to investigate whether these data are indeed traceable. This is an empirical question that has a high priority if one wants to explore the possibility of applying Rawlsian methods in applied ethics. Only if these data are available, it makes sense to construct or assess an equilibrium or consensus. A first step seems to be to apply the method to a real case and to ask the different actors involved about their background theories, moral principles, and considered judgments and to ask them whether they themselves conceive of the outcome as justified (within a reflective equilibrium or overlapping consensus). Of course, such a first exploration may seem slightly artificial with actors being informed about wide reflective equilibrium and the difference between overlapping consensus and *modus vivendi*. However, if this litmus test succeeds, it opens the door to less artificial applications.

3.6.　Conclusions

In this chapter, an inventory is made of 12 applications of Rawlsian approaches. Although sometimes labeled as promising within the field of applied ethics, it was found that Rawlsian approaches are still mainly applied within the political domain (i.e., they refer to the basic institutions of society). Of the 12 examples described in this paper only two did not refer to the political domain. In all applications the methods were used hypothetically. It was found that the range of applications was rather diverse in terms of purpose and content of the method.

Two conceptual (or normative) obstacles for applying Rawlsian methods were identified. The first refers to the issue of inclusiveness. For an equilibrium or consensus to be just it is important that all relevant actors are included and that they can equally engage in debate. The second obstacle is related to the communitarian objection to Rawls that people have to become detached from their personal life. The implication of the latter is that in situations where these personal concerns are important, the Rawlsian approaches lose their power. If the method is used to reach agreement between people with conflicting moral frameworks, the requirement that people are able to deliberate in terms of public reason becomes more important. Besides these normative obstacles, some authors raised doubt with regard to the feasibility of the Rawlsian approaches. It

has been argued that this is an empirical question, which cannot be answered on the basis of hypothetical cases alone.

In addition to the obstacles identified, it was found that in most applications the actors were not actively involved in the assessment of the equilibrium or the consensus. In order to advance the use of Rawlsian approaches in applied ethics, it would be interesting to see whether the relevant data (i.e., the considered judgments, moral principles, and background theories), can indeed be traced such as to be able to (re)construct a reflective equilibrium or consensus. Within the field of business ethics there is growing interest in the testing of ethical frameworks (cf. Giacalone et al.: 2005; Robertson et al.: 2008; Zarkada-Fraser: 2000), which could contribute to understanding how WRE can be established. This course should be explored further and in more depth. Methodological insights from the social sciences are essential to understanding the moral considerations and motivations of people involved. Hence, the empirical turn in applied ethics should not be limited to the use of empirical data but also to empirical methodologies.

If it turns out to be possible to assess the relevant data, the normative issues related to inclusion/exclusion, and to a lesser extent the detachment from personal life, can be addressed. It is important that moral justification should not be left to the ethical researcher alone. All relevant actors should get an active role in the assessment of the equilibrium or consensus.

Appendix

This appendix gives a brief description of the 12 applications.

Same-sex marriage (McClain: 1998)
McClain discusses Rawls' overlapping consensus in the light of the question what type of arguments citizens may appeal to or what kind of reasons they may proffer to support their positions. She illustrates this with the example of same-sex marriage and the requirements of reciprocity (i.e., the requirement to give reasons that can be understood by other citizens and that s/he might reasonably expect other citizens to accept; PR 797). McClain stresses the importance of critical examination of a supposed overlapping consensus and repudiation of it if it impinges on basic rights. Citizens who deliberate on political policy should support their comprehensive beliefs (i.e., arguments derived from philosophical, religious, or moral doctrines) by reference to public reasons and political values.

The status of the EU and international law (La Torre: 1999)
In his paper, La Torre presents a moral framework for the European Community law. Because Community law directly concerns and affects some of the most fundamental interests and values of Member States' citizens, Community Law is in need of a "general" and "thin" criterion by which to settle disputes between the different communities. This criterion should be based on a (minimum) common denominator of the different substantive thick legal and political communities, that is, on an overlapping consensus between the different member states, the author argues.

Curriculum reform (Halliday: 1999)
This paper discusses recent proposals to introduce citizenship and values education in the UK. Although widely welcomed, the values underlying this curriculum reform are open to debate. Halliday argues for a Wittgensteinian reinterpretation of overlapping consensus with regard to the political values in order to arrive at "a conception of citizenship to which all can happily assent" (Halliday: 1999, 47). The consensus consists not so much in doctrines as in beliefs about what ought to be done in particular circumstances. The curriculum reform should not be a State affair but rather something which is done inside the locus of community.

Liberal education (Paris and Kimball: 2000)

Paris and Kimball discuss the method of overlapping consensus in the context of liberal education in the US during the twentieth century. Based on previous work of Kimball, the authors compare a more pragmatic form of Rawls' overlapping consensus with the original approach as suggested by Rawls. Based on this comparison the authors argue that the emerging consensus concerning liberal education at the beginning of the 21st century is an overlapping consensus, that is, "a consensus whose nature is pragmatic, as well as a consensus whose substantive tenets are rationalized by pragmatism" (Paris and Kimball: 2000, 143), which is less "thick" than Rawls' version. This also allows for substantive changes over time if these are required.

European citizenship (Lehning: 2001)

Lehning discusses how questions regarding political identity and citizenship, raised by the creation of the "new Europe," can be addressed by appeal to liberal ideas of justice. The author claims that an overlapping consensus that results in a political conception of justice, shared throughout a political community, does generate a shared identity that will supersede rival identities based on, for instance, ethnicity.

Pediatric cochlear implants (Reuzel et al.: 2001)

Reuzel et al. discuss the method of WRE in the light of interactive technology assessment (iTA). iTA is a kind of technology evaluation that is characterized by active stakeholder participation and deliberation. Reuzel et al. do not apply the method of WRE but propose it as a way to come to a justified agreement on decision making in medical technologies. They argue that not any agreement or reflective equilibrium will do. It is important that the reflective equilibrium is inter-subjective. Stated generally, a WRE is only morally justified if the considerations of all relevant actors are involved, with no exceptions.

International Business vs. Globalization (Ho: 2003)

In her paper on transnational economic activities, Ho uses Rawls' conception of overlapping consensus and original position to defend the shift from international business, which she conceives as a practice of mutual competition and promotion of self-interest between separate national units, to globalization. According to the latter view, the construction of a global moral community in which there is an overlapping consensus is encouraged. This overlapping

consensus will guide economic institutions and business practices such that they can benefit all participants within the global community. Rawls' political conception of fairness can serve as a basis of informed and willing political agreement among free and equal participants through public reason.

Environmental values (Preston: 2004)

Preston tries to extrapolate principles of Rawls' theory of justice to the field of environmental ethics. Although Rawls himself does not consider environmental issues to be part of the public good – and therefore not part of fundamental political justice – it is desirable that environmentalists appeal to public reason to support their claim for environmentally friendly policies. In this article, environmental issues are the stepping stone to explore the limits of fundamental political justice.

Human rights (Tobin: 2005)

In her paper on the universality of human rights, Tobin criticizes the use of Rawls' notion of overlapping consensus to explain "how we can have meaningful agreement about human rights despite extensive cultural and moral diversity" (Tobin: 2005, 33). By discussing the issue of equality in Islamic comprehensive doctrine, Tobin tries to show that the interpretation of equality rights in Islamic comprehensive doctrine and Western, liberal, democratic communities is fundamentally different, which renders meaningful agreement impossible. Instead Tobin proposes to construct universality through actual dialogue both within and between communities.

Muslim minorities (March: 2006)

In his article on liberal citizenship and Muslim minorities, March seeks to establish "what political liberalism demands of Muslim citizens living as minorities in liberal states by way of a *doctrinal affirmation of citizenship*" (March: 2006, 373; italics in original). The objective of the article is to establish when it can be said that there is a consensus on the terms of social cooperation in a liberal society and thus that the comprehensive doctrine in question is providing its adherents with moral reasons for endorsing those terms.

Abortion (Moran: 2006)

The author discusses Rawls' theory in the light of the criticism expressed by former pope John Paul II that our current pluralistic, liberal society, which according to him is based on moral relativism, is in need of a transformation with regard to issues like abortion and euthanasia. Moran uses a revised version of Rawls' notion of overlapping consensus to show that one need not embrace relativism in order to defend liberalism. Although severely revised compared with Rawls' own idea of overlapping consensus, Moran's idea can be considered an application or elaboration of Rawls' original ideas. The main objective of the article is to justify procedural argumentation.

Technical and technological innovations in sport (Sheridan: 2007)

Sheridan explores the use of WRE as a possible decision making method for rational evaluation of technical and technological innovations in sporting practices. The paper describes the implementation of a new seeding system in male tennis at the Wimbledon championships. The WRE method was not used to justify the new seeding rules but only to hypothetically describe how the process of coming to these new rules might have developed.

4 Exploring rationales for distributing responsibilities[70]

4.1. Introduction

Although often aimed at advancing human well-being, the introduction of new technologies may come at the price of negative and unforeseen side-effects and risks. These consequences are hard to predict with any accuracy beforehand and often only materialize during use (Swierstra and Jelsma: 2006). Nevertheless, it is desirable that potential moral concerns (e.g., those regarding safety, sustainability, privacy, distribution of welfare, and social inclusion), be identified in as early a stage as possible and be included as additional design criteria in Research and Development (R&D) and technological design, even if they are not fully known yet (Schinzinger: 1998; Schot and Rip: 1997).

In the literature on the social implications of technology, these concerns are often addressed by proposals for involving stakeholders during design, development, and implementation of new technologies to broaden the scope of these processes. Many researchers argue that making the network inclusive will bring social and moral considerations to the table and make them part of the design and implementation process (see, e.g., Sclove: 1995). In the last decades, many participatory methods have therefore been developed to include the different perspectives of stakeholders in the development of and decision making on technology.[71] These participatory methods can be used for two different reasons (Rowe and Frewer: 2004). They derive either from recognition of the very nature of democracy; or they are a means to enrich the assessment and decision making through involving citizens and stakeholders in the process.

---

[70] This chapter has been accepted for publication in *Science, Technology, & Human Values.*

[71] Cf. Integrative Assessment (Van Asselt and Rijkens-Klomp: 2002), participatory technology development (Schot: 2001) or design (Kensing: 2003; Schuler and Namioka: 1993), stakeholder learning dialogues (Daboub and Calton: 2002), Constructive Technology Assessment (Rip et al.: 1995; Schot and Rip: 1997), Interactive Technology Assessment (Grin and Hoppe: 1995; Reuzel et al.: 2001), Participatory Technology Assessment (Joss and Bellucci: 2002; Schot: 2001), scientists stakeholder workshops (Cohen: 1997; Hanson et al.: 2006), and consensus conferences (Einsiedel et al.: 2001; Joss and Durant: 1995).

In the former case, participation is considered a way to empower citizens and stakeholders; hence, the participatory process is a goal in itself (Dryzek: 1997; Laird: 1993; Perhac: 1998). In the latter case, participation is a way to improve the quality of the decisions (Schot and Rip: 1997; Van Asselt and Rijkens-Klomp: 2002).

Research has shown that the inclusion of relevant stakeholders is not sufficient for getting what I tentatively call a *complete* distribution of responsibilities; that is, a distribution of responsibilities in which for all relevant moral issues at least one person in the network is responsible for addressing it (Doorn: forthcoming).[72] There are two ways in which a distribution of responsibilities can be incomplete. The first type of incompleteness refers to issues that are not "on the agenda." Most literature from Science and Technology Studies (STS) focuses on this type of incompleteness (Cf. Einsiedel et al.: 2001; Saari and Miettienen: 2001). However, even if all people agree that some issue needs to be addressed, this is no guarantee that it *will* in fact be addressed. There may still be gaps in the distribution of responsibilities because people do not consider it their own responsibility to address these issues but rather expect someone else to do it. We could therefore interpret it as a problem of "collective cooperation." In the ethics literature this is known as (an instance of) the problem of many hands; that is, the problem of identifying the person responsible for something if a large number of people is involved (Bovens: 1998; Thompson: 1980).[73]

There are several ways to understand this problem of "collective cooperation." In the last decades, the notion of network has gained an important place in STS literature on cooperation between actors. According to the network approach, positions or roles occupied by social actors and the relations or connections between these positions are the basic units of analysis and not the individuals

---

[72] By "someone being responsible for a moral issue" I mean that someone has the task to see to it that it is taken into account during the development and design phase. This does not mean that all risks should be completely prevented – a requirement which is impossible to live up to – but that at least everything that can reasonably be known should be considered during the design and development phase. What could reasonably be known should fit the standards of reasonable care, as exemplified in, for instance, professional standards.

[73] Although the more common interpretation of the problem of many hands refers to backward-looking responsibility (that is, responsibility for faults or mistakes), it could also be interpreted as the forward-looking responsibility of identifying or agreeing on the person responsible for doing something in the future.

that occupy these roles (Knoke: 1990). Because the actors are constrained by their position in the network, their rationality is bounded rather than comprehensive. Network cooperation could therefore best be understood as a strategic process in which the outcomes are the result of bargaining between agents in the networks (Dowding: 1995). Following this network paradigm, we could view this collective cooperation problem as a problem that stems from opposing network positions and interests. Accordingly, the distribution of responsibilities is a strategic negotiation process in which people frame the problem and set priority's on the basis of their own agenda (see, e.g., Grin and van der Graaf: 1996a; Grin and van der Graaf: 1996b; Henneberg et al.: 2006; Möller: 2010). In the extreme case, responsibilities are only assumed if it fits the actors' interests.

However, recent empirical work indicates that this cooperation narrative is too one-sidedly focused on network positions. Glenna (2010) argues that cooperation conflicts should be analyzed in terms of different ethical traditions rather than differences between groups pursuing their own interests. Glenna analyzed the so-called New York City (NYC) watershed controversy, a controversy that emerged in the 1990s when residents of the NYC watershed filed a lawsuit to block NYC's proposed regulations for the land surrounding the streams and reservoirs that supply NYC's drinking water. Whereas the press and academic literature portrayed the conflict as a controversy between two "roles" (the residents of the watershed as anti-environmentalists and economic growth proponents; the City as the pro-environmental opponent), Glenna interprets the problem as a controversy of conflicting ethical views on justice and responsibility. He argues that the underlying dimensions of ethics and justice tend to be overlooked when analyzing environmental conflicts in terms of these positions, which hampers the actual resolution of the conflict.

In a previous empirical study on responsibility distributions in professional networks, these findings were confirmed. The empirical data described in (Doorn: 2010b) suggest that these cooperation problems can be partly traced back to the different normative background theories people endorse. Although these background theories are not predictive for people's actual judgments on what makes a desirable distribution of responsibilities, they seem to indicate different ways of reasoning and justifying outcomes and therewith to reflect different conceptions of responsibility.

In applied ethics, the reflective equilibrium model has been proposed as an approach to reconcile such a pluralism of ethical views. Based on the work of the

political philosopher Rawls, this model allows for decision making in a pluralist context with different stakeholders, without giving *a priori* priority to any of them. Although Rawlsian approaches have attracted ample attention in the field of applied ethics – they are often recommended in scholarly literature as a proposed way out of an ethical impasse or as a fruitful approach to do further research on (cf. Brand-Ballard: 2003; Gracia: 1995; Van de Poel and Royakkers: 2007) – the actual application of these approaches is still relatively rare. Consequently, insight in how to retrieve the necessary empirical data and how the model actually "performs" is still rather weak.

The aim of the present chapter is to see whether the reflective equilibrium model can explain the distribution of responsibilities in the context of R&D. In this model, the collective cooperation problem is treated as a problem emerging from conflicting conceptions of responsibility rather than conflicting network positions (assuming that these conflicting responsibility conceptions themselves do not stem from the respective network positions). Because the actors' own conceptions of responsibility form the point of departure in the reflective equilibrium model, I focus on the question how the actors within design and research teams conceive of responsibility for addressing the moral and social issues related to technology (including their own) and how they think these responsibilities should be distributed among the team members. By exploring the different rationales for distributing responsibilities, I try to improve our understanding of responsibility distributions in the context of R&D with the ultimate aim of reducing the occurrence of incomplete responsibility distributions.

The empirical data is based on a case study covering the development of a prototype application for in-house monitoring of patients, based on Ambient Intelligence technology (the ALwEN project, where ALwEN is an acronym for Ambient Living with Embedded Networks). This project was studied as part of an ethical parallel study (see, e.g., Van der Burg (2009) for a description of this type of ethical research). The idea of such parallel research is that ethical investigations are carried out parallel to, and in close cooperation with, a specific technological R&D project. The R&D project described here is carried out by a team of commercial and industrial companies, several universities and a clinical research center. In the methodology section, I give a more elaborate description of the project.

The outline of this chapter is as follows. In Section 4.2, I describe the reflective equilibrium model for moral deliberation. This model provides the

theoretical framework for interpreting the empirical data. In this section, I explain the importance of thinking in terms of different responsibility conceptions and the need to distinguish them from responsibility as a concept. In Section 4.3, I describe the methodology of the research and I present a summary of the empirical observations. In Section 4.4, I discuss the empirical data, followed by a concluding Section 4.5, in which I summarize the findings.

4.2.   Moral deliberation in a pluralist society: Rawls' WRE model

In the previous section, I presented the incompleteness of responsibility distributions as a problem stemming from different – sometimes competing – responsibility conceptions. These different conceptions reflect philosophical disagreements on what responsibility amounts to and under what conditions one is responsible for doing something. To understand this, we have to distinguish between "concepts" and "conceptions," a distinction introduced by the political philosopher John Rawls in order to clarify philosophical disputes over the meaning of "justice" [(Rawls: 1999 [1971]); see also (Hart: 1961)]. Rawls explained that people all have different views on how to shape social cooperation; that is, they all defend a different set of principles that indicate what they find just. However, although people may disagree about exactly which principles should define the basic terms of cooperation, they "understand the need for, and are prepared to affirm, a basic set of characteristic principles for assigning basic rights and duties" (Rawls: 1999 [1971], 5).[74] This basic set of characteristic principles (the concept of justice) is the central element which all these different sets of principles (that is, the different conceptions of justice) have in common. It encompasses, for example, the idea that "like cases should be treated alike and different cases differently;" it leaves room, though, for different interpretations of what differences in cases are to be regarded relevant and what differences are not. Similarly, people may use stipulative definitions to indicate when the term responsibility applies and how it connects to other theoretical terms (such as duty or obligation), but these definitional clarifications cannot solve all

---

[74] Rawls refers to the legal philosopher H.L.A. Hart who defends a similar distinction between what he calls a "leading precept" of justice (cf. the concept) and the relevant criteria of justice which "may often vary with the fundamental moral outlook of a given person or society" [cf. the conceptions; (Hart: 1961, 155-159)]. I could not find any earlier text than the one by Rawls in which the terminology "concept" and "conceptions" was used in this particular way.

disagreements on the question "Who is responsible?" because the different conceptions of responsibility reflect different criteria for when it is appropriate or fair to ascribe responsibility (the "fairness requirements"). These different fairness requirements, in turn, can be traced back to the different reasons people have for ascribing for responsibility. One could think of the development of a new risky technology, where people need to distribute the moral responsibilities to prevent harm or to look after certain risks. For the general public, we want to avoid harm as much as possible and for this reason we want all risks to be looked after. Hence, if one has a conception of moral responsibility that is based on the no harm principle, for example, it may seem functional to make anyone in the innovation chain responsible for preventing harm. However, some researchers may judge it unfair that they have to look after (all) the risks, also highly unlikely ones and maybe risks they are themselves not familiar with. They may say that it is preferable to let a more specialized person deal with the risks. From a consequentialist perspective, this may be more efficacious.

When responsibility is not explicitly discussed in the research team, these different conceptions may lead to different expectations and ultimately, gaps in the distribution of responsibilities because people expect others to assume a particular responsibility.

The fact that these conceptions may sometimes conflict is not just a practical problem but also a moral problem; if we look at the example given above, both the conception based on the no harm principle and the consequentialist conception rely on a particular normative background theory, or – as Rawls calls it – "comprehensive doctrine." The term "comprehensive doctrine" is typically used to refer to a religious, philosophical, or other standard moral doctrine that "applies to all subjects and covers all values" (Rawls: 2001, 14). These doctrines have a particular "conception of what is of value in human life or, alternatively, of what is regarded as a fully worthwhile life" (Rawls: 2001, 19).[75] In this paper, I defend the pluralist thesis, which says that at least some of these different comprehensive doctrines are legitimate and that they cannot be reduced to one overarching view or value (Dryzek and Niemeyer: 2006). Hence, if there is a link between people's legitimate comprehensive doctrines and their conception of responsibility, we can say that there is also a pluralism of responsibility

---

[75] Comprehensive doctrines contrast with political conceptions that are not related to any one particular comprehensive doctrines but that are compatible with (one or more) of these comprehensive doctrines (Rawls: 2001, 19).

conceptions. If we want to do justice to this pluralism of responsibility conceptions, we need a procedure or framework that is impartial towards the different conceptions.

In this chapter, I use the Wide Reflective Equilibrium (WRE) approach, initially developed by Rawls (1999 [1971]) and further elaborated by Daniels (1979; 1996), as a framework for analyzing moral deliberation concerning responsibility. In the concept of WRE, a distinction is made between three layers of considerations: (1) descriptive and normative (moral) background theories, (2) moral principles, and (3) considered moral judgments about particular cases or situations. Reflective equilibrium refers to a state of coherence between one's considered judgments and moral principles concerning a certain case. According to Rawls, everyday moral reflection takes place by examining our moral judgments on particular matters (layer 3) with more general or broader beliefs and principles on similar issues (layers 1 and 2). In order to decide how to respond to moral issues, people move back and forth between these various beliefs and considerations, reflect on them, revise them if necessary and try to achieve an acceptable coherence between their moral judgments on particular matters with more general or broader beliefs and principles on similar issues (see Figure 4.1 for a graphical presentation of the WRE model). In this deliberative process, all three layers are open to revision. In the literature, the term WRE can refer both to the state of coherence between these three layers of morality and to the process or method itself of arriving at such equilibrium.

Rawls used the concept of WRE in explicating and defending his theory of justice in the context of political philosophy. He tried to develop a criterion of justice that would be *fair* to all despite the diversity of moral frameworks people endorse (hence, the idea of justice *as* fairness). Although Rawls at first tried to develop a substantive conception of justice – viz. the right to equal basic liberties, the right to fair equality of opportunity and the difference principle (which says that inequalities are allowed only if they work to the benefit of the worst-off group) – he later recognized the plurality of incompatible and irreconcilable moral frameworks within a democratic society and he limited the idea of justice as fairness to a purely procedural conception of justice. People with divergent moral doctrines will most probably not agree on a substantive conception of justice but they can overlap in their acceptance of a procedural conception of justice. For these procedural principles to be justified they must cohere with each individual's background theories and considered judgments. Rawls introduced the term overlapping consensus to refer to these justice

principles that are agreed upon by all citizens and that are part of each individual's WRE. If there is coherence between all layers of the WRE model, the outcome of the reflective process can be considered justified, Rawls argued.

Assuming that the participants working in an R&D project want the moral tasks to be properly addressed, we need a distribution of responsibilities that is both acceptable to all (i.e., a distribution of responsibilities that is *justified* in terms of each individual's WRE) and complete (or at least "complete enough," in the sense of having only tolerable omissions). For this justificatory purpose, we therefore need to know how people rationalize the responsibility they take or have taken, because only then we can see whether the distribution of responsibilities coheres with each individual's principles and normative background theories.
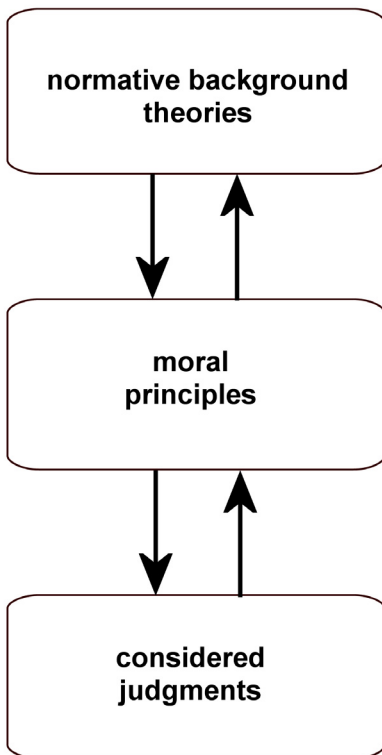
Figure 4.1: Wide Reflective Equilibrium model

Van de Poel and Zwart (2010) argue that learning about one's value systems is a prerequisite for arriving at an overlapping consensus because only then do people become aware of the legitimacy of other people's opinions and normative background theories. In order to encourage these learning processes, the WRE approach can be used constructively, describing the process of working towards a justified outcome.

To summarize, the WRE model provides an interpretative framework for analyzing moral deliberation in a collective context. In addition, the model can be used in a constructive way, where the outcome of the deliberative process can be considered justified if it fits each individual's WRE.

In the next section, I describe a workshop that was organized with actors in a real research team to analyze how the WRE model can explain moral deliberation in the team and how the different rationales for distributing responsibilities were explored.

4.3.   Case and methodology

In this section, I describe the methodology for retrieving the empirical data that are used to test the WRE model and to explore the different rationales for distributing responsibilities. The empirical data are based on a qualitative case study that I carried out in the period September 2008 – August 2010. The case study concerned the ALwEN project, an R&D project aimed at developing a health care application based on Ambient Intelligence technology. In Section 4.3.1, I describe the ALwEN project in more detail.

After an observation period of several months (September 2008 – May 2009), a series of interviews was carried out with 13 representatives of the different institutional partners involved in the project.[76] The interviews were recorded and approved by the interviewees. An extensive overview of the interview results is presented in Doorn (forthcoming). On the basis of the interview results, some striking issues were selected and explored in more detail during an interactive workshop. Ten months after this workshop, two additional interviews were carried out with two project members to evaluate the case study and to discuss the lasting effects of the workshop.

---

[76]   In this thesis, I reserve the term "partner" to refer to institutions involved in the project and the term "actor" for individual people affiliated to the different institutional partners.

This chapter draws heavily on the material collected during the workshop, complemented with the data collected through the interviews. In the following subsections, I describe the ALwEN project and the workshop in more detail.

4.3.1. Development of an Ambient Assisted health care application: The ALwEN project

The ALwEN (Ambient Living with Embedded Networks) project is aimed at developing a prototype application for in-house monitoring of patients, based on Ambient Intelligence technology. Ambient Intelligence reflects a vision of the future of ICT in which intelligence is embedded in virtually everything around us, such as clothes, furniture, etc. The technology consists of Wireless Sensor Networks (WSN), the combination of body sensors, ambient sensors and wireless networks. In order to capture the whole trajectory of fundamental research to the development of a prototype application and ultimately commercial exploitation, several universities, a consortium of 12 Small and Medium Enterprises (SME), two independent industrial research institutes and a scientific research center in rehabilitation technology cooperate. In the project a use case is developed to serve as an example of what can be done with WSNs and to focus the work of the demonstration activities of the project. The use case describes a situation of in-house monitoring of the daily activities of a patient with Chronic Obstructive Pulmonary Disease (COPD), a chronic lung disease.

In the project, end users, including health care professionals, are consulted to clarify their wishes and demands with respect to the environment to be created. In the project proposal, a first experimental set-up of the WSN is scheduled, followed by explorative experiments with real users to determine the functional and technical requirements in more detail. Afterwards the experimental WSN will be evaluated both in terms of the technical specifications and in terms of the objectives set to improve quality of life of the end users.

4.3.2. Description of the workshop

Before describing the workshop, it should be noted that the aim of the workshop was to retrieve the relevant data for testing the WRE model and to explore the different participants' rationales. It was not the explicit aim to brainstorm on relevant moral issues or to make the project as "morally responsible" as possible. With this aim in mind, some salient issues were selected for discussion rather

than an extensive list covering the full project (see also footnote 83). These issues are all related to the "social acceptance" of the technology, which was defined as the explicit goal of the project.[77] However, when the project started, this notion of "social acceptance" was still rather vague. In the interviews, I therefore asked the participants how they conceived of this notion of social acceptance.

It is interesting to note the differences between the technical researchers and engineers on the one hand, and the clinical researchers on the other. For the former, social acceptance (as a goal of the project) was primarily conceived as social *acceptability*; that is, a prospective quality that was to be determined by experts, by whom they meant the clinical researchers and possibly the ethicist as well. For these researchers, the involvement of clinical and ethical experts was therefore crucial for achieving the goal of social acceptance.

The clinical partners, on the other hand, defended a participatory approach, the result of which would lead to social acceptance. As one of the clinical researchers argued:

> In a way you could say "the proof of the pudding is in the eating": if the service is being used by the intended users (including all technical imperfections, user unfriendliness and the fact that they have to pay for it), it is socially acceptable. That means that the use of the service for some other than the intended purpose is still OK. The acceptability lies in the final use.

---

[77] Since "social acceptance" was, in the original research proposal, identified as a crucial element of the success of the project, this notion of "social acceptance" was chosen as the starting point of the author's ethical investigations, including the necessary conditions for getting the technology socially accepted. In the interviews, the representatives of the different institutional partners involved in the project were invited to brainstorm on the relevant "moral issues" pertaining to the project. The interviewees were asked to think of "moral issue" in as broad a way as possible: anything related to risks and moral values (e.g., social acceptance, human well-being, privacy, society, and sustainability) was considered relevant. According to the technological researchers, these issues should be addressed in order to gain social acceptance. I realize that this description of moral issue is not as well-defined as some philosophers would like it to be. However, since the interviews and the workshop were explicitly aimed at tracing the opinions of the engineers themselves, I did not give any constraints on what counts as a moral issue nor did I introduce issues that were not mentioned by the engineers themselves. For a more well-wrought description of when a value can be considered a *moral* value, see (Nagel: 1979, Chapter Nine: The fragmentation of value).

In this view, the technology's acceptability is defined in terms – or maybe one could say, is constructed by – its acceptance. For these researchers, the inclusion of end users was therefore of primary importance.

In the remainder of the text, I use the term "social acceptance" to refer to both the social acceptance and the social acceptability.

The workshop was organized in the university's *Group Decision Room* (GDR), a room with electronic meeting support systems that allow for a participative approach to complex tasks such as anonymous discussion, brainstorming, voting, etc. Each individual has a laptop and can participate in online discussions. The reason for using the GDR was to facilitate the anonymous discussion which was expected to give more frank and honest reflection on delicate topics. Moreover, since all discussions are automatically recorded, no interpretative transcriptions are required. For this workshop, the relevant actors in the technological project were invited.

The following people participated in the workshop:[78]

- Four technical researchers formally responsible for one of the project's so-called work packages ("work package leader"); three of these researchers are affiliated to one of the SMEs and one to a university-based research institute. Two of these researchers worked in software engineering, one in embedded systems engineering, and one in electrical engineering.
- A postdoctoral researcher in computer science and a PhD student in electrical engineering; both from different universities.
- A researcher from the rehabilitation research center with a background in biomedical engineering.
- A systems architect (researcher) from one of the independent industrial research institutes.

Except for the PhD student, all workshop participants were also interviewed. This overview shows that all participants had a background in engineering.[79] In

---

[78] In addition to the eight participants working on the technological project, the author and two colleagues from the author's own department were present to moderate the session.

[79] In the literature, there is an ongoing discussion on the exact definition of engineering and the question what engineers distinguishes from technologists or scientists. Concerning the distinction between scientists and engineers, the different function of knowledge is usually taken as an important difference. Whereas scientists seem primarily driven by the goal of

the remainder of the text I will not further distinguish between the different researchers in order to ensure anonymity. These participants were selected because together they represented all relevant institutional partners in the project. Moreover, these participants were all actively involved in the project and as such well-informed.

The workshop was structured along the different layers of the WRE model described in Section 4.2. Together these three layers of the WRE model give insight in how people reflect on moral issues. A detailed workshop agenda can be found in the appendix. In the remainder of this subsection, I describe how the different layers of the model were assessed. To start with the normative background theories, insights from moral psychology were used to assess this layer. In moral psychology, ethics position theory (EPT) has been developed to empirically account for disagreements among people when discussing questions of ethics. EPT maintains that the individuals' personal normative background theories (or "ethics positions") influence their judgments, actions, and emotions in morally laden situations (Forsyth et al.: 2008). As such it is the empirical counterpart of the philosophical view that people endorse different "comprehensive doctrines." Similar to the different meta-ethical positions in the scholarly philosophical literature on the basis of which one can judge the moral "rightness" or "goodness" of particular actions and consequences (be it a typically Kantian or consequentialist position, or one based on virtue ethics), EPT holds that people's judgments can be traced back to or are informed by their different ethics positions.

EPT distinguishes between two dimensions: particularism (skepticism with regard to inviolate moral principles) and idealism (concern for benign

---

knowledge (as an end in itself), engineers are considered to be primarily driven by a means-end reasoning, often combined with the notion of design (Vincenti: 1990; Davis: 1995). The distinction between technologists and engineers is more subtle. I think the main difference is that the notion of design gets a more prominent place in the work of engineers, as also exemplified in the following definition of engineering as the "practice of organizing the design and construction of any artifice which transforms the physical world around us to meet some recognized need" (Rogers: 1983, 51). With this definition of engineering, the people in the ALwEN project can certainly be considered engineers, although some in different subfields. Since I take engineering to be a profession in the Western world (Davis: 1997; 2009), the foregoing implies that the actors in the ALwEN project have a *professional* responsibility.

outcomes).[80] Particularism refers to the extent to which individuals reject universal moral rules in favor of the particular features of the situation and action they are evaluating. Idealism refers to the degree to which individuals are idealistic in their attitude toward the consequences of actions. People with a relatively high score on the idealism dimension "assume that desirable consequences can, with the 'right' action, always be obtained" (Forsyth: 1980, 176). People with a relatively low score on the idealism dimension are more pragmatic; they assume that harm is sometimes unavoidable and that one, in case of moral dilemmas, needs to choose between the lesser of two evils. Whereas the idealism dimension mainly focuses on consequences, the second dimension (particularism) pertains to the question whether or not moral principles are the basic guides for evaluating actions in terms of right/wrong or good/bad. For people with a relatively high score on the particularism dimension, the contextual particularities of the situation are most important, whereas people with a relatively low score on the particularism dimension tend to rely more on moral principles and norms for evaluating their actions (Forsyth: 1980; 1985; 1992; Forsyth et al.: 2008).

Although Forsyth and colleagues do not provide explicit definitions of the four positions – the descriptions of the two dimensions are supposed to suffice – they make an explicit link with the different positions in moral philosophy, most notably for the two positions with a low score on the particularism dimension. With a low, respectively high score on the idealism dimension, these positions correspond to a consequentialist/utilitarian, respectively Kantian framework.

---

[80] Although basically referring to the same moral concepts, the names used in the moral psychology literature are somewhat different than the terminology utilized in the meta-ethical debate. In moral psychology, the dimensions are labeled relativism and idealism. However, these names are misleading when compared with the meta-ethical debate (see also Doorn: 2010b). In EPT, universalism is seen as the opposite of relativism. However, in moral theory the appropriate counterpart of universalism seems particularism, the view that there are no overriding principles applicable in every case but that moral judgments should be based on a case by case basis (see (Dancy: 2004; Hooker and Little: 2000) for a defense of this view). In this paper, I follow the terminology commonly used in moral philosophy and use the name particularism to refer to this "relativism dimension." A similar remark could be made for the idealism dimension. The name idealism might misleadingly suggest that exceptionists and subjectivists are not driven by a vision of how people should behave or how things should be arranged. Although there was only one participant with a relatively low score on "idealism," this participant clearly showed that he was motivated by some ideal (in this case, the ideal of efficacy), from which a normative appeal can be derived.

These are also the two frameworks ("comprehensive doctrines") that Rawls' typically describes as forming the most abstract layer in his WRE model.[81]

Forsyth and colleagues (1980; 1988) developed the Ethics Position Questionnaire (EPQ) to assess people's score on the idealism and particularism dimension and to assess their ethics position accordingly. The EPQ score is comprised of Likert scale responses to 20 statements. In the workshop, the EPQ was used to assess the participants' personal normative background theories. The advantage of using the EPQ is that it is a well-calibrated instrument that does not require subjective interpretation by the interviewer. The four categories or ethics positions are shown in Table 4.1.

Regarding the second layer of the WRE model, Rawls speaks of moral principles pertaining to a particular case. Although Rawls does not explicitly say how the different conceptions of justice are to be fitted in the WRE model, his definition of justice conception as "a basic set of characteristic principles for assigning basic rights and duties" suggests that these conceptions could be interpreted as an operationalization of this second layer. I therefore situate the different conceptions of responsibility also in this layer. Concerning the different responsibility conceptions, these may be (more or less) well-defined terms in the literature on responsibility, in practice they are probably more ambiguous. I therefore choose to operationalize the responsibility conceptions as responsibility *rationales*, which I define as the recurrent type of argumentation used by the different actors to support their claims. These types of arguments typically reflect the different values or principles that people consider important when distributing responsibilities and as such they seem a sound operationalization of the responsibility conceptions. In the remainder of the text, I use the term responsibility conception to refer to the theoretical responsibility conceptions found in the moral philosophy literature and the term responsibility rationale to refer to the empirically reconstructed responsibility conceptions. The characterization of the types of argumentation (that is, the rationales) is based on an interpretation of the online discussions and a written justification/evaluation of the workshop's outcomes.

---

[81] Although Rawls leaves open the possibility to include more background theories in this layer (other than the comprehensive doctrines themselves), in his writings he seems to focus on the standard theories in moral philosophy.

In the context of responsibility, the considered judgments in the WRE model can be seen as the actual responsibility ascriptions. These were assessed by asking the participants to distribute a list of "moral tasks" among the different actors or project activities.[82] This list included tasks like "making sure that the application does not interfere with everyday life," and "identifying how technological choices affect social acceptance."[83] In the present research, both the list of tasks and the list of project activities was pre-established on the basis of the interviews. According to the technological researchers, these tasks should be addressed in order to gain social acceptance. After discussing the differences of opinion and different conditions for responsibility, the participants were asked to redistribute the tasks. The latter was done to assess whether the participants converged to a consensus.

---

[82] Some definitional clarifications are in place here. In order to avoid a discussion on a too personal level ("*you* should have done that!"), I used the more neutral terms project "activities" or "phases" as the organizational entities to ascribe responsibility to. Concerning the "moral tasks," I used the notion of task-responsibility throughout the workshop, where the task-responsibility for X is defined as "the task to see to it that X" (Van den Hoven: 1998). Although the word task might mistakenly suggest that the items are well-defined assignments, they are rather "ethical challenges" or "issues" to be dealt with. However, to avoid misunderstandings and unnecessary wordiness, I use the word "task" to refer to the task-responsibilities listed above and I reserve the word challenge and issue for addressing the "ethics of the project" on a more abstract level.

[83] The full list of tasks comprised the following elements: making sure that the application does not interfere with everyday life (invisibility of technology); setting the requirements of the security of this application (how secure is secure enough?); striking the right balance between user friendliness, reliability and functionality; making sure that end users (patients, their family & friends, clinicians) are able and willing to use the application; starting a broad societal discussion about the desirability of these kinds of (monitoring) applications; addressing questions related to data storage and data access (legal aspects); inventorying/monitoring potential risks of the present application; identifying how technological choices affect the social acceptance. The list of activities comprised the following elements: project management; research on hardware; research on software; simulation; (formal) modeling, network configuration/optimization; security and privacy; laboratory experimentation; clinical research; clinical experimentation; exploitation of application.

Table 4.1: Taxonomy of Ethical Ideologies (source: Forsyth, 1980)

	Particularism	
Idealism	**High**	**Low**
High	**Situationist** Rejects moral rules; advocates individualistic analysis of each act in each situation; relativistic; no moral compromises.	**Absolutist** Assumes that the best possible outcome can always be achieved by following moral rules.
Low	**Subjectivist** Appraisals based on personal values and perspective rather than universal moral principles; relativistic.	**Exceptionist** Moral absolutes guide judgments but pragmatically open to exceptions to these standards; utilitarian.

4.3.3. Observations

During the workshop, all three layers of the WRE model were assessed. Concerning the layer of normative background theories, four different ethics positions are distinguished in EPT (see also Table 4.1): absolutism, situationism, exceptionism, and subjectivism.[84] It appeared that most participants had a tendency to either the absolutist position or the situationist position. One person had a typical subjectivist position and only one person had an inclination towards exceptionism (combined with absolutism; this person's score on the idealism dimension was not very distinct).

Concerning the layer of the responsibility rationales, three recurrent types of argumentation could be distinguished: one based on fairness considerations ("What is a fair workload?"), one based on pragmatic and goal-directed considerations ("Who is best equipped to perform a certain task?"), and one putting emphasis on the user perspective ("How to involve the user?"). Analysis of the discussions showed that participants with a similar ethics position tended

---

[84] It should be noted that those four categories are not sharply defined boxes. I therefore prefer to about tendencies or inclinations rather than fixed positions. Idealism and particularism are, after all, "spectrum concepts" and not dichotomous either/or concepts.

to follow the same rationale and use comparable arguments to defend their claim. Table 4.2 shows a summary of both the ethics position and the type of argumentation (layers 1 and 2 in the WRE model) for the 8 workshop participants. These data suggest that there was indeed a typical absolutist rationale (with fairness considerations), a typical situationist rationale (with the user and societal perspective), and a typical subjectivist rationale (with pragmatic and goal-directed considerations).

This correlation between ethics position and type of argumentation suggests that the participants' rationales for distributing responsibilities are, at least partly, related to the different normative background theories they endorse.[85] In other words, the normative background theories (here operationalized in terms of the ethics position) do indeed give an indication of the rationales that different people follow.

Considering the third layer (the actual responsibility ascriptions), most participants recognized the multi-faced character of social acceptance and the need to address the listed tasks. Since the participants were asked to distribute the responsibilities twice, we could assess where their opinions on the question "Who is responsible?" converged or diverged. It appeared that there was convergence on most issues. On a general level, the result of the workshop was that the participants converged to the opinion that the focus should shift towards laboratory and clinical experimentation so that moral and social issues could be better investigated and addressed.

An interesting point of observation concerns the participants' opinions on the scope of the project and the inclusion of different tasks. The workshop participants were allowed to say that certain tasks were beyond the scope of the project (or not to be addressed at all). Especially in the second voting round (that is, after the participants had mutually discussed the responsibilities), only few participants labeled some of the tasks as "outside this project." This means that most participants considered it the project team's responsibility (either on an

---

[85] Although the disciplinary background of the participants may provide a set of background theories as well, the participants in the workshop to a large extent shared the same disciplinary background (see 4.3.2). If these "disciplinary background theories" would be predictive for the different rationales, we would expect one dominant rationale rather than these three different rationales. In order to assess the influence of disciplinary background to the responsibility rationales, we would have to ask the same questions to a more diverse (in terms of disciplinary background) project team.

individual basis or collectively) to address the moral tasks, also the broader societal ones (for example, addressing legal questions related to data storage and data access).

In their evaluation of the workshop, most participants indicated that they had become more aware of certain moral issues (e.g., the need to involve end users). There was a general agreement that most moral issues span several activities within the project and that it is therefore difficult to single out one activity where it should primarily be addressed. The primary responsibility was in those cases ascribed to the project management (for coordinating this joint effort), to the experimentation phase (where all activities were supposed to come together), or to the clinical partner. Some participants explicitly mentioned that this workshop made them realize that some moral issues were currently not addressed adequately. The idea that the work should shift from research towards either laboratory or clinical experiments with a (prototype) application was shared by all. Soon after the workshop, a brainstorm meeting was scheduled in which the requirements for clinical experimentation were discussed in more detail.

Table 4.2: Summary of individual actors' ethics position and type of argumentation

Participant	Ethics position	Type of argumentation	Project activity primarily responsible
1	Absolutist (Situationist)	Fairness (workload); workplace relations	Clinical Experimentation
2	Subjectivist	Goal-directed; efficiency	Project Management
3	Situationist	User perspective; societal; action-oriented (initiating new plans)	Clinical Experimentation
4	Situationist	User perspective, societal	Project Management/ Outside project
5	Absolutist	Fairness (workload)	Research on Software
6	Absolutist (Exceptionist)	Fairness (workload); user perspective	Clinical Experimentation
7	Absolutist	Fairness (workload)	Clinical Experimentation
8	Situationist	User perspective; Goal-directed	Project Management/ Clinical Experimentation

4.4.   Discussion

In order to better understand the correlation between the ethics position and the type of argumentation (the rationales), I will now discuss the rationales of three of the workshop participants in more detail and try to reconstruct them in terms of the different layers of the WRE model. I have selected three workshop participants with a distinct ethics position but with a similar role in the project, viz. a formal management role (either of the project as a whole or of one of the work packages) combined with the role of technical researcher and technology producer. These participants were all representatives of one of the SMEs. For reasons of validity, I checked with the participants whether they recognized my interpretation of their rationales and they all approved of the respective descriptions.

The first participant (participant 1 in Table 4.2) combined a relatively high score on the idealism dimension with an average score on the particularism dimension. Hence, in terms of principles, this participant seems to uphold the view that moral dilemmas do not really exist and that right consequences could, with the right actions, always be obtained. Given the average score on the particularism dimension, this participant relies on both moral principles and contextual properties for evaluating his actions.

In the discussions, this participant emphasized the need to involve all activities. Compared to the other participants, he[86] placed more responsibility for ethical issues with the technical activities themselves. When discussing how to distribute the moral tasks, he repeatedly argued that these cannot be addressed in one activity ("The issues do not simply map onto tasks. They need a multi-disciplinary approach to be addressed;" and "There cannot be a primary activity that addresses all these issues. In other words: You simply need all these activities to address these issues!"). In the discussion of the conditions for responsibility, this participant emphasized again the interrelatedness of the different subprojects and the need to know and understand the role of the different project members. Also in the interview, this participant indicated that he highly valued the joint contribution of all researchers and the lack of hierarchy:

---

[86]   In order to anonymize the results, I use the male (possessive) pronoun throughout, also for the answers given by the female participants and interviewees.

> The communication with the PhD students is rather open. There are no hierarchical decisions; PhD students are involved in these decisions as well. This decision making is more of a vague, continuous process. [...] Ultimately, all partners have to learn from each other and to make compromises.

The interviews and the discussions during the workshop suggest that this participant was mainly driven by an ideal of equality and fairness, both in terms of nonhierarchical workplace relations and in terms of division of labor. He was less driven by strategic considerations stemming from his role in the project. In a concluding interview, this participant indicated to prefer multidisciplinary projects over monodisciplinary ones because multidisciplinary projects will, in the end, result in better outcomes.[87] In this light, his remark "[...] all partners have to learn from each other and to make compromises," could be seen as a confirmation of his high score on the idealism dimension. By mentioning "compromises," he seemed to refer to joint and mutually inspired outcomes rather than (suboptimal) negotiated agreements.

As already shown in Section 3.3, other participants with a comparably high score on the idealism dimension seemed to emphasize the fairness of the workload as well. Especially the participants who combined a relatively high score on the idealism with a relatively low score on the particularism dimension (see Table 4.1; absolutism) made "fairness" a guiding principle for distributing the moral responsibilities. Most of them agreed that it would be unfair, to make the people in one particular activity responsible for addressing all the ethical issues.

The second participant I would like to discuss (participant 2 in Table 4.2) combined a relatively low score on the idealism dimension with a relatively high score on the particularism dimension, which makes him a typical subjectivist (see Table 4.1). Compared to the other participants in the workshop, this participant considered more moral tasks as not to be addressed in the project. More than others, he emphasized that the main goal of the project is technology development and not so much the development of a prototype application. For the tasks that were to be addressed in the project, he emphasized the role of the project management. At the same time, he argued that the project management

---

[87] By a multidisciplinary team, this participant meant a team consisting of both technical researchers and researchers with a background in the humanities or social sciences.

also needs a mandate to further delegate certain responsibilities. He indicated that this was missing in some situations. The relatively low score on the idealism dimension indicates a tendency to use pragmatic arguments and recognition of the need to make moral compromises, which was confirmed by his contributions to the discussions. For example, this participant said it was best to build something that works first and to take that as a starting point for putting more focus on the ethical aspects. This sharply contrasts with the "idealist" view on the consequences of our actions, which holds – in its most extreme form – that these consequences would never need to be suboptimal. The relatively high score on the particularism dimension suggests that this participant's "moral compass" is based on context rather than moral principles. In the discussions, this participant seemed to be guided by the question "Who is best equipped to perform a certain task?" which does indeed indicate a sensitivity to context.

The last participant (participant 3 in Table 4.2) was driven by yet another rationale. In the discussions, he mentioned the lack of user involvement several times and in his evaluation afterwards he indicated that it should be a wise idea to add ethics as part of the project plan and technical requirements in future projects. He explicitly mentioned the need for involving stakeholders in writing such a project plan. In the discussion on where to address the moral tasks, he emphasized the need for experiments. For him, it was important that the ethics in the project would be a real element and not just "a bunch of related (paper) work." For this participant, the main concern seemed to be the lack of user involvement ("It strikes me that the end user in *all* those issues is again lacking [emphasis in original]", "Where is the customer?"). Just as the first participant, this person had a relatively high score on the idealism dimension. But for him, the high score on idealism was combined with a relatively high score on particularism, which indicates a tendency towards "situationism" (see Table 4.1). In other words, this participant derived his conclusions on moral questions from context rather than moral principles.

Elsewhere, I have argued that absolutists are driven by the question "How to realize the ideal world?" and situationists by the question "What is (morally) expected in this context?" (Doorn: 2010b). This distinction was also visible in answers given by the absolutist and situationist participants. Whereas the answers of the typical absolutists referred to some ideal (viz., the ideal of fairness), the answers given by the typical situationists were mainly focused on the "here and now" of this particular project. (This does not mean that the absolutists ignored the present project but rather that they were also concerned

about how to embed ethics in future projects *in addition to* giving opinions about the present project.) It is striking that two participants with a tendency to absolutism mentioned the need for a "broad societal discussion" about the desirability of Ambient Intelligence technologies in general (which is similarly related to the world outside this project). In a closing interview, this same absolutist participant argued for "ethics education," both parallel to the project and in the engineering curriculum. Whereas an other, situationist, participant considered ethics courses or training as maybe too far removed from the engineering work floor (that is, context), this absolutist participant was clearly in favor of teaching ethics to support engineers in their work.[88]

With the above discussion, I do not want to suggest that some participants gave better answers or showed morally superior behavior compared to other participants – most participants confirmed the relevance of the workshop and the concern for discussing ethical issues, so in that sense they all showed an awareness of the relevance of the ethical aspects. They rather used different types of arguments. They also differed in their opinion about which issues should be addressed first. Initially, the distributions of responsibilities as proposed by the different workshop participants varied therefore also quite significantly.

With the rationales interpreted as the empirical counterpart of the theoretical responsibility conceptions, the data confirm the scholarly philosophical literature on responsibility, where a relation was found between responsibility conception on the one hand and normative background theory on the other (Doorn: 2011). How do these data fit into the WRE model? Recall that the responsibility rationales or conceptions could be interpreted as the middle layer of Rawls' WRE model. This suggests that there is indeed coherence between the first (normative background theories) and second layer (moral principles) of the model. We could say that the normative background theories explain the type of arguments (i.e., the rationales) that are used by the different participants. The immediate follow-

---

[88] One of the reviewers correctly suggested that ethics education may also be an explanatory factor in this regard. Although I did not explicitly ask the participants in the workshop after their ethics education, I got the impression that only one of them had previously read ethics literature. Given that the introduction of ethics education at the technical universities in the Netherlands is of only recent date we can assume that the "senior" team members did not have a formal training in (engineering) ethics. Regarding the "junior" team members, most of them received their training outside the Netherlands (Eastern Europe and Latin America) and, as far as I am aware, ethics is not part of the engineering curricula in those regions.

up question is then how these rationales relate to the actual responsibility ascriptions (the third layer in Figure 4.1). It was found that the different rationales were not predictive for the responsibility ascriptions. People using the same type of arguments (for example, reference to the fairness of the responsibility ascription) may come to different responsibility ascriptions and, vice versa, people with different responsibility rationales might come to similar responsibility ascriptions. Hence, there is no correlation between people's responsibility rationales (second layer) and responsibility ascriptions (third layer). In political theory, it is generally seen as an advantage of Rawls' theory that it allows for different levels of abstraction. Disagreements on the more abstract levels do not undermine the possibility of agreements on the more concrete levels, or vice versa. McCarthy (1994), for example, argues that the more difficult it becomes to agree on general interests and shared values, the higher the required level of abstraction of the overlapping consensus. However, this case shows that it could equally well work the other way around. People can disagree on the abstract level on responsibility conceptions and use different arguments to distribute the responsibilities, they may still agree on the particular responsibility ascriptions (third layer), though on different grounds. It was witnessed that the participants converged to a consensus regarding how the responsibilities are to be distributed among the team members. Given the procedural nature of the WRE model, this is exactly the strength of the model, proponents like McCarthy argue. Notwithstanding the diversity in terms of responsibility rationales with rather different foci (consequences, fairness, tasks, duties, professionalism), the participants tended to be sensitive to one another's arguments. Although it proved difficult to attain consensus on all points, the opinions of the different participants tended to converge between the two "distributing exercises." Whereas the first distribution of responsibilities showed a significant scatter of tasks over different project activities and partners, the second distribution showed more responsibility for the project management and the clinical partners.

These data suggest that there is a complex interplay between the different layers of the WRE model. The implicit assumption in the WRE model is that people are prepared to revise their original opinions, which was also witnessed during the workshop. However, this raises the question how the participants are able to listen to one another's arguments when they have these different rationales. A tentative answer to this question is that the WRE model is a *coherence* model, in which every element is open for revision. In their

deliberation, people do not work top-down or bottom-up but rather seek coherence between the different layers. This not only means that none of the elements in the WRE has priority over other elements (and that all elements in the model are open for revision) but also that people are able to "include" arguments that are supportive for the coherence, even if those arguments do not correspond to their own argumentation style. In its extreme form, coherence includes the possibility that people are able to revise their normative background theories (the first layer in Figure 4.1). In practice, it is doubtful whether people will fully revise their ethics position, even if that would increase the coherence. People in favor of deontological ethics (cf. absolutism) will most probably not become full-fledged utilitarians. However, they may become more nuanced in how they interpret their normative background theories. An absolutist, for example, may become aware that context *does* matter, even though her original inclination is to deny context.

The fact that the normative background theories and the responsibility rationales are not predictive for the actual responsibility ascriptions indicates that moral reasoning is not a deductive process that goes one-directionally from the abstract to the more concrete levels. Also the people who have a low score on the particularism dimension (which suggests a top-down derivation from abstract to the concrete) did not take these abstract layers for granted but appeared to be able to reflect on them. One typical absolutist noted that there was indeed a feedback from the (concrete) discussions to his (abstract) theories and principles. Hence, the coherentist WRE model seems a more realistic model for interpreting moral deliberation than a deductive model that derives the outcome top-down from the normative background theories or moral principles to the actual responsibility ascriptions. The WRE model can explain how people, even when disagreeing on some remaining issues, may achieve a consensus on points that fit in each individual's WRE. When this coherence is achieved, the outcome can be considered fair or justified. As such, the outcome of the WRE approach is more stable than a negotiated compromise.

So far, I have interpreted the responsibility ascriptions and the deliberation concerning the responsibilities in terms of the WRE model. I showed how the normative background theories can explain people's rationales for distributing responsibilities. Taken as a whole, the WRE model is a powerful descriptive device, which can show the relations between people's different layers of morality. However, the case presented in this chapter is relatively stable and the question arises how the model performs in other contexts. In situations where

the cooperation is more vulnerable, the cooperation may become more strategic in nature. One could think of a controversial research topic (for example, the development of military equipment) or a situation where the internal competition between the institutional partners is high. In the ALwEN project, the cooperation between the SMEs was made more stable by working in a "pre-competitive" stage. When the competition between the partner institutions represented in the team is high, the cooperation becomes vulnerable because people may encounter a tension due to conflicting loyalties. This also holds for the type of partners involved. Larger industrial companies may be more inclined towards strategic cooperation. Dependent on the composition of the research team and the research topic, the cooperation may be more or less following the network cooperation paradigm criticized by Glenna or the actor-centered paradigm described in this thesis. It is worth investigating this hypothesis further.

4.5.   Conclusions

In this chapter, I have explored different actors' rationales for distributing responsibilities in an R&D project and interpreted these in terms of Rawls' WRE model. On the basis of a series of interviews and a workshop with project members, I discussed different tasks related to the social acceptance of the technology under development. This study indicates that there is a correlation between the actors' ethics position and their responsibility rationale. Actors with a similar ethics position referred to the same type of arguments. The actors' ethics positions were found not to be predictive for the actual responsibility ascriptions. Rawls' WRE model proved a powerful descriptive framework for analyzing moral deliberation. The data suggest that people seek coherence between different layers rather than deductively deriving their concrete opinions on moral issues from more abstract normative background theories and principles. It proved that people's normative background theories can explain people's rationales for distributing responsibilities.

On a methodological level, the operationalization of the WRE model proved a useful first step for applying the model in real situations. Validated instruments from moral psychology were used for assessing the data, which allowed for a minimum of speculative (third-person) interpretations.

The central theoretical contribution of this case relates to the different levels in moral deliberation, in particular the relation between the normative

background theories and the responsibility rationales and the fact that people seek coherence between the layers rather than work through them one-directionally. By distinguishing between rationales for distributing responsibilities and the actual distributions, possible sources of misunderstanding can be identified. The benefit from acknowledging these different rationales is that it enables actors to recognize the legitimacy of other people's opinions, ultimately contributing to a responsibility distribution that is both complete and accepted by all as justified.

Appendix

The workshop agenda consisted of five parts. Before start of the workshop, the participants were asked to fill in the Ethics Position Questionnaire. Afterwards, the participants were asked to individually reflect on the session and the end result.

Workshop program:
- Distribution of responsibilities (I): In what phase of the research, design or use of the technology should the parties involved address the moral tasks (i.e., tasks related to risks and social acceptance)? The participants are allowed to ascribe two types of responsibilities: primary responsibility (P) and secondary responsibility (S). The latter is defined as "relevant for this particular activity but not the activity where the primary responsibility lies." The participants are also allowed to say that a particular task was outside the scope of the present project. The participants are asked to distribute the tasks individually.
- Conditions for responsibility: What conditions need to be fulfilled before we can ascribe a particular responsibility to a person? The conditions are discussed online and afterwards ranked by each participant individually.
- Online anonymous discussion on the basis of disagreements about responsibility distribution (I).
- Distribution of responsibilities (II): Similar to Part 1, but now informed by discussion and "intermezzo" on responsibility conditions.
- Oral discussion of the last disagreements about responsibility distribution (II).

5   A procedural approach to distributing responsibilities[89]

5.1.   Introduction

Technological research is increasingly carried out in networks of organizations with different kinds of actors involved. These networks often lack a strict hierarchy and a clear task division (cf. Callon et al. 1992; Rogers and Bozeman: 2001; Saari and Miettinen: 2001). Consequently, decisions are subject to negotiation instead of top-down decision making. This increases the likelihood of the problem of many hands, which is the difficulty, even in principle, to identify the person responsible for some outcome (Bovens: 1998; Thompson: 1980). The occurrence of this problem in Research and Development (R&D) is especially undesirable since the introduction of technologies can be accompanied by risks and unforeseen side-effects as well, often with high impact (e.g., the use of asbestos, CFCs, DDT, nuclear waste and the greenhouse effect). If no-one is responsible for addressing these issues, the implementation of technologies might result in harmful consequences for society.

Research has shown that the problem of many hands can be partly traced back to different views on responsibility (Doorn: 2010b; this thesis, Chapter 4). In a pluralist society, people have different views on what responsibility amounts to and under what conditions one is responsible. Whereas some people defend a virtue ethical approach to responsibility, others take a deontological or consequentialist stance (see, e.g., Nihlén Fahlquist: 2006b; Williams: 2008; or Goodin: 1995 for a discussion of some of these approaches). Responsibility conceptions can differ in at least two ways. First, people may have a different understanding of what responsibility actually *means* (e.g., giving an account of something, to compensate for potential loss, to have a task to do something, to take care of something). Secondly, people may have different conceptions on

---

when a person *is* responsible. In this chapter, I focus on the second type of diversity: diversity in opinions on when a person is responsible.

The different conceptions may lead to different distributions of responsibilities. People defending a virtue ethical approach to responsibility, for example, may consider it a researcher's responsibility to show the merits of a technology to the broader society, whereas people with a more duty-based conception of responsibility may think in terms of a formal task description and consider this particular responsibility not to be part of that description.

In order to do justice to this pluralism of responsibility conceptions, there is a need for a distributing procedure that leads to a workable agreement but that, at the same time, leaves room for different responsibility conceptions without favoring any one in particular. Simply distributing on the basis of majority rule is potentially unfair to groups representing minority views.

In this chapter, I develop an approach that is based on procedural political theory. The underlying thought is that people do not have to agree on substantive conditions which tell when a person is responsible as long as they agree on the *procedure* for distributing the responsibilities (and given that they have a shared understanding of what responsibility means. The latter is important to prevent people from talking at cross-purposes). If such a procedure, or its outcome, is accepted by all people involved as representing the "fair terms of cooperation," this might help reconcile the pluralist responsibility conceptions and, ultimately, alleviate the problem of many hands. In order to test the applicability of political theory to responsibility distributions, the model of procedural justice is applied to a real case. The guiding question is whether a procedural approach contributes to reconciling the pluralist responsibility conceptions.

The outline of this chapter is as follows. Following this introduction, I sketch a procedural approach to justice based on Rawls' political liberalism. After explaining the approach, I describe two procedural norms that are derived from policy and innovation theory. Subsequently, I apply the approach to an empirical case in order to see whether the method contributes to reconciling the pluralist responsibility conceptions. In the final section, conclusions are given together with recommendations for further research.

5.2.   A procedural approach to justice: Rawls' political liberalism

Professional responsibility[90] and the distribution thereof is a topic that has gained increasing attention in recent years. Not only the scholarly literature on professional ethics but also professional settings themselves often reveal a large variety in responsibility conceptions. If we look at moral responsibility in general, we already see many different perspectives. Adherents of virtue ethics or care ethics, for example, emphasize the agent's character and the morally relevant features of a situation, herewith trying to answer the question what a responsible person in this situation would do (Ladd: 1991; Van Hooft: 2006). The main question in duty ethics is what the agent's duty is and what rules she should follow (Van Hooft: 2006, 9-17). Yet another approach is a consequentialist conception of responsibility, such as defended by, for example, Goodin (1995). According to this approach, responsibility should be conceived as largely a matter of result-oriented tasks. If we translate these different perspectives to the notion of *professional* responsibility, the differences are at least as pressing. Professional responsibility in duty ethics, for example, is often defined in terms of preventing wrong-doing. Others define professional responsibility in terms of professional codes and standards (Davis: 2002), licensing (Neuman: 1991), virtues (Oakley and Cocking: 2001; Harris 2008), knowledge (Alpern: 1983; Brannigan 2005), or even the "existential pleasure" of carrying on one's profession (Florman: 1976).

In professional networks, these different perspectives may all be represented by the different actors constituting the network. Moreover, if we recognize the political ideal of pluralism, more than one of these different perspectives may be *legitimate.*[91]

---

[90] Following the definition of profession in note 4, we could distinguish between profession and occupation, and consequently, between professional responsibility and occupational responsibility. Since I take the people in the ALwEN team to be members of a profession, I use the term professional responsibility and not occupational responsibility.

[91] As explained in Chapter 1, pluralism can be understood as the acknowledgment of diverse and competing values and visions of the good life. It is assumed to be the cornerstone of democracy because it distributes power over multiple centers, herewith countering authoritarianism (Dryzek and Niemeyer: 2006). According to the pluralist thesis, conflicting private values cannot simply be reduced to single public values.

In R&D networks, this pluralism in responsibility conceptions leads to the problem of how to distribute responsibilities. Since professional networks often lack strict hierarchical relations, decision making is done on the basis of mutual negotiations rather than top-down decision making. It remains therefore open how responsibilities should be distributed. In the ALwEN project, a team of engineers was involved. Even if people would agree what this team's joint responsibility involves (e.g., the task to prevent certain risks stemming from a technology), it is not obvious how this responsibility should be distributed among the engineers constituting the research team. Should it be done in as early a stage of technology development by the team member doing fundamental research or by the team members commercially exploiting the technology? The answer to this question is partly dependent on the responsibility conception one endorses. The pluralist thesis implies that the diverse and competing visions of responsibility cannot be reduced to one overarching conception. Hence, people should somehow find a consensus concerning how responsibilities are to be distributed. However, what counts as a *justified* consensus remains open; not any consensus will do. Even in the absence of a strict hierarchy, power relations may still be present. Critics of consensus policy often warn that the promotion of consensus is coercive, notwithstanding its democratic aims. The promotion of consensus runs the risk of negotiating the interests of the most powerful. If one actor defends a virtue ethical approach to responsibility but agrees to distribute responsibility according to tasks in order to gain something else in return, it is questionable whether the agreement counts as a *justified* consensus. In order to assess which kind of consensus can be considered justified (where justified is understood as "doing justice to pluralism without favoring one view over the other"), we need a framework that incorporates both the ideal of consensus and that of pluralism.

In political theory, the idea of procedural justice has emerged as a way to provide such a framework. The term procedural justice refers to the way procedures (e.g., decision making procedures) are structured so that their outcomes can be considered fair. The term is especially relevant in pluralist societies where people often cannot agree on substantive views on what justice amounts to. An example of procedural justice is the principle that those who are affected by a certain decision be afforded the opportunity to participate in the decision making.

In recent decades, different solutions have been proposed to find a workable middle ground between the ideals of consensus and pluralism, all balancing substantive views on justice with procedural requirements. A highly developed and differentiated procedural political theory is Rawls' political liberalism.[92] Rawls attempts to propose the formal conditions under which the decision making can be deemed fair. His theory is particularly attractive because it provides both an elaborated *justificatory* framework and a *constructive* framework for encouraging reflection (Doorn: 2010a).

Central in Rawls' theory are the concepts of overlapping consensus and wide reflective equilibrium (WRE). Rawls' aim was to develop a criterion of justice that would be agreed upon by all under conditions that are fair to all (JaF 15). Although Rawls at first wanted to develop a substantive theory of justice for a relatively homogeneous well-ordered society, he revised this idea of a well-ordered society in his later work. Recognizing the permanent plurality of incompatible and irreconcilable moral frameworks within a democratic society, he introduced the concept of overlapping consensus. People are able to live together despite conflicting moral values and ideals as long as they share a moral commitment to the society's basic structure.

People with different comprehensive doctrines must be able to justify for themselves the acceptability of the claims of political justice (RH 143, ToJ 17, PL 28). Rawls introduced the idea of reflective equilibrium to refer to this individual justification. In this idea, a distinction is made between three levels of considerations: (1) considered moral judgments about particular cases or situations, (2) moral principles and (3) descriptive and normative background theories. Assuming that all people want to arrive at a conception of justice that yields definite solutions and that is complete, in the sense that it is more than a mere collection of accidental convictions, people should aim at coherence between the considerations at the different levels. We speak of *equilibrium* if the different layers cohere and are mutually supportive; it is called *reflective* if the equilibrium is arrived at by working back and forth between the different considerations and if all are appropriately adjustable in the light of new situations or points of view; and it is called *wide* if coherence is achieved between

---

[92] I do not want to suggest that Rawls' theory is the only procedural theory. Deliberative democracy, such as defended by, among others, Cohen (1989; 1997) and Elster (1986; 1998), is an other example of a highly developed procedural theory. The concept of deliberation can also be linked to the work of the German philosopher Jürgen Habermas (1990a).

all three levels of considerations and not only between the considered judgments and moral principles (in which case we speak of *narrow* reflective equilibrium). Although people with different comprehensive doctrines might arrive at different WREs, they likely have an overlap when it comes to the basic principles of fairness. This "shared module" is what Rawls calls the overlapping consensus (JaF 32).

Dryzek and Niemeyer argue that, notwithstanding its focus on formal structure, the establishment of an overlapping consensus still requires agreement on the substantive values underlying the procedure.[93] The management of pluralism requires a shared tradition (such as liberalism) or a shared set of values to acknowledge the legitimacy of other comprehensive doctrines (Dryzek and Niemeyer: 2006, 636). As such, the approach seems still biased towards liberalism. However, Rawls makes a distinction between different forms of justification, allowing some to be more substantive than others. The complete idea of *justice as fairness*[94] will most probably not be part of shared WRE but in a plural society it can still be endorsed by reasonable comprehensive doctrines as a *political* conception of justice, that is, as a basis of

---

[93] A similar criticism comes from Habermas, who argues that Rawls introduces a particular conception of the moral person into his theory. According to Habermas, it is especially the sense of fairness and the capacity of the good which are in need of prior justification (Habermas: 1995, 112).

[94] The term *Justice as Fairness* is used by John Rawls to refer to his distinctive theory of justice in which he developed two principles for organizing modern welfare state. The first principle, known as the equal liberty principle, states that each person is to have an equal right to the most extensive basic liberty compatible with similar liberty for others. The second principle describes two conditions that are to be satisfied in case of social and economic inequalities: (a) The inequalities are to be attached to positions and offices open to all under conditions of fair equality of opportunity (the fair equality of opportunity principle); and (b), The inequalities are to be to the greatest benefit of the least advantaged members of society (the difference principle) (JAF 42/ PL 5-6). At the philosophy seminar of the Royal Institute of Technology (Stockholm), where I presented a draft version of this chapter, I was rightly pointed to the fact that Rawls derives fairness from his veil of ignorance, which assumes that people do not know which comprehensive doctrines they adhere to. Under this condition of ignorance, justice implies fairness and vice versa. This means that only for the first kind of justification, the terms justice and fairness could be used interchangeably. However, in order to be consistent with the responsibility terminology and everyday language, I take a more lenient stance and use the term fair or fairness also to refer to the outcomes of the other types of justification (where Rawls would probably prefer the term justified rather than fair; the same holds for the "appropriateness" of responsibility ascriptions, which he would probably judge in terms of being justified or not rather than being fair or not).

social unity in a constitutional democracy with a plurality of reasonable but incompatible – religious, philosophical and moral – doctrines. People with divergent comprehensive doctrines can overlap in their acceptance of a conception of justice, because they are likely to share at least some beliefs about reasonable pluralism. They do not have to agree on all particular decisions but they do agree on "principles of fairness" related to the political realm, which get shape as the society's basic institutions. Being the focus of an overlapping consensus, these principles specify the fair terms of cooperation among citizens and the conditions under which a society's basic institutions can be deemed just (PL 133). Rawls calls this *pro tanto* justification, which draws on public reasons or arguments only (i.e., values, judgments, principles and background theories valid for the public domain). It is done "without looking to, or trying to fit, or even knowing what are, the existing comprehensive doctrines" (RH 145). An individual citizen can then try to fit this political conception of justice into his own comprehensive doctrine. This is what Rawls calls full justification, which is carried out by an individual citizen as a member of civil society and in which the citizen accepts a political conception and fills out its justification by embedding it in his own comprehensive doctrine. This latter justification does not require adherence to liberalism.

Central in most liberal theories of justice is the notion of "public reason." This holds for Rawls as well. Compared to, for example, deliberative democracy theorists – and Habermas in particular – Rawls has a restricted notion of public reason. Habermas, for example, defends a conception of public reason which includes all unofficial arenas of public discourse; these unofficial arenas in fact ground democratic self-government and political autonomy (McCarthy: 1994, 49). For Rawls, however, public reason is limited to the official institutions. Since his procedural approach to justice aims at "uncovering a public basis of justification on questions of political justice given the fact of reasonable pluralism," it should proceed from "what is, or can be, held in common; and so [...] begin from shared fundamental ideas implicit in the public political culture in the hope of developing from them a political conception that can gain free and reasoned agreement in judgment" (PL 100-101). Hence, the function of public reason is not so much to be critical but rather to be constructive. Public reason, therefore, needs to start from shared ideas and organize those into a political conception that can serve as the focus of an overlapping consensus, which in turn can enhance stability. Rawls connects his conception of reasonableness to T.M. Scanlon's principle of moral motivation, which is one of the basic

principles of contractualism (Scanlon: 1982, 104, 115). The principle tells us that we have a "basic desire to be able to justify our actions to others on grounds they could not reasonably reject" (PL 49-50; fn 2).

Similar to deliberative approaches in Technology Assessment (TA), which are based on deliberative democracy procedural theory, Rawlsian concepts have also found their way to more applied contexts. Especially in the context of applied ethics, the tension between diverging moral frameworks is an urgent problem; for example, how to sustain the conditions of the good life in a globalizing world (Dower: 2004; Hardin: 1999), or how to decide on issues related to abortion (Little: 1999) or living organ donation (Hilhorst: 2005). Rawlsian approaches seem promising for answering these kinds of questions since they offer a methodological alternative to the extreme positions of ethical generalism and (specified) principlism on the one hand and particularism on the other (Daniels: 1996; St. John: 2007; Van den Hoven: 1997).[95] Rawlsian justification avoids the drawbacks of both extremes because it aims at coherence between the abstract theoretical principles and the more particular considered judgments without giving priority to any of them. As such Rawlsian approaches seem to offer a promising decision making procedure within applied ethics.[96] Especially the concept of WRE seems an attractive method for real-life justification. Even without reference to political concepts as overlapping consensus, the notion of

---

[95] Proponents of the first category argue that applied ethics is essentially the application of general moral principles (Beauchamp and Childress: 1994, 112; DeGrazia: 1992; Lustig: 1992) or theories (Gert et al.: 1997; Hare: 1988) to particular situations. This position is criticized for mistakenly assuming that valid principles can be formulated that govern all rational persons. Moreover, the critics argue, procedures for deducing answers to moral questions is impossible, unnecessary, and undesirable. These critics argue for situational adequacy, that is, an ideal of doing justice to persons in a particular historical context. The problem with particularism, on the other hand, is that it runs the risk of lacking moral justification. In most situations where ethical reflection is at stake, people should be able to justify their actions in terms of moral principles. However, if particularism is carried through to the extreme, it becomes difficult to provide public justification of moral judgments (Van den Hoven: 1997, 240-241).

[96] In addition to this justificatory application, Rawlsian approaches are sometimes used in a constructive way as well. In the latter case, they are used as a framework for structuring discussion and debate, with the aim of coming to a justified agreement. The method could then be used, for example, as a *means* to attain a coherent basis for decision making in ethical committees or to gain support for particular decisions in the context of public policy (Holmgren: 1987). This second way of applying Rawlsian approaches is comparable to the constructive application of deliberative approaches, albeit the Rawlsian ones take the moral background theories and principles more explicitly into account.

WRE can help explain why people consider certain things fair or unfair. Moreover, since the Rawlsian approach takes the different layers of morality explicitly into account, the approach seems to provide a powerful tool for encouraging reflection. The criticism that the Rawlsian procedural approach to justice requires that people share the tradition of liberalism does not seem valid. It is sufficient that people acknowledge that reasonable pluralism is the permanent condition and that the concept of reasonableness replaces that of moral truth. This is not the same as sharing the comprehensive view of liberalism. In a professional setting where people are motivated to work towards a fair distribution of responsibilities, this demand of "reasonableness" is probably a realistic one. McCarthy (1994) argues that it is a strength of Rawls' theory that he allows different levels of abstraction. The more difficult it becomes to agree on general interests and shared values, the higher the level of abstraction of the overlapping consensus. However, it could also work the other way around; in case of responsibility distributions, people can disagree on the abstract levels of responsibility conceptions and principles, but agree on particular responsibility ascriptions.

In the next section, I develop this Rawlsian procedural approach further to assess the fairness of responsibility distributions.

5.3. Procedural fairness in responsibility distributions: two norms

In their paper on reflective equilibrium in R&D networks, Van de Poel and Zwart (2010) derive two procedural norms that follow from applying the Rawlsian method of WRE to actual cases: reflective learning and inclusiveness. According to the authors, these norms, which are also used in the literature on policy and innovation networks, contribute to achieving a justified overlapping consensus. Before explaining the relation between these norms and the procedural approach, I first discuss the two norms in somewhat more detail.

5.3.1. Reflective learning

Since the last decades, several interactive and participatory methods have been proposed to successfully implement and develop new technologies (where successful is understood as "sustainable," "responsible," or some other desirable adjective). Most often these processes are shaped and evaluated in terms of the

degree of learning experienced within the network or organization of relevant actors.

Most scholarly literature on learning goes back to the work of Fischer (1980; 1995) and Schön (1983). Fischer conceptualized his "levels of argumentation" (he does not refer to learning or reflection explicitly) within the context of policy making. Schön refers to the professions of engineering, architecture, management, psychotherapy, and town planning to show how professionals meet challenges by engaging in a process of "reflection-in-action."[97] A distinction is generally made between two levels of learning or reflection: lower-order versus higher-order discourse (Fischer: 1980) or reflection (Schön: 1983), single-loop versus double-loop learning (Argyris and Schön: 1978; Sabatier and Jenkins-Smith: 1993) or adaptive versus generative learning (Senge: 1990). Although the contexts and the exact definitions differ, the distinction between the two types of learning in all cases is more or less similar. In the lower-order category, the learning process is a kind of technical or instrumental learning. It is reactive, short-term focused, within a context of fixed objectives (as applied to policy), a context of fixing new problems within the same problem definition and procedures (as applied to organization), or a context of technological design optimization (Brown et al.: 2003; Hoogma et al.: 2001). In the higher-order category of learning, the objectives, problem definitions and procedures are not tested but questioned and explored (Hoogma et al.: 2001). It therefore involves the redefinition of policy goals and changes in norms and values (Brown et al.: 2003). This higher-order learning is also more long-term focused. In the remainder of the text, I will use the term "reflective learning" to refer to these higher-order learning processes.

The effect of learning in organizations can be conceived as a threefold shift (Brown et al.: 2003): (1) a shift in framing of the problem; (2) a shift in principle approaches to solving the problem and in weighing of choices between alternatives, and (3) a shift in the relationships among actors in a professional

---

97   Note that this is a sociological definition of profession and not a normative one, such as proposed in the professional ethics literature (cf. the definition given by Michael Davis, in which profession is defined as "a number of individuals in the same occupation voluntarily organized to earn a living by openly serving a certain moral ideal in a morally-permissible way beyond what law, market, and morality would otherwise require" (Davis: 1997, 417; 2002: 3). Although, both Schön and Davis emphasize the importance of the professional community, the explicit normative element in Davis' definition is missing in Schön's account.

network as well as the broader sphere.[98] It is especially this third shift (a shift in the relationships among actors) together with the object of reflective learning (appreciative systems and overarching theories) which makes reflective learning such an important phenomenon in the context of responsibility distributions. In the discussion of procedural justice, it was explained how WRE can be used to decide on issues in a context of reasonable pluralism (i.e., in a situation with diverse and competing interests). Reasonableness requires that people recognize the legitimacy of other actors in the network with other moral views. Lower-order learning occurs when people become aware of their position in the network and the possible differences in actor roles, agendas, perceptions, values and interests among the actors. The awareness of these differences enhances the instrumental rationality of the actors in the sense they realize that the other actors enable or constrain the achievement of certain goals (Van de Poel and Zwart: 2010, 181). In case of reflective learning, actors are not only aware of these differences but they also recognize the *legitimacy* of these other views. Reflective learning therefore includes reflection on the desirable properties of the network as a whole. Additionally, it might help distinguishing between private and public values, that is, between arguments that are and that are not legitimate or important for an actor fulfilling a specific role in the network. Reflective learning might thus contribute to achieving an overlapping consensus concerning a fair distribution of responsibilities among actors within a network displaying a large variety of value systems and background theories (ibid.).

5.3.2. Inclusiveness and openness

The second norm that Van de Poel and Zwart (2010) distinguish is "inclusiveness" or "openness," which can be described as the norm that all relevant actors are included in a network. Van de Poel and Zwart explain which actors can be considered a relevant actor in terms of the Rawlsian criterion of public reason. Each actor that can legitimately claim to have a "reasonable stake" or a "reasonable interest," where reasonableness means that it can be argued upon on the basis of public reasons, can be considered a relevant actor. Since this point of relevance will probably always be a point of debate, the authors add

---

[98] Brown et al. refer to "professional network," but following Michael Davis' definition of "profession" the term "occupational network" seems better in place because the network does not necessarily consist of people from one profession only.

the criterion of openness, which serves to warrant the possibility that new aspects become relevant (182).[99] However, openness has an additional, more institutional feature. The criterion of openness calls for an open discourse, which means that it is not only important that all relevant actors are included, but that they have equal opportunities for participating in and contributing to the decision making process as well. If a group of actors with different fields and levels of expertise are engaged in a conversation, it is important that the vocabulary used by the experts is understandable to all. The criterion of openness also requires that people feel free to bring in unwelcome arguments. If some actors are discouraged to do so and remain silent, the overlapping consensus that is arrived at cannot be justified as being fair. Together, inclusiveness and openness determine when an overlapping consensus can be considered fair. They prevent "unjustified shortcuts to a wide reflective equilibrium or overlapping consensus" (ibid.). The latter could be the case when people with unwelcome arguments are excluded from the network.

Van de Poel and Zwart exert on explaining why this notion of *justified* overlapping consensus does not imply that they smuggle in some substantive notion of public reason. As explained in Section 5.2, critics of consensus theory argue that, under the sway of deliberation, the goal of consensus can all too easily be equated with the interests of the powerful (cf. Mouffe: 1999, 2000; Young: 1996, 2000). Hence, we can understand that not any consensus is a democratic outcome. In other words, we cannot avoid introducing some criterion to distinguish a valid consensus from an invalid one. Although Young goes further (in that she doubts every instance of consensus), Rawls would probably agree that reference to consensus requires due care in order to distinguish it from a mere compromise or *modus vivendi* (JaF 191). In case of the latter, people come to an agreement on the basis of some negotational process in which power relations and mutual dependencies play a crucial role. For the actors, the outcome may be a satisfactory one; they decide so on the basis of pragmatic and sometimes prudential reasons. However, even if all actors agree, this does not equate such a bargained compromise with a morally justifiable

---

[99] This resembles the Habermassian understanding of justice as an ongoing exercise of political autonomy, which is always incomplete and subject to shifting historical circumstances (Habermas: 1995, 131). For Habermas, no conception of justice can ever be final and some questions should therefore explicitly be left open (118). The composition of the network seems a plausible instance of such an "open" question.

consensus. The procedural criterion of inclusiveness and openness provides justificatory force to the consensus that is achieved in the network. The fact that this criterion is based on a certain notion of public reason is not so much problematic but rather an intrinsic element of the method. Managing pluralism requires a distinction between "valid" and "nonvalid" reasons, between "public" and "nonpublic" ones. To base the demarcation on a notion of public reason that others "could not reasonably reject" – to use Scanlon's wording – seems a strength rather than a weakness.[100]

5.3.3. Relation between the two norms and fairness in responsibility distributions: sufficient or necessary conditions

With the two procedural norms described in the previous sections, we can now analyze whether these two norms are indeed beneficial to reconciling different responsibility conceptions. This requires a series of steps. The first is to see whether people can agree on a distribution of responsibilities and are able to give a *pro tanto* justification. If that is the case, we have achieved a consensus. The next step is then to see whether this distribution of responsibilities is also coherent with everyone's individual conception of responsibility; in other words, whether it fits within each individual's own WRE. If that is the case, we can speak of a *justified* overlapping consensus of the responsibility distribution.

Although Van de Poel and Zwart say that the two procedural norms are contributory to getting a justified overlapping consensus, their description of the norm of inclusiveness suggests that at least this norm is a necessary one (and not just contributory); without the norm of inclusiveness being fulfilled, no responsibility distribution can be justified as procedurally fair. However, although fulfilling this norm is necessary, it is probably not sufficient. People also have to recognize the legitimacy of other actors' opinions and the need to justify their own standpoint in terms of public reason. To account for the latter,

---

[100] In the Introduction and in Section 5.2, I explained how Rawls interpretation of public reason differs from Habermas' notion of public reason in that the former is more restricted. Since Rawls' notion is assumed to be more *constructive* and Habermas' notion more *reconstructive* (Habermas: 1995, 131), it is probably dependent on the field of application which interpretation could best be applied. In case of establishing a distribution of responsibilities that is justifiable to all actors involved, Rawls' constructive notion seems more adequate. However, if one wants to organize participatory meetings in which deliberation serves to map out divergence in opinions, Habermas' reconstructive notion seems more adequate.

reflective learning processes may indeed be contributory. If there is a direct correlation between an agent's responsibility conception and what she considers a "fair" responsibility distribution, reflective learning is not just contributory but even necessary. In the case described in the next section, I will analyze whether this correlation is indeed present.

5.4.   Discussion

In this section, I analyze the ALwEN case in terms of the two procedural criteria developed by Van de Poel and Zwart. For a description of the case, the reader is referred to Section 4.3.

When analyzing the empirical results, we have to keep in mind that several things run together. First, the ethical parallel research itself probably has some effect on the way the research is carried out and how the different responsibilities are distributed. The technical researchers are probably more attentive to moral issues due to the presence of an ethicist at their project meetings. Secondly, the workshop was structured along the lines of the Rawlsian WRE approach so that the different elements in the workshop were not only used to *assess* the individuals' moral opinions but also to *encourage reflection* (see also Section 3.3 and note 17 on the dual use of Rawlsian approaches). When we try to analyze the resulting distribution of responsibilities in terms of the Rawlsian procedural framework and try to see whether this approach does indeed reconcile the tension between the different conceptions of responsibility, it is somewhat difficult to separate the effect of the workshop itself from the effect of the procedural approach. However, notwithstanding these multiple effects, we can still derive some interesting points from the ethical parallel research and the workshop.

First, the workshop prompted discussion on the distribution of responsibilities in the project. In their evaluation of the workshop, most participants indicated that they had become more aware of certain moral issues (e.g., the need to involve end users). There was a general agreement that most moral issues span several activities within the project and that it is therefore difficult to single out one activity where it should primarily be addressed. The primary responsibility was in those cases ascribed to the project management for coordinating this joint effort, to the experimentation phase where all activities were supposed to come together, or to the clinical partner. Some participants explicitly mentioned that this workshop made them realize that some moral

issues were currently not addressed adequately. The idea that the work should shift from research towards either laboratory or clinical experiments with a (prototype) application was shared by all. Soon after the workshop, a brainstorm meeting was scheduled in which the requirements for clinical experimentation were discussed in more detail. Hence, one effect of the workshop was certainly to pay more attention to the end users and to involve them in the research.

Secondly, although the participants endorsed rather different conceptions of responsibility with different foci (consequences, fairness, tasks, duties, professionalism), they tended to be sensitive to one another's arguments. Although it proved difficult to attain consensus on all points, the opinions of the different participants tended to converge between both "distributing exercises" (remember that the participants were asked to distribute the responsibilities twice, with a discussion in between). Whereas the first distribution of responsibilities showed a significant scatter of tasks over different project activities and partners, the second distribution showed more responsibility for the project management and the clinical partners. Since all discussions and responsibility ascriptions were done anonymously, this can be considered a genuine convergence and not the result of group pressure. The participants were also asked about the fairness of the resulting distribution of responsibilities. Interestingly, although the participants perceived the end result in rather different ways, they all seem to interpret the end result more or less as a consensus on how the responsibilities are to be distributed. Some interpreted the outcome of the workshop as the insight that the "ethics" of the project is, in the end, a joint effort, whereas others interpreted it as primarily a responsibility of the clinical partners or the project management to *coordinate* this joint effort. However, all participants agreed that, in the end, all project members should have a commitment to the project as a whole (including the moral aspects).

Thirdly, when asked whether the workshop would affect the work in the project, most participants indicated that it would indeed have implications for their work, though for some only minor ones. All participants expected a shift in focus from research towards either laboratory or clinical experiments with a (prototype) application. One participant expected that the enduring impact of the workshop would be to make more explicit what the project in fact aims for. Before the workshop took place, the goal of the project was still rather ill defined. Additionally, the opinions on what is part of the project became clearer and also more inclusive. Some researchers initially considered most moral issues as being beyond the scope of the present project. However, during the discussion

and in the second "distributing exercise," most issues were included in the scope of the project, with a central role for the project management.

When we assess the project in terms of the two procedural criteria developed by Van de Poel and Zwart, we can identify the following points. First, both levels of learning seem to occur. The various participants' remark that they became more aware of ethical issues is a clear sign of first-order learning. However, the discussions indicate that this workshop prompted second-order learning processes as well. Some senior participants worried about the fairness of the load for the PhD and postdoctoral researchers, which indicates an openness to other people's interests. Moreover, the emphasis that the work requires a *joint* effort, spanning all the project activities, also points to (second-order) reflective learning processes. Lastly, the fact that the problem definition itself became object of discussion is also an indication of reflective learning.

In terms of inclusiveness, the project clearly aims to be inclusive. It was deliberately chosen to include a clinical partner in the project as well, herewith attempting to make the project more than just a technological project. However, the cooperation between the technical partners and the clinical partner proved difficult in practice. During the workshop it was also mentioned that the user involvement was in fact rather weak. In that sense, the project was less inclusive than aimed for at the start. However, soon after the workshop, more tangible attempts were made to include end users. Since the researchers sincerely aimed at openness and inclusiveness and since they did not raise formal obstacles for including more people, we can conclude that this criterion is, at least partly, fulfilled.

What does the foregoing teach us about the necessity of the two procedural norms: are these norms indeed required? Regarding inclusiveness, the answer is obviously yes. If the criterion of inclusiveness is released, the method loses its justificatory force. In practice, it will be difficult to involve all relevant people in the decision making directly. However, in a case such as the current project, the interests of those people that are affected by the technology should at least be represented. If we look at the end users, for example, it is important that their interests are looked after. Even though they do not have to be involved in the actual division of labor, the ultimate division of labor should include the task to look after their interests. So, though indirectly, they should be included or represented in the decision making process.

The second norm is learning. During the workshop, it was investigated to what extent the moral background theories (Column 2 in Table 4.2) were predictive for the actual distributions of responsibility (Column 4 in Table 4.2). The empirical findings of the case suggest that there is no correlation between these two layers of morality. People with similar moral background theories might come to different responsibility distributions and people with different moral background theories might come to similar responsibility distributions. This suggests that reflective learning (here, a willingness to change one's moral background theory) is not required to come to a similar distribution of responsibilities. However, without reflective learning, people will probably not recognize the legitimacy of other people's arguments in the first place. So, reflective learning is probably still required to agree on the possibility and legitimacy of disagreement. People do not have to change their own conception of what responsibility amounts to, but they do have to acknowledge that their conception is one among many. In the empirical case, reflective learning processes were present, especially in the discussion of the fairness of responsibility ascriptions. It is questionable whether the outcome would have converged as it did now without these reflective learning processes. This shows that both norms are indeed beneficial for getting a justified overlapping consensus and that the norm of inclusiveness is also required.

5.5.   Concluding remarks

In this chapter, I developed a procedure for distributing responsibilities based on Rawls' political liberalism. The procedural model was applied to a technological project that is currently being carried out. This project was studied as part of an ethical parallel study. An interactive workshop was organized to discuss the responsibilities for moral issues in the project. During the workshop, it appeared that the team members endorse a large variety of responsibility conceptions and rationales for distributing them.

The case shows that, in a pluralist setting, a procedural approach can be useful for prompting discussion on the legitimacy of the different conceptions and the question what a fair distribution of responsibilities amounts to. Although a full overlapping consensus regarding the distribution of responsibilities is probably too demanding, the case shows that the tension between the different conceptions can be alleviated by structuring the discussion along the lines of the different layers of the Rawlsian WRE approach because

this encourages participants to think in terms of "fair" workload and the legitimacy of other people's arguments. Although some differences in opinion remained, the effect of the workshop was that the work became more focused and that certain moral issues that were until then not recognized became part of the work. The two procedural norms (reflective learning and inclusiveness), as proposed by Van de Poel and Zwart, were both (partly) fulfilled.

Three points deserve further investigation. First, because the workshop was structured along the lines of the WRE approach, it is difficult to assess whether it is the workshop itself or the "procedural approach" that encourages reflection and alleviate the tension between the different responsibility conceptions. If the workshop was structured in a different way, not focusing on the different layers of morality, would the result have been the same? This question cannot be answered on the basis of this single case alone. Related to this point is the question whether the method should be applied in its full justificatory function or mainly as a constructive approach. Both questions need further research.

Secondly, the present case does neither confirm nor refute that (higher-order) reflective learning processes are indeed indispensible for recognizing the legitimacy of other people's conceptions. Reflective learning proved, strictly speaking, not a necessary condition: it may be theoretically possible to think of a situation where people commit to reasonable pluralism without any instance of reflective learning. However, in practice it is highly unlikely that people will recognize the legitimacy of other people's responsibility conceptions in the absence of reflective learning processes. Hence, although reflective learning is not logically necessary, in practice it probably *is* required.

Thirdly, due to the divergent interpretations of the final distribution of responsibilities, this final distribution cannot straightforwardly be interpreted in terms of an overlapping consensus or in terms of individual WREs. In that sense it is maybe somewhat artificial to talk about "procedural justice" in this context. The workshop did not explicitly derive or discuss procedural justice or cooperation norms. However, the fairness of responsibility distributions was explicitly discussed, including the question whether the final responsibility distribution could be considered fair. Together this seems a first step to deriving procedural justice norms.

More studies are needed for further developing the present approach to discuss responsibility. Remaining questions are the role of reflective learning processes and the different aims of the approach. My hypothesis is that, the more challenging the moral disagreements are, the more important these

reflective learning processes become and the more important it becomes to systematically touch upon the different layers of morality. Alternatively, a discussion might easily arrive at an impasse in which opposing opinions are merely expressed rather than being listened to.

6 Towards a new role for engineering ethicists[101]

6.1. Introduction

Arising in the early 1970s, the field of engineering ethics is relatively new within the broader field of applied or professional ethics. Due to a number of scandals, technology was being widely criticized by philosophers and non-academics alike (Durbin: 1997). Contrary to the field of bioethics, which became a flourishing branch of applied ethics, engineering ethics has not succeeded in providing philosophical foundations for their field until the late 1990s (Durbin: 1997, 79). Until then, engineering ethics focused on questions about alleged wrongdoing: who can be held responsible for technological failure or disasters.[102] The assessment of responsibility typically focused on questions related to liability and blameworthiness. However, this view on engineering ethics has increasingly attracted criticism for being too narrowly focused on wrongdoing and less on more positive standards of what responsible engineering requires (Durbin: 2008; Lynch and Kline: 2000; Pritchard: 2001; Pritchard and Holtzapple: 1997; Vanderburg: 2000).

In this "narrow approach" the topic of responsibility is often discussed in terms of the traditional concept of individual responsibility. However, by applying traditional individualistic concepts to the collective undertaking of engineering and technology development the conclusion of this assessment is often a negative one: none of the actors participating in the engineering enterprise can be fairly held responsible, because none of the actors has individually fulfilled all of the responsibility conditions.

The rationale for ascribing responsibility in the way mentioned above is related to blame. As a counterpart to this blame-oriented perspective, one can

---

[101] This chapter originally appeared as an article in the *Bulletin of Science, Technology & Society* 30 (3) 2010. The paper is co-authored with Jessica Nihlén Fahlquist.

[102] See the discussions of the Bhopal disaster, the explosion of the *Challenger*, or the sinking of the *Herald of Free Enterprise* in typical engineering ethics text books (Davis: 2005; Harris et al.: 2005[1995]; Martin and Schinzinger: 2005; Whitbeck: 1998).

also distinguish a more consequentialist perspective. In this case responsibility is not so much related to the fulfillment of certain criteria, but on the consequences that follow from assigning responsibility. Advocates of this latter approach argue that the scholarly literature on engineering ethics seems to be biased towards the blame-oriented perspective on responsibility rather than the consequentialist perspective, hereby overlooking the opportunities for incorporating ethical reflection *during* technology development (Pritchard: 2001). In this chapter, we describe an attempt to rethink engineering design and to put this "preventive" engineering ethics (Vanderburg: 2009) into practice.

The outline of this chapter is as follows. We first briefly elaborate on the different senses of responsibility. We distinguish between an outsider's perspective and an insider's perspective. In the subsequent section, we give a brief history of technology management, including the recent trend of ethical parallel or embedded ethical research. Using an example of the development of a new sewage water treatment plant, we show how the engineering ethicist, by engaging in the engineering practice, can improve technological design, both in functional terms and in moral terms. This requires a new role for the engineering ethicist. Rather than acting as a policeman, discussing from the outside what to do and what to avoid, the ethicist's role can be to focus on "how" technology is to be developed and elucidate issues that would otherwise be overlooked. This chapter ends with practical recommendations for technology development and technological design.

6.2. Holding responsible versus assuming responsibility

Few concepts in moral philosophy are more slippery than that of responsibility (Miller: 2004). Different authors have distinguished different taxonomies, most going back to the famous parable of Hart (1968) and the captain of the ship.[103] In

---

[103] The story runs as follows (with numbers in brackets corresponding to the taxonomy in the main text):

> As captain of the ship, X was responsible [5] for the safety of his passengers and crew. But on his last voyage he got drunk every night and was responsible [1] for the loss of the ship with all [others] aboard. It was rumored that he was insane, but the doctors considered that he was responsible [2] for his actions. Throughout the voyage he behaved quite irresponsibly [6], and various incidents in his career showed

a rather rough classification one could distinguish at least the following six senses of the term's use (more fine-grained taxonomies are possible as well; for the present purpose this classification will do):

(1)   Responsibility-as-causation: being the cause of some event.
(2)   Responsibility-as-capacity: having the capacity to initiate or prevent a situation or an event.
(3)   Responsibility-as-blameworthiness: being eligible for blame because what one did was wrong.
(4)   Responsibility-as-liability: being legally required to pay the damages/repair.
(5)   Responsibility-as-role: having the task to do something.
(6)   Responsibility-as-virtue: being a responsible person.

In the literature on moral responsibility the discussion is often about the third sense of responsibility: responsibility-as-blameworthiness. Because moral responsibility, in this sense, is related to the reactions of praise and blame (Fischer and Ravizza: 1993; Miller: 2004; Strawson: 1974), the ascription of responsibility is only warranted if the person is an appropriate candidate for reactive attitudes, that is, if the accompanying reactive attitudes and their consequences are merited or deserved (see Eshleman: 2008; Magill: 2000; Wallace: 1994; Watson: 1996; Zimmerman: 1988). For this reason, responsibility is supposedly only ascribed when certain conditions are met.

Although academics disagree on the exact formulation and the question whether these conditions are necessary and sufficient for someone to be held responsible, the following conditions together capture the general notion of when it is fair to *hold* an agent morally responsible for their actions or the consequences thereof:

---

that he was not a responsible person [6]. He always maintained that the exceptional winter storms were responsible [1] for the loss of the ship, but in the legal procedures brought against him he was found criminally responsible [4] for his negligent conduct, and in separate civil proceedings he was held legally responsible [4] for the loss of life and property. He is still alive and he is morally responsible [3] for the deaths of many women and children. (Hart: 1968, 11).

(1) Moral agency: This means that the agent must have adequate possession of her mental faculties and be responsive to moral reasons at the moment of engaging in the action. Young children and people whose mental faculties are permanently or temporarily disturbed will most of the times not be held fully responsible for their behavior. However, to put oneself knowingly and voluntarily into a situation of limited mental capacity (by drinking alcohol or taking drugs for example) does not in general exempt one from being responsible for the consequences of one's behavior (see Fischer and Ravizza: 1998; Frankfurt: 1971; Wallace: 1994). Some people phrase this condition in terms of intention, meaning that the action was guided by certain desires and beliefs (Corlett: 2006).

(2) Voluntariness or freedom: The action resulting in the outcome was voluntary, which means that the actor is not responsible for actions done under compulsion, external pressure or hindered by circumstances outside the agent's control. The agent must not only be in the position to determine her own course of action (cf. condition 1), but also to act according to it (cf. Aristotle: 1985; Fischer and Ravizza: 1998; Frankfurt: 1988).

(3) Knowledge of the consequences: The agent knew, or could have known, the outcome. Ignorance due to negligence, however, does not exempt from responsibility (Corlett: 2006; Swierstra and Jelsma: 2006).

(4) Causality: The action of the actor contributed causally to the outcome; that is, there has to be a causal connection between the agent's action and the damage done (Hart and Honoré: 1985).

(5) Transgression of a norm: the causally contributory action was in some way faulty; that is, the actor in some way contravened a norm and by doing so caused some negative outcome (Bovens: 1998; Feinberg: 1970).

Together these conditions determine the *scope* of responsibility, assuming the agent does not voluntarily choose to take on more. An agent is "morally [...] responsible for some act, omission or attempt to the extent that [she is] guilty of committing a harmful wrongdoing intentionally, knowingly and voluntarily, and that [she is] 'at fault' in doing so" (Corlett: 2006).

If we look at technological failure, these conditions are seldom met. Different people jointly contribute to a certain outcome, which makes it difficult to causally trace the outcomes back to actions of the different people involved. Equally problematic is the assessment of individual action in terms of the intention and freedom conditions. To what extent are individuals free if they work in collective settings? And how should the intention for one single contribution be interpreted in terms of the final outcome? In the literature this is called the problem of many hands, which is first defined as such by Thompson (1980). It refers to the difficulty to identify the person responsible for some outcome if a large number of people is involved in an activity. Thompson formulated the problem within the context of the moral responsibility of public officials. Because many different officials, at various levels and in various ways, contribute to policies and the decisions of the organization, it is difficult to ascribe moral responsibility for the organization's conduct in the last instance. For those outside of an organization who want to hold someone responsible for a certain conduct, it is particularly difficult or even impossible to find any person who can be said to have independently formed and carried out a certain policy or taken some decision. Hence, if we focus on these conditions, often no-one can fairly be held responsible. But sometimes it is the joined acting of individuals within a collective that bring about negative consequences, precisely because collectives can potentially create greater harms than individuals working independently. Collectives are able to act in ways that the individuals could not manage on their own. Some people therefore propose to hold the collective as a whole morally responsible. In that case, all individuals within the collective are held equally responsible (May and Hoffman: 1991). This ascription of responsibility to the whole collective is criticized by some for being morally unsatisfactory. People are then being held responsible for the conduct of others, which is unfair (Lewis: 1948/1991).

As the previous discussion shows, responsibility-as-blameworthiness is primarily about holding *others* responsible for certain negative outcomes (Davis: forthcoming). However, there is also a more active counterpart and that is assuming responsibility for oneself. This relates to the fifth and sixth sense of responsibility: *assuming* responsibility (be it prospectively for certain tasks or retrospectively for past actions) or what Bovens (1998) calls "active responsibility." From an insider's perspective the problem of many hands does not manifest itself, or at least not as severely, as from the outsider's perspective. The rationale for assuming responsibility is efficacy, which means that it should

contribute to the solution of the problem at hand.[104] Davis (forthcoming) discusses this insider's perspective to engineering and he argues that "the practical criterion most appropriate to engineering as such is neither explanation [...] nor blame [...], but [...], whether or not of their own making, [...] to fix it if they can." In a more forward-looking sense it seems intuitively plausible to relate this efficacy rationale somehow to the potentially harmful consequences of technology and the prevention thereof. Hence, we take this rationale of efficacy to imply that in order for a technology to be socially and morally desirable or at least acceptable, all potential implications for society (e.g., human health and the environment) should be taken into account during the design phase. That is, for every potential implication, whether this is a risk or some other problematic issue, someone should be ascribed the responsibility to address this issue. This does not mean that all risks should be completely avoided – a requirement which it would be impossible to live up to – but that everything which can reasonably be known should be considered during design and development phase.[105]

Before we illustrate how this active responsibility works in engineering practice, let us first turn to the broader field of technology management and engineering ethics.

6.3.   A history of technology management and engineering ethics

In their eminent article on engineering practice and engineering ethics, William Lynch and Ronald Kline argue for a more sociologically informed engineering ethics. Insights from Science and Technology Studies (STS) could contribute to the "initial recognition of ethical problems in ill-structured, real-world situations" (Lynch and Kline: 2000, 209) and as such provide an understanding of the role of "workplace organization and culture in facilitating or impeding remedial action" (195) rather than focus on whistle blowing and questions related to blame. A similar plea for sociological investigations can be found in the work of Willem Vanderburg (2000), who argues that these kinds of investigations may contribute to a more preventive orientation in engineering.

---

[104] Note that responsibility-as-blameworthiness could be guided by consequentialist considerations as well. The consequentialist argument for blaming is based on the idea of deterrence of wrongful conduct.

[105] For the moment we leave it open how to determine "what can reasonably be known." For the remainder of the argument it is not required to exactly define it.

Lynch and Klyne are optimistic about this, what they call, "sociologically sensitive ethics," which could ultimately help "identifying and responding to ethical issues in engineering" (Lynch and Kline: 2000, 208). Swierstra and Jelsma welcome the authors' criticism of the current practice of blaming engineers for disasters. However, their conclusion is less optimistic. They argue that "the STS literature empirically demonstrates that in common cases of modern technological projects, the necessary conditions for individual moral agency and responsibility are lacking" (Swierstra and Jelsma: 2006, 311). Swierstra and Jelsma think that engineering ethics is too much focused on conditions of individual moral responsibility. They therefore propose to rethink the "relationship between individual moral agency and responsibility on one hand and on the individual's enabling and constraining environment on the other" in order to come to an "institutional ethics rather than an individual one" (312).

Both articles provide valuable insights for the topic of responsibility. Lynch and Kline's argument that engineering ethics should be sociologically informed is in line with the recent "empirical turn" in applied ethics.[106] Swierstra and Jelsma have a strong point in saying that the focus on *individual* moral responsibility seems unsuitable for engineering practice. However, in both papers the position of the ethicists remains that of an outsider. They do not get involved in the engineering practice but assess the outcomes of the engineer's work from an external perspective. This external or outsider's perspective has a long history in technology management, as we will show in the present section.

Historically, the focus of technology management in the 19th and early 20th century was on the establishment of agencies to regulate technology, which was typically done by outsiders as a response to existing undesirable consequences. As such the room for influence was limited. Especially in the aftermath of World War II, science dynamics was almost autonomous with science policy uncritically accommodating ongoing science (Rip: 2006, 84-85). At the end of the 1960s, it was realized that technological development, though intended to positively contribute to society, often had unintended negative consequences as a side-effect. Because of the lack of power of the downstream approaches, more anticipatory attempts to govern R&D emerged, starting with so-called *Technology Assessment* (TA) approaches, which were originally aimed at detecting these

---

[106] Cf. Section 1.3.2 of the Introduction.

negative societal implications beforehand (Berloznik and Van Langenhove: 1998). The anticipatory nature of TA allowed for agenda-setting and the inclusion of "downstream" considerations into "upstream" decision making contexts (Fisher et al.: 2006, 487). However, an important drawback of the early TA approaches was the uncertainty associated with the initial forecasting. Thus, actors involved in the management of technology were faced with what has become known as the "control dilemma" or "Collingridge-dilemma":

> Attempting to control a technology is difficult, and not rarely impossible, because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow. (Collingridge: 1980, 19)

To overcome this dilemma, more constructive and participatory TA approaches have emerged, notably Constructive TA (CTA; Schot and Rip: 1997), and Interactive TA (iTA; Grin and Hoppe: 1995; Reuzel et al.: 1999), here onwards referred to as second generation TA. These approaches seek to address a broader range of issues in the assessment but also to include more participants. Contrary to the traditional TA approaches, in which technology development was conceived as an autonomous force, these newer approaches try to influence technological design decisions as well (Fisher et al.: 2006). They do so by involving all possible actors or stakeholders that participate in the development, implementation, and consuming of technologies. However, these second generation TA approaches still remain "a practice organized outside the laboratory, ... [since they do not] involve the scientists while doing their job but after their job is done" (Berloznik and Van Langenhove: 1998, 24-25). Accordingly, despite these attempts to feed public engagement back into design choices, the influence of second generation TA on everyday practices of science remains fuzzy and unclear (Wilsdon: 2005, 24).

The latest trend in TA therefore aims at the inclusion of societal considerations during the phase of R&D itself. These third generation approaches are labeled Real-Time TA (RTTA; (Guston and Sarewitz: 2002; Sarewitz: 2005) or Integrated TA (ITA; (Berloznik and Van Langenhove: 1998). Contrary to previous TA approaches, the focus of third generation TA approaches is on "opening up the innovation process, rather than managing it after-the-fact" (Sarewitz: 2005, 20). Similar to second generation TA approaches, RTTA and ITA seek to build learning into the implementation process, but by

staying close to the technological development process itself these newest approaches have more impact on technological development. Instead of merely addressing the impacts of technology, these approaches aim at "shaping the trajectory of technological development" (Wilsdon: 2005, 24) in order to improve both the societal consequences and the decision making about science and technology (Sarewitz: 2005). Hence, TA should be integrated in the practices of R&D and become, as Berloznik and Van Langenhove (1998) call it, "a built-in monitoring of the R&D process" (27). The ideal of TA can then be conceived as a research practice, internal to the process of R&D itself, which allows for more reflexive participation by the scientists and engineers. The results of these reflections and studies feed back into the ongoing research. Ultimately, this may lead not only to more societal responsible research but possibly to more efficient and effective research as well (Berloznik and Van Langenhove: 1998, 30).[107]

These internal governance attempts are labeled midstream modulation (Fisher et al.: 2006). In the past years, several attempts have been made to implement and further develop these midstream or third generation TA approaches. In the United States, the National Science Foundation has funded approximately 6 million dollars to prototype this research at the Center for Nanotechnology in Society at Arizona State University (Sarewitz: 2005). Preliminary results show "a receptiveness to collaboration on TA activities, rooted in a desire to contribute to societally beneficial outcomes" (Guston and Sarewitz: 2002) and an awareness of the possibility of modulating the decisions accordingly. Moreover, the interactions were found to add value to research instead of hampering research (Fisher: 2007; Fisher and Mahajan: 2006).

In the Netherlands, the most important public financer of technology research STW and the Netherlands Organization for Scientific Research (NWO) started a pilot of four ethical research projects that are carried out parallel to the technical research (Driessen: 2009; Van der Burg: 2009; Zwart et al.: 2006). The idea behind this type of research is that ethical investigations are carried out parallel to, and in close cooperation with, a specific technological R&D project. The ethicists interact with the technological researchers, allowing the ethicists to co-shape new technological developments. Whereas the pre-World War II

---

[107] Related to these third generation TA approaches, there is a call for a more preventive orientation in engineering and the regulation thereof (Vanderburg: 2000). This requires not only an engineering curriculum reform (Vanderburg: 2009) but also a broadening of the engineering profession itself (Vanderburg: 2006).

management was aimed at increasing the pace of technology development and the early TA attempts at inhibiting the pace, the third generation TA attempts are aimed at informing the direction of technology development (Van de Poel: 2008, 29-30).

In the next section we will bring together the discussion of responsibility ascriptions and technology management by discussing one of the ethical parallel researches carried out in the Netherlands.

6.4.   Development of a new sewage treatment technology

Let us summarize our arguments so far. We first argued that there are roughly two perspectives for responsibility: an outsider's perspective, which often focuses on responsibility-as-blameworthiness, and an insider's perspective, which is about assuming responsibility. We then gave a history of technology management. It was found that with the most recent trend in technology management, the focus had shifted from management before or after to management *during* technology development, or from setting the *pace* of technology development towards informing the *direction* of technology development.[108]

One of the ethical parallel studies carried out in the Netherlands several years ago concerned the development of a new sewage treatment technology; the so-called granular sludge sequencing batch reactor (GSBR).[109] In this technological project different people contributed, varying from engineering researchers working for a technical university to more practically oriented partners working at a Dutch Waterboard and a commercial consultancy firm. The ethical parallel research consisted of a qualitative study, based on interviews, document analysis, attendance of technical meetings, and the organization of an interactive session

---

[108] It is interesting to note that the engineering profession itself has also broadened its conception of the engineer's "professional responsibility" to all stages of technology development and use (e.g., the use of the life cycle analyses and a focus on sustainable development; Mehalik: 2000). The codes of ethics or codes of conduct formulated by engineering professions also point to a broader conception of the engineer's responsibility. The Dutch Royal Society for Engineers, for example, explicitly states in its code of conduct that engineers "when making technical decisions take into account the safety and health of the general public and the environment" (KIVI NIRIA: 2006).

[109] For a more detailed analysis of the responsibility ascriptions within this project, see Doorn (2011). See Text box 2.1 for a more elaborate description of the GSBR technology.

in the Group Decision Room (GDR; an electronic brainstorming facility) with the different stakeholders, where questions related to risks and responsibilities were addressed.

During the ethical parallel research, it was observed that some of the risks due to so-called secondary emissions (i.e., unwanted but not yet regulated substances in the effluent), were not addressed by any of the engineers and researchers involved in the development of the technology. The users of the technology delegated the risk of secondary emissions to the research phase, for which they were not primarily responsible, and most of the researchers allocated the risk to a phase for which they in turn bore no responsibility. Nobody therefore assumed responsibility for dealing with this risk. The arguments put forward by the researchers and users were that the impact of the risks due to these secondary emissions was negligible and that problems were expected to be solvable in the next phase of the research. This was based on the presumed – but never proven – similarity between biological processes in traditional sewage plants and the biological processes in the GSBR technology. As a result, the issue who is responsible for checking or preventing secondary emissions never became an object of discussion. The engineering ethicists brought this issue to the table, whereupon the consultancy firm together with the university applied for subsidies to carry out research into the secondary emissions.

This example shows that engineering ethicists can play an important role in distributing responsibilities in R&D projects.[110] Rather than giving orders on what to do and what to avoid, the ethicist's role can be to elucidate issues that would otherwise be overlooked (in this case, issues stemming from a knowledge gap). Ideally, their contribution could consist in trying to enable engineers to assume their professional responsibility.[111]

How does this in practice work? In the example of the sewage treatment plant the gap in the responsibility distribution was based on the lack of knowledge as a result of which the engineers could not adequately assess to what

---

[110] That this new role is also a desirable one is at least suggested by the results of the interviews that I carried out with some of the members of the ALwEN project. They argued that an ethicist has knowledge that they themselves are lacking and that it is desirable that ethicists play an active role in elucidating those issues (see also Text box 7.1).

[111] Since this new role is more that of an insider than an outsider, it is important that the ethicist has familiarity with engineering practice. I therefore prefer to speak of an engineering ethicist rather than a professional ethicist or applied ethicist because the latter do no necessarily know the context, making it difficult to fulfill this new role.

extent the secondary emission were in fact dangerous. Or, in other words, they were not able to discharge the responsibility to address the risk due to the secondary emissions. On the basis of the principle of moral consistency we could derive a responsibility for the engineers to check whether they are able to perform their tasks (Marcus: 1980; by Van den Hoven (1998) referred to as a meta-task responsibility).[112] Although the principle of moral consistency is usually formulated in terms of avoiding *conflicting* responsibilities (i.e., moral overload), the case for meta-task responsibilities is even stronger than the one for avoiding moral overload. After all, in case of conflicting moral demands, the primary responsibility is countered by another moral demand, which is lacking in case of meta-task responsibilities. This meta-task responsibility applies to "ordinary" moral responsibility and there is no reason to assume that it does not apply to professional responsibility (in this case, the professional responsibility of engineers).

Hence, we could say that if engineers have a particular responsibility, they also have the meta-task responsibility to check whether they are able to discharge their responsibility or to perform that particular task (in this case, to check whether they have the required knowledge). Hence, the ethicist's role could be to elucidate the missing conditions and help engineers see relevant factors that might otherwise have been overlooked.

To recapitulate, we have argued that engineering ethicists and technology managers have a new role to play in R&D projects, namely to make sure that efforts are made to establish responsibility distributions that are efficacious and that all relevant issues are addressed. In order to do that, it would be useful to come up with a list of guidelines to consider when assigning and distributing responsibility. Accordingly, we have a tentative suggestion of what such a list could look like. This list is intended to be a first proposal, not as the final and only possible list. Our main aim in creating such a list is to show that there are important considerations to make when assigning responsibility and that this should not be ignored. Engineering ethicists who take part in the R&D phase are well equipped to help catch these considerations. What the specific guidelines should be is something that should be discussed and our list could be a way of starting a discussion. We believe that the guidelines given in Table 6.1 are

---

[112] Moral consistency is a topic which, for reasons of space, I cannot discuss in any detail. In short, the discussion of moral consistency deals with the question whether or not there is one action-guiding principle that can serve to settle disputes in case of a moral dilemma.

reasonable, but it should be kept in mind that they obviously have to be applied and interpreted as well as adjusted to specific contexts.

Table 6.1: Guidelines for distributing and assigning responsibilities

Guideline
1. Making sure the agent has capacity, resources, and power to perform tasks entailed by responsibility
2. Avoiding suboptimal distributions of responsibility
3. Minimizing negative effects on internal commitments
4. Minimizing negative effects on external commitments

The last two guidelines are – again – related to the principle of moral consistency. Although we probably cannot rule out conflicts all together, it is worthwhile to take this issue of consistency into consideration. This would, for example, mean that in a situation where Agent A is substantially better equipped than Agent B to do R, it might still be preferable to assign (the responsibility to do) R to Agent B. This will be the case when assigning R to A would make A unable to perform her other tasks. Assigning R to A might not only be unfair but also ineffective because some of A's other responsibilities may be equally important. For example, A may have a previous responsibility S to reduce risks to the public. Adding R should not prevent A from doing S.

Let us look at the example of the sewage treatment technology once more to illustrate these guidelines.

(1) Responsibility should be assigned only to agents who have the power, resources and capacity to perform the tasks entailed by that responsibility: The issue could not be resolved within the project itself so funds were requested such that additional research could be carried out. The importance of further investigating this issue was recognized by the researchers.

(2) When there are several possible candidates, responsibility should be assigned to the agent who is best equipped to perform the tasks entailed by that responsibility; that is, suboptimal distributions of responsibility should be avoided: The issue about the secondary emissions was primarily about biological mechanism so the researchers at the laboratory for biotechnology are the most obvious

candidates in terms of capacity to do this (and not the researchers from, say, civil engineering).

(3) A balance should be struck in order to minimize the negative effects on the agent's (who is assigned responsibility) other commitments in the project: It was indeed necessary to acquire *additional* funding and not just to refocus the research because this might come at the cost of other – maybe more important – research topics such as the focus on primary emissions.

(4) A balance should be struck in order to minimize the negative effects on the agent's (who is assigned responsibility) commitments outside the project: Project members also have commitments towards the companies they represent and to professional peers. Especially the relations with the latter are based on a combination of trust and criticism, and cooperation and competition (Van de Poel: 2008). Responsibility should not be assigned in a way destructive to these relations.[113]

In addition to making sure responsibility is assigned in an effective way, engineering ethicists have, as we saw in the GSBR case, an important task in finding or acknowledging potential risks. This entails facilitating an adequate risk management. In this way, engineering ethicists could be seen as "the public's advocate" in the sense of protecting the public against threats and to reduce the risks of harm to people and the environment through encouraging discussion and management of those risks.

By making the technological research team aware of these responsibility issues, some of the technological researchers took the initiative to incorporate the secondary emissions in the research project. As such the effect of the ethicists' involvement on the engineering practice was not blaming or sanctioning but rather that of co-shaping. The ethical parallel research did not as much pose limits to the technology development but guided it.

Following this approach we come to a new role for technology managers and engineering ethicists. Are they traditionally driven by the question *whether or not*

---

[113] Note that this holds for relations outside the professional sphere as well. People have a reasonable right to a private life (e.g., family, hobbies, etc.). Responsibility should not be assigned in a way destructive to this private life. This does not amount to time only, but also to moral issues concerning the technology under development (e.g., stem cell research).

to authorize, approve, or adopt a certain technology or by the question of who is to blame for potential mistakes, in the new approach (RTTA, midstream modulation, ethical parallel research) the guiding question is *how* research is to be carried out (Fisher and Mahajan: 2006, 3). This requires a different role for ethicists. They become insiders or (integrated) participants in the research, not driven by a "negative heuristic," which expresses what not to do (Vanderburg: 2009). Their role can be to elucidate issues that are otherwise overlooked.

However, one important point of concern that needs to be addressed is an issue raised already in the eighteenth century by Immanuel Kant (1795). In the second supplement of his essay on Perpetual Peace, Kant argues that philosophers should not become kings because of the danger to become corrupted by power. It is not the intentional misuse of power that Kant fears, but the loss of "untrammeled judgment of reason." This suggests that ethicists should on the one hand become an insider, but that they should maintain their critical stance. They should maintain their independence, also in order to assure that there is room left to give and accept criticism (Van de Poel: 2008, 36). A combination of personal skills and institutional safeguards is probably required to deal with this challenge. Personal skills are required to maintain the relations of trust and criticism, to temper the eagerness to report moral concerns and discuss those with the technical researchers first. Clear arrangements about reporting the ethical parallel research, including the right to "react to claims made by the ethicists or complaints about unwarranted claims" (ibid.), should be part of the institutional safeguards to make the cooperation a successful one.

6.5.   Conclusions

In this chapter, we discussed two responsibility perspectives in the light of technology development: an outsider's ("holding responsible") and an insider's perspective ("assuming responsibility"). It was found that the outsider's perspective was driven by a different rationale than the insider's perspective. By linking the more academic discussion on responsibility to the recent trend within engineering ethics to manage technology during the R&D phase of technology development, we developed a new role for engineering ethicists, namely, to become that of insider in the research team, who can help elucidate issues that are otherwise overlooked.

By means of the example of recent ethical parallel research, this new role for engineering ethicists was further illustrated. It was shown that the involvement

of ethicists led to the identification of gaps in the distribution of responsibilities (related to a lack of knowledge). As a result, funds were acquired to carry out additional research. As such, the analysis of the responsibilities by the ethicists led to an improvement of the division of labor amongst the technological researchers and engineers, which in turn led to an improved technological design.

A tentative list of relevant considerations was developed. This list could serve as a tool for assigning and distributing responsibilities. It was shown that capacity, power, and resources are necessary conditions for discharging one's responsibility. Unless an individual or a group of individuals have the resources and power to do what is required, it is not reasonable to expect them to do it. Additionally, guidelines were given to take efficacy in consideration as well.

In their new role, engineering ethicists can facilitate a discussion of potential risks and other ethically relevant issues at a time the changes in the technological design can still be made. Subsequently, they can guide the discussion on how to assign and distribute the responsibility for addressing these risks.

7 Conclusions

The main objective of the research presented in this thesis has been to alleviate the problems of many hands in R&D networks. This problem was shown to be especially relevant in the context of R&D networks. On the basis of an ethical parallel research, different rationales for distributing responsibilities were explored and analyzed in terms of Rawls' WRE model. In this concluding chapter, I answer the research questions as formulated in the introductory chapter and present the main conclusions of my research, including a critical reflection on the generalizability of the findings and recommendations for further research. This chapter concludes with my personal vision on applied ethics in general and engineering ethics more in particular. An evaluation of the ALwEN project and the ethical parallel research is included in Text box 7.1.

7.1.   Research questions

In the introductory chapter, I presented the main research question as follows: How can a procedural approach, based on Rawls' idea of WRE, be used to resolve the tension between conflicting responsibility notions, which lies at the core of the problem of many hands? In order to answer this question, let us go back to the successive subquestions.

The research question addressed in Chapter 2 was: What are the different responsibility conceptions and what are the accompanying criteria or requirements for the responsibility ascriptions (Q1)? Three different theoretical conceptions were identified: a merit-based, a rights-based, and a consequentialist conception, with the accompanying criteria fairness, informed consent and efficacy respectively. It was shown that these different conceptions correspond to different moral background theories ("comprehensive doctrines"). This suggests that they are – at first glance – all legitimate. After all, in the philosophical literature, the different moral background theories have all been defended and till date, none of the moral theories has been definitely rejected by its respective opponents. In other words, given the current state of the moral philosophical debate, the burden of proof lies with the person rejecting one of the theories as illegitimate rather than with the person defending it. The fact that there is – in theory – a pluralism of responsibility conceptions accounted for the use of

procedural theory to reconcile the tension between the different responsibility conceptions.

Looking at the distributing rationales as displayed by the engineers and technological researchers in the ALwEN project (research question Q3), we see that the theoretical conceptions identified in Chapter 2 were all partly "confirmed" by the empirical data (Chapter 4). The engineers and technical researchers did indeed tend to think in terms of "fairness" and "efficacy," the criteria that correspond to the merit-based and consequentialist conception. The type of argumentation labeled as "user-perspective" or "societal perspective" seems to correspond most to the rights-based conception of responsibility. The correlation between the type of argumentation (fairness, efficacy, user perspective, and societal perspective) and the individual ethical positions indicate that the different responsibility conceptions are indeed legitimate, because these ethical positions represent different – but legitimate – moral background theories. These different conceptions lead to different opinions on who is responsible for doing a particular task, which may lead to gaps in the responsibility distribution because all people expect someone else to take up a particular task. This suggests that the problem of many hands can indeed partly be traced back to this pluralism of responsibility conceptions.

Concerning the application of Rawls' procedural theory to practical problems (research question Q2), I tried to learn from the obstacles as identified in Chapter 3. This resulted in an inventory of recent applications of Rawls' theory and this taught me that the stakeholders deserve a more prominent role in assessing whether or not an outcome is justified terms of their own WRE. Assessing whether or not someone has attained a WRE is indeed something one can only judge for oneself. In this regard, I departed from the approaches that use reflective equilibrium methodology mainly to identify and analyze the different moral intuitions pertaining to a particular case (cf. Van Thiel: 2009, Chapter 6). Rational deliberation cannot be reduced to a set of relations among propositions but is instead an act of embodied reflection. Individual emotions and circumstantial judgments play a role in perceiving the particularities of a situation. Together, emotions and perception may prompt reflection and elicit a response. This indicates that the process of searching for reflective equilibrium is essentially a process embodied in a sovereign deliberator (Richardson: 1997, 188-189). Hence, if the reflective equilibrium approach is to fulfill its justificatory purpose, the establishment of this equilibrium cannot be judged from a third-person perspective. For this reason, I asked the people involved in

the ALwEN project themselves to assess whether the outcome of the workshop where we distributed the responsibilities fitted their own moral frameworks and principles (in short, whether they thought the outcome was "fair").

In Chapter 5, I addressed the research question whether a procedural approach to fairness resolves the tension between competing conceptions of responsibility (Q4). For answering this question, I draw heavily on the empirical material collected through the workshop I organized for the members of the ALwEN project. In Section 1.3.2, I argued why I followed a so-called "good reasoning-justified outcome strategy," as advocated by Van Thiel and Van Delden (2009) and DePaul (1993). This strategy seeks reliability from the deliberative process itself in order to warrant a justified outcome. In the research presented in this thesis, I introduced the procedural norms inclusiveness and reflective learning to assess the reliability of the process and to provide justificatory force.

Inclusiveness requires incorporating the "broadest evidence" available (Daniels: 1996, 2-3). For an equilibrium or consensus to be just, it is important that all relevant actors are included and that they can equally engage in the debate. If the criterion of inclusiveness is released, the method loses its justificatory force. In practice, it will be difficult to directly involve all relevant people in the decision making. However, in a case such as the ALwEN project, the interests of those people that are affected by the technology should at least be represented. If we look at the end users, for example, it is important that their interests are looked after. Even though they do not have to be involved in actually distributing the responsibilities, the ultimate responsibility distribution should include the task to look after their interests. So, though indirectly, they should be included or represented in the decision making process.

Regarding the procedural norm of reflective or higher-order learning, the empirical data suggest that, strictly speaking, reflective learning (here, a willingness to change one's moral background theory) is not required to come to a similar distribution of responsibilities. However, I argued that without reflective learning, people would probably not recognize the legitimacy of other people's arguments and that reflective learning is therefore still required to agree on the possibility and legitimacy of disagreement. People do not have to change their own conception of what responsibility amounts to, but they do have to acknowledge that their conception is one among many.

As explained in Chapter 3, Rawlsian approaches can also serve a constructive purpose and it is especially with the "good reasoning-justified outcome" strategy that the two purposes (justificatory and constructive) should be seen as intertwined rather than distinct. After all, part of the justificatory power is derived from the quality of the deliberative process. As such, encouraging reflection (conform certain deliberation rules) is supportive for getting a justified outcome. For encouraging reflection, the method seems very useful. It prompts discussion and it encourages people to think what they find fair and defend these claims in terms of public reason. The strength of the method lies in incorporating all layers of morality and revising them when necessary and not limiting the revision to, say, considered judgments. For justificatory purposes, the WRE model is a matter of varying degrees. It is possibly too demanding to require that all layers are in full equilibrium. However, not having attained full equilibrium does not mean that the participants assess the outcome as fully unfair.

In the ALwEN project, the team was inclusive and they showed reflective learning processes as well. This warrants the conclusion that the "good reasoning-justified outcome" strategy was successful, both in terms of efficacy (the final distribution of responsibilities covered most moral issues) and fairness (the outcome was accepted by all as fair). In that sense the procedural approach is a promising for resolving the tension between competing responsibility issues.

The last research question stands somewhat apart from the main line of argumentation and it concerns the role of an engineering ethicist in R&D: What role could engineering ethicists play in R&D? Concerning the procedural approach, the fact 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 role is a modest one and this role could be performed by other ethicists as well. It is not up to the ethicist to say whether or not an outcome (here: responsibility distribution) is fair, precisely because the ethicist does not have direct access to these first-person considerations. However, an engineering ethicist can have a profound role in putting salient issues on the agenda. It *is* the ethicist's expertise to say that certain issues need to be addressed before a technology can be safely developed any further. However, given that there is a whole R&D team working on it, the ethicist cannot decide whose particular responsibility it is. The fairness of the

distribution of these responsibilities among the members of the research team is something that the engineers and the technical researchers and developers should agree upon themselves. The WRE approach can be supportive in this process. By discussing the fairness of the responsibility ascriptions, the participants in the research team are encouraged to think about the legitimacy of other people's claims and the scope of their work. If the team members find out that none of them can be fairly ascribed a particular responsibility, they should consider extending their team with professionals who are able to do assume the responsibility. In the ALwEN project, the clinical partner was initially not part of the project team. However, once the team realized that their project involved clinical questions that require due attention, they invited the clinical partner to join the research team.

My discussion of the procedural approach indicates that an ethicist's role can be normative without being "moralizing," to phrase Swierstra and Jelsma's interpretation of moralism as "the attitude of burdening individuals with moral demands and responsibilities they cannot possibly meet" (Swierstra and Jelsma: 2006, 314). The ethicist's role can be normative in the sense of putting issues on the table, pointing at gaps in the distribution of responsibilities or gaps in the composition of the research team itself, and even deciding on ethical issues. This latter normative stance was even asked for by two of the participants in the ALwEN project whom I interviewed after finishing my research. When asked how they would ideally view the involvement of an ethicist, one of them said:

> Considering the role of an ethicist, I think he can be normative. After all, ethics is his expertise. So if I, as a technical specialist, see a schematic diagram that could be improved I say so and discuss it with my fellow-workers and of course try to convince them to change it. Analogously, the ethicist is the specialist on ethical questions and should have the responsibility to make decisions regarding these ethical points. It is maybe not very common to explicitly include an ethicist but it might be a good idea. In 99% of the cases you can maybe trust the commons sense of the technical people but in some projects it may definitely be worth to include an ethicist for dotting the i's.

Taking up this normative stance requires that the ethicist is well-informed about the project, including some of its technical details. Of course, an ethicist does not need to be as specialized as the technical team member themselves, but some familiarity with the topic and the application domain is required. Without this familiarity, the ethicist's involvement runs the risk of becoming either irrelevant (for example when the ethicist does not sufficiently understand the

application of the technology) or moralizing (when the ethicist does not understand what could reasonably be asked from the technical researchers and what not). However, in Section 6.4, I also discussed the concern of a too tight link between philosophers (here: ethicists) and the technical researchers. Although it is in general favorable that ethicists become insiders in the project team, they should keep their critical stance. Institutional safeguards and personal skills are required to adequately deal with this balancing act of involvement and critical independence. For the ethicist, awareness of the limits of the scope of one's judgments is probably a first step ahead. Ethicists should be careful not to adopt a moralizing attitude. I argue that the procedural approach to assessing the fairness of responsibility distributions developed in this thesis fits this new and delicate role of ethicists perfectly well. On the one hand, it leaves room for putting moral issues on the agenda that are otherwise overlooked. In that sense, the ethicist can adopt critical and normative position. On the other hand, the ethicist does not have the last word in ascribing the responsibility for addressing these issues. She cannot decide on the question "Who will do what?" Instead, she can contribute to making things done, either by raising awareness for important moral issues or by explicating that the research team needs the expertise of others and as such, make the research team more inclusive.

Having discussed the successive subquestions, we can now turn to the overall research question, which I formulated as follows: How can a procedural approach to fairness, based on Rawls' idea of WRE, be used to resolve the tension between conflicting responsibility conceptions, which lies at the core of the problem of many hands? In the ALwEN project described in this thesis, the procedural approach was useful for prompting discussion on the legitimacy of the different conceptions and the question what a fair distribution of responsibilities amounts to. Although it was not possible to achieve a full overlapping consensus regarding the distribution of responsibilities, the case study showed that the tension between the different conceptions can be alleviated by structuring the discussion along the lines of the different layers of the Rawlsian WRE approach, because this encourages participants to think in terms of a "fair" workload and the legitimacy of other people's arguments. Although some differences in opinion remained, the lasting effect of the workshop was that most issues that were at first not addressed were later taken up by one of the partners in the research project. In that sense, the procedural approach did indeed alleviate the problem of many hands.

I deliberately phrased my research question as a "how" question; I think that the procedural approach is not simply a "tool" or "recipe" that one can uncritically apply without paying attention to the circumstantial particularities. It is for this reason that I also addressed the role of the engineering ethicists. When doing embedded or parallel ethical research, an ethicist should balance between uncritical support and unrealistic moralism. When applied with due care (that is, focusing on the composition of the research team, missing issues on the agenda, and the quality of the deliberative process, without falling in the trap of moralism), the procedural approach developed in this thesis can help resolving responsibility issues.

7.2. Generalizability

When basing one's empirical conclusions primarily on one case study, the question of generalizability is of course an important one: To what extent can the empirical results be generalized or transferred to other projects or situations? In addition, there is also the question of normative generalizability: To what extent can this approach be used to resolve other moral issues?

Concerning empirical generalizability let me start by noting that the ALwEN project is not an exotic R&D project. To the contrary, following the description of in Chapter 1 (footnote 3), we can see that the ALwEN project is almost prototypical for R&D: it has a focus on *applied rather than fundamental research* (witnessing the involvement of three technical universities and two industrial research institutes and only one general university), it involves *several actors* (with the consortium of SMEs counted as one, there were 12 parties involved), the project has an *innovative* character by focusing on system rather sensor behavior, it is aimed at developing *new Ambient Intelligence based applications* with the ultimate aim of *commercial exploitation*, and a large part of the project is dedicated to *experimentation*. The ALwEN team is slightly exceptional in the cooperation between the SMEs. The consortium of SMEs deliberately seeks cooperation in a "pre-competitive" stage; this means that for joint research projects only those application domains are selected that none of the individual partners has a particular stake in.

Of course, the fact that the ALwEN project is a prototypical R&D project does not guarantee that people in similar R&D projects will have behave similarly (that is, that the network will fulfill the criteria of inclusiveness and higher order learning and that the team includes all relevant moral issues in their scope of

work). I expect this to a large extent to be dependent on the type of project and the type of cooperation. When the project concerns a relatively controversial research topic (for example, the development of military equipment), the project team may be more inclined to exclude particular organizations, especially those that are critical towards the application domain. The same holds when the cooperation is vulnerable, for example because the internal competition is high. In the ALwEN project, the cooperation between the SMEs was made more stable by the "pre-competitive" cooperation. When the competition between the partner institutions represented in the team is high, the cooperation becomes vulnerable because people may encounter a tension due to conflicting loyalties. In those situations, the project team will probably be less inclined to invite an ethicist in their midst. This also holds for the type of partners involved. Larger industrial companies are usually less open towards outsiders (which is to some extent understandable given the scope of their work and stakes of these types of companies).[114] However, especially in the more controversial projects, the need for ethical reflection is probably at least as urgent. In these cases, more formal supervision may be required, for example by making assessment procedures obligatory.[115] Although necessary in some cases, I think that those more formal management procedures are unavoidably reactive rather than pro-active in addressing ethical issues and therefore less effective. I would therefore prefer to follow a cooperative stance if possible. Dependent on the composition of the research team and the research topic, the approach followed in this thesis is a feasible one.

The question to what extent the procedural approach could be used to resolve other ethical issues is largely dependent on the type of questions one wants to resolve. As explained in the Chapters 1 and 3, the procedural approach is a

---

[114] In the ALwEN project, this was confirmed by the strict formal rules that were followed in reporting the findings of my ethical investigations. Whereas the universities and SMEs agreed that only the ALwEN project leader checked the results of my study before publication, the multinational company Philips Research required a more thorough checking procedure. It should be noted, though, that, in effect, Philips did not ask for any substantive changes in any of the publications on this research.

[115] In the chemical industry, for example, the European chemical legislation REACH (Registration, Evaluation, Authorization and restrictions of Chemicals) was implemented to provide an answer to the controversy pertaining to the control of chemicals and the fact that the chemical industry itself was unable to adequately address the potential negative side effects of new chemical substances (Clausen and Hansson: 2007).

method aimed at resolving issues in a situation of pluralism. Moral problems that are *not* related to pluralism are probably better addressed differently. In Chapter 3, an application of Rawls' theory aimed at the establishment of fair seeding rules in tennis was mentioned as an example where the method did not prove useful. In this context, the underlying values did not really pertain to pluralism so the method was found to be too far removed from the persons' particular practices. This detachment from one's particular practice and the desire to justify one's actions on "grounds that others could not reasonably reject" is a requirement that makes particularly sense when people have a different point of departure. The tennis context could be considered a "private sphere" where political theory is not the most suitable way of approaching problems. But also in the context of collective or corporate responsibility, there may be problems or dilemmas that cannot be solved adequately with procedural theory. Philip Pettit, for example, introduced the notion of "discursive dilemma" to refer to situations where the final decision is based on a conceptual sequencing of the issues to be decided. Dependent on the question whether consensus is to be sought on the final decision or on the individual issues (premises), the outcome may be different (List and Pettit: 2004; Pettit: 2001).[116] As a result, the outcome of the procedure may lead to undesirable results for which none of the individual actors can be held responsible (Pettit: 2007). In this case, it is an ambiguous or inconsistent decision making procedure that may lead to the problem of many hands and not a pluralism of responsibility conceptions. This means that there are instances of the problem of many hands that cannot be dealt with by the procedural approach developed in this chapter.

Lastly, and here I come back to the role of the ethicist again, I think that in the type of project discussed in this thesis, familiarity with the context is so important that it requires specialized knowledge to effectively participate in these technological projects. It is important for two reasons. In the discussion of the last research question (What role could engineering ethicists play in R&D?), I already addressed the need for familiarity with the research topic to be able to pick out the relevant issues. After all, an ethicist can only contribute insofar as

---

[116] In one of his papers, Pettit uses these discursive dilemmas explicitly to criticize deliberative democracy. The basis of the problem is "the fact that socially aggregating the conclusion-judgments gives us a different result from socially aggregating the premise-judgments." All consensus theories, including Rawls' reflective equilibrium methodology, are in varying degrees vulnerable to this criticism (Pettit: 2001, 273).

she is able to become familiar with the problem. In my ethical investigations, it helped that I have a technical background and that I was therefore able to grasp the most important characteristics of the technology. Additionally, being an engineer enabled me to create a relation with the ALwEN team members. I was able to talk the same "technical language," which may have contributed to the engineers' willingness to participate in the interviews and the workshop. As explained in the previous section, working as an embedded or parallel ethicist requires special skills to create an atmosphere of trust. Being able to talk the same "language" is, if not a necessary, at a least desirable condition. Hence, regarding ethical parallel research in general and the procedural approach in particular, successfully working as an applied ethicist requires knowledge of the field to be studied. I think that this approach could very well be extended to other projects and other contexts and that it could be applied by other ethicists as well, given that the particular ethicist is familiar with the field, both content-wise and in terms of speaking the "language."

7.3. Further research

As every research, this thesis is aimed at giving answers but, at the same time, raises questions for further research. Since this research was explorative in nature, more research is needed to further develop and assess the procedural approach. This not only requires doing more case studies but also using more rigorous methodology to compare the different cases. In this last section, I would like to address two conceptual issues that specifically deserve closer attention.

The first point concerns the outcome of the reflective equilibrium approach. When applying the reflective equilibrium methodology to real life examples, we want to know when the application is successful; that is, whether or not people have attained a reflective equilibrium. In the existing literature on coherentist justification, it is by now realized that the concept of coherence cannot straightforwardly be applied without further instructions as to how people should evaluate their own set of beliefs with respect to coherence (DeGrazia: 2003; Rauprich: 2008; Van Thiel and Van Delden: 2010). Although most people agree that coherence is in some way related to the way a body of beliefs hangs together (Bonjour: 1985, 93-95), the nature of coherence is poorly described. Van Thiel and Van Delden provide an excellent overview of the existing literature on coherentist justification, including a preliminary tool for measuring coherence

(though the term "measuring" seems somewhat exaggerated). What is missing, however, is an analysis of what counts as agreement. During my own empirical research, I found that most "decisions" cannot be framed in terms of simple closed questions that allow for binary answers (agree or disagree; proceed or abort). Agreement is a matter of degree: To what extent recognize people the way a problem is framed? How should it ideally be addressed? What types of solutions are relevant? Contrary to the hypothetical examples in the ethical literature, where the points of discussion are often presented as simple yes-or-no decisions, decision making in real life is often more complex. Rather than talking about isolated decisions, it is about forming a kind of road map: How are we going to tackle things? What falls within the scope of the project and what will be excluded? In the ALwEN project, it was more or less agreed during the workshop that the project management should take the lead in organizing ethics-related activities but it did not result in a clear to-do list. Does this mean that no genuine agreement had been reached? My tentative answer would be that the people had reached an agreement, most particularly on the aims of the project and the framing of the problem (see also the discussion in Section 4.4). Also the convergence of opinions on the "scope" of the project (which activities are and which are not part of the project) is an example of agreement reached in the ALwEN project. Just like there are levels of coherence, there are levels of agreement, ranging from abstract initiatives to get everyone rowing in the same direction, to very concrete agreements on specific tasks. I think more practical insight is needed on the type of questions that are suitable for the WRE approach and also the type of outcomes (that is, the type of agreements) that could be reached. In the sphere of responsibility distributions, agreement on the scope of the project is, in my view, an example of a tangible and feasible agreement.

The second point concerns the objective or purpose of using reflective equilibrium methodology. In Chapter 3, I presented a taxonomy in which I distinguished between a justificatory, a descriptive, and a constructive use of the reflective equilibrium methodology. As often with analytical categories, the classification is sometimes a bit artificial. Concerning the constructive and justificatory uses of the method, these are not mutually exclusive. To the contrary, when the goal of the method is to reach a justified agreement, encouraging reflection is part of the strategy. In other words, the constructive use of the method is (partly at least) a derived aim of the justificatory use. This is also in line with the good reasoning-credible output strategy. I therefore wholeheartedly agree with Van Thiel who argues that "deliberation towards a

point of reflective equilibrium is an argumentative process. The justificatory power of the resulting RE [reflective equilibrium, ND] depends both on the strength of the moral view and on the argumentative force that is displayed in the reasoning process" (Van Thiel: 2009, 5). According to this "good reasoning-justified outcome strategy," deliberation is a derived subgoal that constitutes the justification of the outcome of the deliberative process.[117] In my own research, I started with the (rather theoretical) aim of reconciling different responsibility conceptions with a procedural justificatory model. During my research, however, I became aware of the importance of the deliberative process itself and the way it was encouraged by the method. The objective of using the WRE model therefore shifted from purely justificatory to encouraging discussion and attaining a justified agreement (a "hybrid" model). However, if the constructive use prevails over the justificatory one, the opponents of empirical ethics may be right in stating that the method lacks normative force. I think that the relation between this constructive and justificatory purpose needs therefore further exploration, especially in terms of the norms that make the deliberative process an adequate one (or, in terms of the "good reasoning-justified outcome strategy," a good one). In this thesis, I took the procedural norms inclusiveness and reflective learning to be the guiding ones. Especially the norm of higher order learning was also instrumental to deliberation and achieving a consensus. The question how this norm relates to the Rawlsian idea of "public reason" deserves closer attention, especially when the deliberative process is not in accordance to the demands of "public reason" but prompts reflective learning processes nevertheless.

---

[117] In this light it is also good to note that reflective equilibrium is a model for moral justification and not a (coherence) theory of moral truth. In applied ethics, reflective equilibrium reasoning is neither infallible nor aimed at moral truth. Van Thiel describes reflective equilibrium as "a model for moral inquiry that leads a thinker to a justifiable moral judgment or (modest) theory. The justification of the views attained through RE is thus not based on the claim that they are true. The objective of justification is reflective testing of all relevant considerations in order to produce a coherent moral view that boosts our confidence that we are not mistaken" (Van Thiel: 2009, 77).

7.4.   The role of the engineering ethicist revisited

Having discussed these last conceptual points, I would like to conclude with my vision on engineering ethics. Based on my experience in the ALwEN project, I see a positive role for embedded or parallel ethicists. I think that the future for applied research is multidisciplinary in nature, involving both social and natural scientists and researchers with a background in humanities. Of course, it is not always necessary to give all three disciplines a prominent role. However, the combination of the empirical social sciences and the more conceptual humanities promises to provide a valuable surplus to (technological) research. An ethicist or social scientist can provide a different perspective, represent the "general public," and bring to the fore issues that are otherwise overlooked.

In my ethical parallel research, I had a double role: I was both an observer with a specific research question in mind and an ethicist. In my role of observer, I deliberately tried to intervene as little as possible in the cooperation between the technical researchers in order to make the observations as "objective" as possible. After all, I was mainly interested in the researchers and engineers' own opinions. For this reason, I choose to present myself as an engineer understanding the "language" of engineers making my presence as "natural" as possible.

However, being an ethicist, I also had the desire to make the project as "responsible" as possible. This could potentially have required interventions during the project meetings – this was sometimes even explicitly asked by the engineers and technological researchers themselves when they asked my opinion on certain topics. There is potentially a tension between this role as the neutral observer and the intervening ethicist. Since I was primarily interested in the reasons why the researchers distributed the responsibilities the way they did and the underlying argumentation, I managed to keep the interventions to a minimum. In practice, the course of the project was such that, except for the organization of the workshop where I tried to facilitate a discussion on certain moral issues, the researchers themselves were very much aware of the moral issues pertaining to the technology and my main contribution was to make the engineers actually address them. Therefore, I think the collected data concerning the distribution rationales provides sufficient scientific rigor to base my (empirical) conclusions on.

When ethical investigations are not bounded by constraints that follow from, for example, a dissertation trajectory, an ethicist could take up a more profound role in the project. Still, I think that, dependent on the kind of questions one

wants to study, there does not need to be a tension between intervention and methodological soundness. Even when ethical parallel research requires active intervention, this does not necessarily disqualify the empirical results. I think the main requirement is to be aware of one's normative assumptions. In my research, my normative assumption (or maybe better: my personal agenda) was to make my research societally relevant, in the sense of trying to *improve* engineering practice rather looking for persons to blame for immoral behavior. The methodological literature on action research provides useful guidelines for balancing scientific rigor with bringing about change (Greenwood and Levin: 2007[1998]; Herr and Anderson: 2005; Reason and Bradbury: 2008[2001]).

More generally, I think the way forward for engineering ethics – or any branch of applied ethics – is to take their work beyond normative analysis and focus on how to bring about moral change in one's field of study. At this point, engineering ethicists could maybe learn from their colleagues working in medical ethics, where medical ethicists are represented in hospital decision-making bodies. The latter requires a thorough understanding of this field and sensitivity to context. This implies using insights from the social sciences, either by involving social scientists or by doing social scientific research oneself. I think this multidisciplinary approach is crucial. It requires a different mindset though. Only when social scientists dare giving up their fear of normativity and ethicists dare embracing the real world, will applied ethics become a flourishing (research) field that can benefit society.

Text box 7.1: Evaluation of ALwEN project and ethical parallel research

Already with the composition of the research team, the ALwEN consortium tried to differentiate itself from other projects by capturing the whole trajectory of fundamental research to the development of a prototype application and ultimately commercial exploitation. In order to do so, four universities, two independent industrial research institutes, one clinical partner and a consortium of 12 SMEs cooperated. At the start, the ambitions of the ALwEN team were high. In the project proposal it had set itself the goal of bringing the engineering science for such a technology to the level of commercial product viability. The aim was to develop a prototype Ambient Assisted Living (AAL) type application to monitor and assist the activities of the elderly in the context of an elderly home. In this pilot application, so it is mentioned in the project proposal, concepts and techniques required to safeguard security and privacy of the information collected through use of WSN could be tested and further developed. Rather than focusing on isolated aspects of the technology, the ALwEN consortium aimed at a more systematic and integral approach to scientifically understand all interactions, interferences, and cross-relations of WSN technology, such as to find the right balance and trade-offs on the system level.

In practice, the cooperation between the different partners proved difficult. Especially regarding the cooperation between the clinical partner and the technical partners, the team members adopted an attitude of waiting. The technical partners seemed to be waiting for instructions "how to establish social acceptance," whereas the clinical partners seemed to be unaware of the possibilities of WSN technology. The introduction of the COPD use case improved the communication somewhat, but the cooperation remained difficult. Since my own research question concerned the cooperation between partners and the distribution of responsibilities among them, the interview series I carried out in spring 2009 seemed to improve matters somewhat in the sense that the people became more aware of the main question: Who is responsible for addressing these issues or for initiating activities to address them? In the GDR workshop, held in October 2009, this was the guiding question. The main conclusion of the workshop was that it is hard to single out one partner responsible for all "ethical issues." The people agreed that it really is a joint effort and that it is primarily up to the management to coordinate this joint effort. Another conclusion was that the project, so far, was too much focused on fundamental research and too little on more applied research, including experimentation. This result prompted some refocus of the work and soon after the workshop, a meeting was scheduled in which plans for more realistic experiments were made. In this meeting, both the technical and non-technical requirements pertaining to the use case were discussed. In the first half of 2010, these real-life experiments were prepared and they are expected to be carried out in the beginning of 2011.

Text box 7.1 (continued)

With approximately one year left, the ALwEN project could be characterized as a highly ambitious and inquisitive project. Notwithstanding the different background and accompanying difficulties in their cooperation, the project members still aim at putting high priority to the social acceptance of the end product. In a concluding interview, I asked two team members with a formal role in the management of the project after their experiences with the ethical parallel investigations. They both expressed their appreciation of the involvement of an ethicist in the project and they argued that it had helped them giving "ethics" a more profound role in the project. One interviewee argued that the involvement of an ethicist can help making technical people more aware of things they otherwise overlook. Regarding future projects, they both thought it should be common practice to give an ethicist a formal role in technical projects during the whole course of the project. Both interviewees indicated that they see ethics as a relevant, but for themselves unknown, field of expertise. Since they considered themselves lacking the ethical expertise, they thought future projects would gain in quality by composing multidisciplinary teams. They consider ethics not as instrumental to successful technology implementation but rather an end in itself. Ideally, ethics should be seen as a "non-functional requirement that you cannot ignore," one of the interviewees remarked.

To summarize, even if the ALwEN project does not fully meet the high standards set at the start – something which cannot be judged yet – the cooperative attitude towards my ethical investigations indicate a sincere willingness to make technology societally relevant and acceptable. On a more general level, the enduring impact could be that the technical project members get a more concrete image of engineering ethics and the possibility to include ethicists in their future projects.

Bibliography

Alexy, R. (1990[1978]) 'A theory of practical discourse [trans. David Frisby]', in: Benhabib, S. and Dallmayr, F. (eds.), *The Communicative Ethics Controversy*, MIT Press, Cambridge, MA: 151-190.

Allhoff, F. and Vaidya, A. J. (eds.) (2008) *Professions in Ethical Focus: An Anthology*, Broadview Press, Peterborough, On.

Alpern, K. D. (1983) 'Moral responsibility for engineers', in: *Business and Professional Ethics Journal* 2 (2): 39-47.

Argyris, C. and Schön, D. A. (1978) *Organizational Learning: A Theory of Action Perspective*, Addison-Wesley, Reading.

Aristotle (1985) *Nicomachean Ethics [Trans. Terence Irwin]*, Hackett, Indianapolis.

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

Beauchamp, T. L. and Childress, J. F. (1994) *Principles of Biomedical Ethics*, Oxford University Press, New York.

––––– (2001[1994]) *Principles of Biomedical Ethics [Fifth Edition]*, Oxford University Press, New York.

Berloznik, R. and Van Langenhove, L. (1998) 'Integration of Technology Assessment in R&D management practices', in: *Technological Forecasting and Social Change* 58 (1-2): 23-33.

Berry, C. (2006) 'Corporate manslaughter', in: *Medicine Science and the Law* 46 (1): 2-6.

Bessette, J. M. (1980) 'Deliberative democracy: The majority principle in republican government', in: Goldwin, R. A. and Schambra, W. A. (eds.), *How Democratic is the Constitution*, American Enterprise Institute for Public Policy Research, Washington, DC: 102-116.

Birnbacher, D. (1999) 'Ethics and social science: Which kind of co-operation?', in: *Ethical Theory and Moral Practice* 2 (4): 319-336.

Bisarya, R. K. and Puri, S. (2005) 'The Bhopal gas tragedy: A perspective', in: *Journal of Loss Prevention in the Process Industries* 18 (4-6): 209-212.

Blackburn, S. (1993) 'Can philosophy exist?', in: Couture, J. and Nielsen, K. (eds.), *Méta-Philosophie: Reconstructing Philosophy? New Essays on Metaphilosophy*, University of Calgary Press, Calgary: 163-187.

Blattberg, C. (2003) 'Patriotic, not deliberative, democracy', in: *Critical Review of International Social and Political Philosophy* 6 (1): 155-174.

Bohman, J. (1998) 'The coming of age of deliberative democracy', in: *Journal of Political Philosophy* 6 (4): 400-425.

Bonjour, L. (1985) *The Structure of Empirical Knowledge*, Harvard University Press, Cambridge, MA.

Borry, P., Schotsmans, P. and Dierickx, K. (2004a) 'Editorial - Empirical Ethics: A challenge to bioethics', in: *Medicine, Health Care and Philosophy* 7 (1): 1-3.

——— (2004b) 'What is the role of empirical research in bioethical reflection and decision-making? An ethical analysis', in: *Medicine, Health Care and Philosophy* 7 (1): 41-53.

——— (2005) 'The birth of the empirical turn in bioethics', in: *Bioethics* 19 (1): 49-71.

Bovens, M. (1998) *The Quest for Responsibility. Accountability and Citizenship in Complex Organisations*, Cambridge University Press, Cambridge.

Brand-Ballard, J. (2003) 'Consistency, common morality, and reflective equilibrium', in: *Kennedy Institute of Ethics Journal* 13 (3): 231-258.

Brannigan, V. M. (2005) 'Teaching ethics in the engineering design process: A legal scholar's view ', in: *IEEE Antennas and Propagation Magazine* 47 (1): 146-151.

Brown, H. S., Vergragt, P., Green, K. and Berchicci, L. (2003) 'Learning for sustainability transition through bounded socio-technical experiments in personal mobility', in: *Technology Analysis & Strategic Management* 15 (3): 291-315.

Callon, M., Laredo, P., Rabeharisoa, V., Gonard, T. and Leray, T. (1992) 'The management and evaluation of technological programs and the dynamics of techno-economic networks: The case of the AFME', in: *Research Policy* 21 (3): 215-236.

Carr, S. (1975) 'Rawls, contractarianism and our moral intuitions', in: *The Personalist* 56 (1): 83-95.

Castleman, B. I. and Purkavastha, P. (1985) 'The Bhopal disaster as a case study in double standards ', in: Ives, J. H. (ed.), *The Export of Hazard*, Routledge, Boston.

Clausen, J. and Hansson, S. O. (2007) 'Eurocodes and REACH: Differences and Similarities', in: *Risk Management* 9 (1): 19-35.

Cohen, J. (1989) 'Deliberation and democratic legitimacy', in: Hamlin, A. and Pettit, P. H. (eds.), *The Good Polity: Normative Analysis of the State*, Blackwell Publishing, Oxford.

——— (1997) 'Procedure and substance in deliberative democracy', in: Bohman, J. and Rehg, W. (eds.), *Deliberative Democracy*, MIT Press, Cambridge, MA.

Cohen, J. and Rogers, J. (2003) 'Power and reason', in: Fung, A. and Wright, E. O. (eds.), *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*, Verso, New York: 237-258.

Cohen, S. J. (1997) 'Scientist-stakeholder collaboration in integrated assessment of climate change: Lessons from a case-study of northwest Canada', in: *Environmental Modelling and Assessment* 2 (4): 281-293.

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

Corlett, J. A. (2006) *Responsibility and Punishment*, Springer Science+Business Media, Dordrecht.

Coughlan, R. (2005) 'Codes, values and justifications in the ethical decision-making process', in: *Journal of Business Ethics* 59 (1-2): 45–53.

d'Astous, A. and Legendre, A. (2009) 'Understanding consumers' ethical justifications: A scale for appraising consumers' reasons for not behaving ethically', in: *Journal of Business Ethics* 87 (2): 255-268.

Daboub, A. J. and Calton, J. M. (2002) 'Stakeholder learning dialogues: How to preserve ethical responsibility in networks', in: *Journal of Business Ethics* 41: 85-98.

Dancy, J. (2000) *Normativity*, Blackwell Publishing, Oxford.

——— (2004) *Ethics without Principles*, Oxford University Press, Oxford.

——— (2006) 'Ethical non-naturalism', in: Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, New York.

Daniels, N. (1979) 'Wide Reflective Equilibrium and theory acceptance in ethics', in: *Journal of Philosophy* 76: 256-282.

——— (1996) *Justice and Justification: Reflective Equilibrium in Theory and Practice*, Cambridge University Press, Cambridge.

Davis, M. (1995) 'An historical preface to engineering ethics', in: *Science and Engineering Ethics* 1 (1): 33-48.

——— (1998) *Thinking like an Engineer*, Oxford University Press, Oxford.

——— (2002) *Profession, Code, and Ethics*, Ashgate Publishing, Farnham, Surrey.

——— (2003) 'What's philosophically interesting about engineering ethics?' in: *Science and Engineering Ethics* 9 (3): 353-361.

——— (ed.) (2005) *Engineering Ethics*, Ashgate Publishing, Aldershot.

——— (2009) 'Is engineering a profession everywhere?' in: *Philosophia* 37 (2): 211-225.

——— (forthcoming) '"Ain't No One Here But Us Social Forces": Constructing the professional responsibility of engineers', in: *Science and Engineering Ethics*: (DOI: 10.1007/s11948-010-9225-3).

De Vries, R. and Gordijn, B. (2009) 'Empirical ethics and its alleged meta-ethical fallacies', in: *Bioethics* 23 (4): 193-201.

DeGrazia, D. (1992) 'Moving forward in bioethical theory: Theories, cases, and specified principlism', in: *Journal of Medicine and Philosophy* 17 (5): 511-539.

——— (2003) 'Common morality, coherence and the principles of biomedical ethics', in: *Kennedy Institute of Ethics Journal* 13 (3): 219-230.

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

Doorn, N. (2010a) 'Applying Rawlsian approaches to resolve ethical issues: Inventory and setting of a research agenda', in: *Journal of Business Ethics* 91 (1): 127-143.

——— (2010b) 'A Rawlsian approach to distribute responsibilities in networks', in: *Science and Engineering Ethics* 16 (2): 221-249.

——— (2010c) 'A procedural approach to distributing responsibilities in R&D Networks', in: *Poiesis & Praxis. International Journal of Technology Assessment and Ethics of Science* 7 (3): 169-188.

——— (2011) 'Responsibility ascriptions in technology development and engineering: Three perspectives', in: *Science and Engineering Ethics*: DOI: 10.1007/s11948-009-9189-3.

——— (forthcoming) 'Responsibilities within an Ambient Assisted Living healthcare project', in: Proceeding of the Conference *ICT that makes the difference*, November 22-25, 2009, Brussels.

Doorn, N. and Nihlén Fahlquist, J. A. (2010) 'Responsibility in engineering: Towards a new role for engineering ethicists', in: *Bulletin of Science, Technology & Society* 30 (3): 222-230.

Doorn, N., Raven, R. P. J. M. and Royakkers, L. M. M. (2011) 'Distribution of responsibility in socio-technical networks: the Promest case', in: *Technology Analysis & Strategic Management* 23 (4): 453-471

Dower, N. (2004) 'Global economy, justice and sustainability', in: *Ethical Theory and Moral Practice* 7 (4): 399-415.

Draulans, V. (2010) 'Introduction', in: *Ethical Perspectives* 17 (2): 139-144.

Driessen, A. (2009) 'Ethical aspects of research in ultrafast communication', in: Sollie, P. and Düwell, M. (eds.), *Evaluating New Technologies*, Springer Science+Business Media B.V., Dordrecht: 11-19.

Dowding, K. (1995) 'Model or metaphor? A critical review of the policy network approach', in: *Political Studies* 43 (1): 136-158.

Dryzek, J. S. (1997) *The Politics of the Earth: Environmental Discourses*, Oxford University Press, Oxford.

Dryzek, J. S. and Niemeyer, S. (2006) 'Reconciling pluralism and consensus as political ideals', in: *American Journal of Political Science* 50 (3): 634-649.

Durbin, P. T. (1997) 'Engineering ethics and social responsibility: Reflections on recent developments in the USA', in: *Bulletin of Science, Technology & Society* 17 (2-3): 77-83.

——— (2008) 'Engineering professional ethics in a broader dimension', in: *Interdisciplinary Science Reviews* 33 (3): 226-233.

Düwell, M. (2009) 'What does medical ethics need empirical methods for? An inquiry in normative ethics', in: *Ethik in Der Medizin* 21 (3): 201-211.

Ebertz, R. P. (1993) 'Is reflective equilibrium a coherentist model?', in: *Canadian Journal of Philosophy* 23 (2): 193-214.

Einsiedel, E. F. (2002) 'Assessing a controversial medical technology: Canadian public consultations on xenotransplantation', in: *Public Understanding of Science* 11 (4): 315-331.

Einsiedel, E. F., Jelsoe, E. and Breck, T. (2001) 'Publics at the technology table: The consensus conference in Denmark, Canada, and Australia', in: *Public Understanding of Science* 10 (1): 83-98.

Elster, J. (1986) 'The market and the forum', in: Elster, J. and Hylland, A. (eds.), *Deliberative Democracy*, Cambridge University Press, Cambridge: 103-132.

——— (1992) *Local Justice. How Institutions Allocate Resources and Necessary Burdens*, Cambridge University Press, New York.

——— (1998) 'Deliberation and constitution making', in: Elster, J. (ed.), *Deliberative Democracy*, Cambridge University Press, New York: 97-122.

Eshleman, A. (2008) Moral responsibility. Zalta, E. N. *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition).* URL = <http://plato.stanford.edu/archives/fall2008/entries/moral-responsibility/>.

Feinberg, J. (1970) *Doing and Deserving. Essays in the Theory of Responsibility*, Princeton University Press, Princeton.

Fischer, F. (1980) *Politics, Values, and Public Policy*, Westview Press, Inc., Boulder, CA.

——— (1995) *Evaluating Public Policy*, Nelson-Hall, Chigaco.

Fischer, J. M. and Ravizza, M. (1993) 'Introduction', in: Fischer, J. M. and Ravizza, M. (eds.), *Perspectives on Moral Responsibility*, Cornell University Press, Ithaca: 1-41.

——— (1998) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, Cambridge.

Fisher, E. (2007) 'Ethnographic invention: Probing the capacity of laboratory decisions', in: *NanoEthics* 1 (2): 155-165.

Fisher, E. and Mahajan, R. L. (2006) *Midstream modulation of nanotechnology research in an academic laboratory*, ASME International Mechanical Engineering Congress and Exposition (IMECE2006), Chicago, Illinois, USA.

Fisher, E., Mahajan, R. L. and Mitcham, C. (2006) 'Midstream modulation of technology: Governance from within', in: *Bulletin of Science Technology Society* 26 (6): 485-496.

Fishkin, J. S. (1991) *Democracy and Deliberation: New Directions for Democratic Reform*, Yale University Press, New Haven.

Fishkin, J. S. and Laslett, P. (2003) *Debating Deliberative Democracy*, Blackwell Publishing Limited, Oxford.

Florman, S. C. (1976) *The Existential Pleasures of Engineering*, St. Martin's Press, New York.

Forst, R. (2001) 'The Rule of Reasons. Three Models of Deliberative Democracy', in: *Ratio Juris* 14 (4): 345-378.

Forsyth, D. R. (1980) 'A taxonomy of ethical ideologies', in: *Journal of Personality and Social Psychology* 39 (1): 175-184.

———— (1985) 'Individual differences in information integration during moral judgment', in: *Journal of Personality and Social Psychology* 49 (1): 264-272.

———— (1992) 'Judging the morality of business practices: The influence of personal moral philosophies', in: *Journal of Business Ethics* 11 (5-6): 461-470.

Forsyth, D. R., Bye, J. L. and Kelley, K. N. (1988) 'Idealism, relativism, and the ethic of caring', in: *Journal of Psychology* 122 (3): 243-248.

Forsyth, D. R., O'Boyle, E. H. and McDaniel, M. A. (2008) 'East meets West: A meta-analytic investigation of cultural variations in idealism and relativism', in: *Journal of Business Ethics* 83 (4): 813-833.

Frankena, W. K. (1939) 'The naturalistic fallacy', in: *Mind* 48 (192): 464-477.

Frankfurt, H. (1971) 'Freedom of the will and the concept of a person', in: *Journal of Philosophy* 68: 5-20.

———— (1988) *The Importance of What We Care About*, Cambridge University Press, Cambridge.

French, P. A. (1984) *Collective and Corporate Responsibility*, Columbia University Press, New York.

———— (1991) 'The corporation as a moral person', in: May, L. and Hoffman, S. (eds.), *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*, Rowman & Littlefield Publishers, Savage, MD.

Gert, B. (2004) *Common Morality: Deciding What to Do*, Oxford University Press, Oxford.

Gert, B., Culver, C. M. and Clouser, K. D. (1997) *Bioethics: a Return to Fundamentals*, Oxford University Press, Oxford.

Giacalone, R. A., Paul, K. and Jurkiewicz, C. L. (2005) 'A preliminary investigation into the role of positive psychology in consumer sensitivity

to corporate social performance', in: *Journal of Business Ethics* 58 (4): 295-305.

Glenna, L. L. (2010) 'Value-laden technocratic management and environmental conflicts: The case of the New York City watershed controversy', in: *Science, Technology, & Human Values* 35 (1): 81-112.

Goodin, R. E. (1985) *Protecting the Vulnerable. A Reanalysis of our Social Responsibilities*, Chicago University Press, Chicago.

——— (1995) *Utilitarianism as a Public Philosophy*, Cambridge University Press, Cambridge.

Gracia, D. (1995) 'Hard times, hard choices: Founding bioethics today', in: *Bioethics* 9 (3-4): 192-206.

Greenwood, D. J. and Levin, M. (2007[1998]) *Introduction to Action Research: Social Research for Social Change. 2nd Edition*, SAGE Publications, Thousand Oaks.

Grin, J. and Hoppe, R. (1995) 'Toward a comparative framework for learning from experiences with interactive Technology Assessment', in: *Industrial & Environmental Crisis Quarterly* 9 (1): 99-120.

Grin, J. and van der Graaf, H. (1996a) 'Technology Assessment as learning', in: *Science, Technology, & Human Values* 21 (1): 72-99.

——— (1996b) 'Implementation as communicative action: An interpretative understanding of interactions between policy actors and target groups', in: *Policy Sciences* 29: 291-319.

Guston, D. H. and Sarewitz, D. (2002) 'Real-time Technology Assessment', in: *Technology in Society* 24 (1-2): 93-109.

Gutmann, A. and Thompson, D. (1996) *Democracy and Disagreement*, Harvard University Press, Cambridge, MA.

——— (2004) *Why Deliberative Democracy*, Princeton University Press, Princeton.

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

——— (1990b) 'Discourse ethics: Notes on a program of philosophical justification', in: Benhabib, S. and Dallmayr, F. (eds.), *The Communicative Ethics Controversy*, MIT Press, Cambridge, MA: 60-110.

——— (1995) 'Reconciliation through the public use of reason: Remarks on John Rawls's Political Liberalism', in: *The Journal of Philosophy* 92 (3): 109-131.

Haidt, J. (2001) 'The emotional dog and its rational tail: A social intuitionist approach to moral judgment', in: *Psychological Review* 108 (4): 814-834.

Halliday, J. (1999) 'Political liberalism and citizenship education: Towards curriculum reform', in: *British Journal of Educational Studies* 47 (1): 43-55.

Hanson, C. E., Palutikof, J. P., Dlugolecki, A. and Giannakopoulos, C. (2006) 'Bridging the gap between science and the stakeholder: The case of climate change research', in: *Climate Research* 31 (1): 121-133.

Hansson, S. O. (2006) 'Informed consent out of context', in: *Journal of Business Ethics* 63 (2): 149-154.

——— (2009) 'Risk and safety in technology', in: Meijers, A. W. M. (ed.), *Handbook Philosophy of Technology and Engineering Sciences*, Elsevier, Oxford.

Hardin, R. (1999) 'From Bodo ethics to distributive justice', in: *Ethical Theory and Moral Practice* 2 (4): 399-413.

Hare, R. M. (1952) *The Language of Morals*, Clarendon Press, Oxford.

——— (1988) 'Why do applied ethics?', in: Rosenthal, D. M. and Shehadi, F. (eds.), *Applied Ethics and Ethical Theory*, University of Utah Press, Salt Lake City: 71-83.

Harris, C. E. (2008) 'The good engineer: Giving virtue its due in engineering ethics', in: *Science and Engineering Ethics* 14 (2): 153-164.

Harris, C. E., Pritchard, M. S. and Rabins, M. J. (2005[1995]) *Engineering Ethics: Concepts and Cases (Third Edition)*, Wadsworth, Belmont, CA.

Hart, H. L. A. (1961) *The Concept of Law*, The Clarendon Press, Oxford.

——— (1968) *Punishment and Responsibility. Essays in the Philosophy of Law.*, Clarendon Press, Oxford.

Hart, H. L. A. and Honoré, T. (1985) *Causation in the Law*, Clarendon Press, Oxford.

Henneberg, S. C., Mouzas, S. and Naudé, P. (2006) 'Network pictures: Concepts and representation', in: *European Journal of Marketing* 40 (3/4): 408-429.

Herr, K. and Anderson, G. L. (2005) *The Action Research Dissertation: A Guide for Students and Faculty*, SAGE Publications, Thousand Oaks.

Hilhorst, M. T. (2005) 'Directed altruistic living organ donation: Partial but not unfair ', in: *Ethical Theory and Moral Practice* 8 (1-2): 197-215.

Ho, A. (2003) 'International business vs. globalization: Implications for business ethics', in: *Business and Professional Ethics Journal* 22 (2): 51-69.

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

Holmgren, M. (1987) 'Wide Reflective Equilibrium and objective moral truth', in: *Metaphilosophy* 18 (2): 108-124.

Honoré, T. (1999) *Responsibility and Fault*, Hart, Oxford.

Hoogma, R., Weber, M. and Elzen, B. (2001) *Integrated long-term strategies to induce regume shifts to sustainability*, Towards Environmental Innovation Systems, Eibsee, Germany.

Hooker, B. and Little, M. (eds.) (2000) *Moral Particularism*, Oxford University Press, Oxford.

Hume, D. (1739) *A Treatise of Human Nature*, John Noon, London.

Jamali, D. (2008) 'A stakeholder approach to Corporate Social Responsibility: A fresh perspective into theory and practice ', in: *Journal of Business Ethics* 82 (1): 213-231.

Joss, S. and Bellucci, S. (eds.) (2002) *Participatory Technology Assessment. European Perspectives*, Athenaeum Press, Gateshead, Tyne and Wear, UK.

Joss, S. and Durant, J. (eds.) (1995) *Public Participation in Science: The Role of Consensus Conferences in Europe*, Trustees of the Science Museum, London.

Kant, I. (1795) *Zum ewigen Frieden: Ein philosophischer Entwurf*, Friedrich Nicolovius, Königsberg.

Kekes, J. (1986) 'Moral intuition', in: *American Philosophical Quarterly* 23 (1): 83-93.

Kensing, F. (2003) *Methods and Practices in Participatory Design*, ITU Press, Copenhagen.

Kimball, B. A. (1995a) 'Introduction', in: Orrill, R. (ed.), *The Condition of Liberal Education*, College Entrance Examination Board, New York: xxii-xxiii.

———— (1995b) 'Toward pragmatic liberal education', in: Orrill, R. (ed.), *The Condition of Liberal Education*, College Entrance Examination Board, New York: 3-127.

KIVI NIRIA (2006) Gedragscode / Code of Conduct, The Netherlands.

Koepsell, D. R. (2009) *Who Owns You: The Corporate Gold Rush to Patent Your Genes*, Wiley-Blackwell, Chichester.

Knoke, D. (1990) *Political Networks: The Structural Perspective*, Cambridge University Press, Cambridge.

Kutz, C. (2000) *Complicity: Ethics and Law for a Collective Age*, Cambridge University Press, Cambridge.

La Torre, M. (1999) 'Legal pluralism as evolutionary achievement of community law', in: *Ratio Juris: An International Journal of Jurisprudence and Philosophy of Law* 12 (2): 182-195.

Ladd, J. (1982) 'Collective and individual moral responsibility in engineering: Some questions', in: *IEEE Technology and Society Magazine* 1 (2): 3-10.

———— (1989) 'Computers and moral responsibility: A framework for an ethical analysis', in: Gould, C. C. (ed.), *The Information Web*, Westview Press, Boulder, CA: 207-229.

———— (1991) 'Bhopal: An essay on moral responsibility and civic virtue', in: *Journal of Social Philosophy* 32 (1): 73-91.

Laird, F. N. (1993) 'Participatory analysis, democracy, and technological decision making', in: *Science, Technology, & Human Values* 18 (3): 341-361.

Lehning, P. B. (2001) 'European citizenship: Towards a European identity?', in: *Law and Philosophy* 20 (3): 239-282.

Lewis, H. D. (1948/1991) 'Collective responsibility', in: May, L. and Hoffman, S. (eds.), *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*, Rowman & Littlefield Publishers, Savage, MD.

List, C. and Pettit, P. (2004) 'Aggregating sets of judgments: Two impossibility results compared', in: *Synthese* 140 (1-2): 207-235.

Little, M. O. (1999) 'Abortion, intimacy, and the duty to gestate', in: *Ethical Theory and Moral Practice* 2 (3): 295-312.

Lustig, B. A. (1992) 'The method of principlism: A critique of the critique', in: *Journal of Medicine and Philosophy* 17 (5): 487-510.

Lynch, W. T. and Kline, R. (2000) 'Engineering practice and engineering ethics', in: *Science, Technology, & Human Values* 25 (2): 195-225.

MacIntyre, A. (1984[1981]) *After Virtue: A Study in Moral Theory*, University of Notre Dame Press, Notre Dame.

Mackie, J. L. (1978) *Ethics: Inventing Right and Wrong*, Penguin Books, Harmondsworth.

Magill, K. (2000) 'Blaming, understanding, and justification', in: van den Beld, T. (ed.), *Moral Responsibility and Ontology*, Kluwer Academic Publishers, Dordrecht.

Manders-Huits, N. and Van den Hoven, M. J. (2009) 'The moral relevance of technological artifacts', in: Sollie, P. and Düwell, M. (eds.), *Evaluating New Technologies*, Springer Science+Business Media B.V., Dordrecht: 51-60.

Manin, B. (1987) 'On legitimacy and political deliberation', in: *Political Theory* 15 (3): 338-368.

March, A. F. (2006) 'Liberal citizenship and the search for an overlapping consensus: The case of Muslim minorities', in: *Philosophy & Public Affairs* 34 (4): 373-421.

Marcus, R. B. (1980) 'Moral dilemmas and consistency', in: *The Journal of Philosophy* 77 (3): 121-136.

Martin, M. W. and Schinzinger, R. (2005) *Ethics in Engineering*, McGraw-Hill, New York.

May, L. (1992) *Sharing Responsibility*, Chicago University Press, Chicago.

May, L. and Hoffman, S. (1991) 'Introduction', in: May, L. and Hoffman, S. (eds.), *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*, Rowman & Littlefield Publishers, Savage, MD.

McCarthy, T. (1994) 'Kantian constructivism and reconstructivism: Rawls and Habermas in dialogue', in: *Ethics* 105 (1): 44-63.

McClain, L. C. (1998) 'Deliberative democracy, overlapping consensus, and same-sex marriage', in: *Fordham Law Review* 66 (4): 1241-1252.

McMillan, J. and Hope, T. (2008) 'The possibility of empirical psychiatric ethics', in: Widdershoven, G. A. M., McMillan, J., Hope, T. and Van der Scheer, L. (eds.), *Empirical Ethics in Psychiatry*, Oxford University Press, Oxford: 9-22.

Mehalik, M. M. (2000) 'Sustainable network design: A commercial fabric case study', in: *Interfaces* 30 (3; Special issue: Sustainable Business): 180-189.

Milewa, T. (2006) 'Health technology adoption and the politics of governance in the UK', in: *Social Science & Medicine* 63 (12): 3102-3112.

Miller, D. (2001) 'Distributing responsibilities', in: *The Journal of Political Philosophy* 9 (4): 453-471.

——— (2004) 'Holding nations responsible', in: *Ethics* 114: 240-268.

Molewijk, B. and Frith, L. (2009) 'Empirical ethics: Who is the Don Quixote?', in: *Bioethics* 23 (4): II-IV.

Molewijk, B., Stiggelbout, A. M., Otten, W., Dupuis, H. M. and Kievit, J. (2004) 'Empirical data and moral theory: A plea for integrated empirical ethics', in: *Medicine, Health Care and Philosophy* 7: 55-69.

Möller, K. (2010) 'Sense-making and agenda construction in emerging business networks: How to direct radical innovation', in: *Industrial Marketing Management* 39 (3): 361-371.

Moore, G. E. (1903) *Principia Ethics*, Cambridge University Press, Cambridge.

Moran, J. (2006) 'Religious reasons and political argumentation', in: *Journal of Religious Ethics* 34 (3): 421-437.

Mouffe, C. (1999) 'Deliberative democracy or agonistic pluralism?', in: *Social Research* 66 (3): 745-758.

——— (2000) *The Democratic Paradox*, Verso, London.

Musschenga, A. W. (2005) 'Empirical ethics, context-sensitivity, and contextualism', in: *Journal of Medicine and Philosophy* 30 (5): 467-490.

Nagel, T. (1979) *Mortal Questions*, Cambridge University Press, Cambridge.

Neumann, P. G. (1991) 'Inside risks: Certifying professionals', in: *Communications of the ACM* 34 (2): 130.

Nielsen, K. (1982a) 'Grounding rights and a method of reflective equilibrium', in: *Inquiry* 25 (3): 277-306.

——— (1982b) 'Considered judgements again', in: *Human Studies* 5 (2): 117-129.

Nihlén Fahlquist, J. (2006a) 'Responsibility ascriptions and public health problems: Who is responsible for obesity and lung cancer?', in: *Journal of Public Health* 14 (1): 15-19.

——— (2006b) 'Responsibility ascriptions and Vision Zero', in: *Accident Analysis and Prevention* 38: 1113-1118.

───── (2009) 'Moral responsibility for environmental problems: Individual or institutional?', in: *Journal of Agricultural and Environmental Ethics* 22 (2): 109-124.

Nijsingh, N. and Düwell, M. (2009) 'Interdisciplinarity, applied ethics and social science', in: Sollie, P. and Düwell, M. (eds.), *Evaluating New Technologies*, Springer Science+Business Media B.V., Dordrecht: 79-92.

Nissenbaum, H. (1994) 'Computing and accountability', in: *Communications of the ACM* 37 (1): 73-80.

───── (1996) 'Accountability in a computerized society', in: *Science and Engineering Ethics* 2 (1): 25-42.

Norman, W. (1998) '"Inevitable and Unacceptable": Methodological Rawlsianism in Anglo-American political philosophy', in: *Political Studies* 46 (2): 276-294.

Nowell-Smith, P. H. (1954) *Ethics*, Penguin Books, Baltimore.

Nozick, R. (1974) *Anarchy, State, and Utopia*, Basic Books, New York.

O'Fallon, M. J. and Butterfield, K. D. (2005) 'A review of the empirical ethical decision-making literature: 1996–2003', in: *Journal of Business Ethics* 59 (4): 375-413.

Oakley, J. and Cocking, D. (2001) 'Virtue Ethics and Professional Roles', Cambridge University Press, Cambridge.

OECD (2010) OECD Factbook 2010: Economic, Environmental and Social Statistics, OECD Publishing, DOI: 10.1787/factbook-2010-en.

Paris, D. C. and Kimball, B. A. (2000) 'Liberal education: An overlapping pragmatic consensus', in: *Journal of Curriculum Studies* 32 (2): 143-158.

Pellegrino, E. D. (1995) 'The limitation of empirical-research in ethics', in: *Journal of Clinical Ethics* 6 (2): 161-162.

Perhac, R. M. (1998) 'Comparative risk assessment: Where does the public fit in?' in: *Science, Technology, & Human Values* 23 (2): 221-241.

Perrott, D. L. (1982) 'Changes in attitude to limited liability: The European experience', in: Orhnial, T. (ed.), *Limited Liability and the Corporation*, Croom Helm, London: 81-121.

Pettit, P. (2001) 'Deliberative democracy and the discursive dilemma', in: *Nous*: 268-299.

───── (2007) 'Responsibility incorporated', in: *Ethics* 117 (2): 171-201.

Pigden, C. R. (1989) 'Logic and the autonomy of ethics', in: *Australasian Journal of Philosophy* 67 (2): 127-151.

Pogge, T. W. (2007) *John Rawls: His Life and Theory of Justice*, Oxford University Press, Oxford.

Preston, T. (2004) 'Environmental values, pluralism, and stability', in: *Ethics, Place and Environment* 7 (1-2): 73-83.

Pritchard, M. S. (2001) 'Responsible engineering: The importance of character and imagination', in: *Science and Engineering Ethics* 7 (3): 391-402.

Pritchard, M. S. and Holtzapple, M. (1997) 'Responsible engineering: *Gilbane Gold* revisited', in: *Science and Engineering Ethics* 3 (2): 217-230.

Rauprich, O. (2008) 'Common morality: Comment on Beauchamp and Childress', in: *Theoretical Medicine and Bioethics* 29 (1): 43-71.

Rawls, J. (1951) 'Outline of a decision procedure for ethics', in: *Philosophical Review* 60 (2): 177-197.

——— (1993) *Political Liberalism*, Columbia University Press, New York.

——— (1995) 'Political Liberalism: Reply to Habermas', in: *The Journal of Philosophy* 92 (3): 132-180.

——— (1997) 'The idea of public reason revisited', in: *University of Chicago Law Review* 64.

——— (1999) *Collected Papers*, Harvard University Press, Cambridge, MA.

——— (1999 [1971]) *A Theory of Justice*, The Belknap Press of Harvard University Press, Cambridge, MA.

——— (2001) *Justice as Fairness: A Restatement.*, The Belknap Press of Harvard University Press, Cambridge, MA.

Reason, P. and Bradbury, H. (eds.) (2008[2001]) *Handbook of Action Research: Participative Inquiry and Practice [Second Edition]*, SAGE Publications, Thousand Oaks.

Renn, O. (1999) 'A model for an analytic-deliberative process in risk management', in: *Environmental Science & Technology* 33 (18): 3049-3055.

——— (2006) 'Participatory processes for designing environmental policies', in: *Land Use Policy* 23 (1): 34-43.

——— (2008) 'An ethical appraisal of hormesis: Toward a rational discourse on the acceptability of risks and benefits', in: *Human & Experimental Toxicology* 27 (8): 627-642.

Reuzel, R. P. B., Van der Wilt, G. J., Ten Have, H. A. M. J. and De Vries Robbé, P. F. (1999) 'Reducing normative bias in health Technology Assessment: Interactive evaluation and casuistry', in: *Medicine, Health Care and Philosophy* 2 (3): 255-263.

Reuzel, R. P. B., Van der Wilt, G. J., Ten Have, H. A. M. J. and Robbe, P. E. D. (2001) 'Interactive Technology Assessment and Wide Reflective Equilibrium', in: *Journal of Medicine and Philosophy* 26 (3): 245-261.

Richardson, B. and Curwen, P. (1995) 'Do free-market governments create crisis-ridden societies', in: *Journal of Business Ethics* 14 (7): 551-560.

Richardson, H. S. (1997) *Practical Reasoning about Final Ends*, Cambridge University Press, Cambridge.

Rip, A. (2006) 'A co-evolutionary approach to reflexive governance - and its ironies', in: Voß, J. P., Bauknecht, D. and Kemp, R. (eds.), *Reflexive Governance for Sustainable Development*, Edward Elgar, Cheltenham: 82-100.

Rip, A., Misa, T. J. and Schot, J. (eds.) (1995) *Managing Technology in Society. The Approach of Constructive Technology Assessment*, Cassell Publishers Limited, London.

Robertson, C. J., Olson, B. J., Gilley, K. M. and Bao, Y. (2008) 'A cross-cultural comparison of ethical orientations and willingness to sacrifice ethical standards: China versus Peru', in: *Journal of Business Ethics* 81 (2): 413–425.

Rogers, G. F. C. (1983) *The Nature of Engineering: A Philosophy of Technology*, Macmillan, London.

Rogers, J. D. and Bozeman, B. (2001) '"Knowledge Value Alliances": An alternative to the R&D project focus in evaluation', in: *Science, Technology, & Human Values* 26 (1): 23-55.

Rowe, G. and Frewer, L. J. (2004) 'Evaluating public-participation exercises: A research agenda', in: *Science, Technology, & Human Values* 29 (4): 512-557.

Saari, E. and Miettinen, R. (2001) 'Dynamics of change in research work: Constructing a new research area in a research group', in: *Science, Technology, & Human Values* 26 (3): 300-321.

Sabatier, P. A. and Jenkins-Smith, H. C. (1993) *Policy Change and Learning: An Advocacy Coalition Approach*, Westview Press, Inc., Boulder.

Sarewitz, D. (2005) 'This won't hurt a bit: Assessing and governing rapidly advancing technologies in a democracy', in: Rodemeyer, M., Sarewitz, D. and Wilsdon, J. (eds.), *The future of technology assessment*, Woodrow Wilson International Center for Scholars, Washington, DC: 14-21.

Scanlon, T. M. (1982) 'Contractualism and Utilitarianism', in: Sen, A. K. and Williams, B. A. O. (eds.), *Utilitarianism and Beyond*, Cambridge University Press, Cambridge: 130-128.

Schinzinger, R. (1998) 'Ethics on the feedback loop', in: *Control Engineering Practice* 6: 239-245.

Schmidt, V. H. (1994) 'Bounded justice', in: *Social Science Information* 33: 305-333.

Schön, D. A. (1983) *The Reflective Practitioner. How Professionals Think in Action.*, Basic Books, New York.

Schot, J. W. (2001) 'Towards new forms of participatory technology development', in: *Technology Analysis & Strategic Management* 13 (1): 39-52.

Schot, J. W. and Rip, A. (1997) 'The past and future of constructive Technology Assessment', in: *Technological Forecasting and Social Change* 54 (2/3): 251-268.

Schuler, D. and Namioka, A. (eds.) (1993) *Participatory design: Principles and practices*, Erlbaum, Hillsdale, NJ.

Sclove, R. E. (1995) *Democracy and Technology*, The Guilford Press, New York.

Sencerz, S. (1986) 'Moral institutions and justification in ethics', in: *Philosophical Studies* 50: 77-95.

Senge, P. M. (1990) 'The Leader's New Work: Building Learning Organizations', in: *Sloan Management Review*: 7-23.

Sheridan, H. (2007) 'Evaluating technical and technological innovations in sport: Why fair play isn't enough', in: *Journal of Sport & Social Issues* 31 (2): 179-194.

Singer, P. (1974) 'Sidgwick and Reflective Equilibrium', in: *The Monist* 58: 490-517.

St. John, J. (2007) 'Problems with theory, problems with practice: Wide Reflective Equilibrium and bioethics', in: *South African Journal of Philosophy* 26 (2): 204-215.

Strawson, P. F. (1974) 'Freedom and Resentment', in: Strawson, P. F. (ed.), *Freedom and Resentment and other Essays*, Methuen, London: 1-25.

Sunstein, C. R. (2005) *Laws of Fear*, Cambridge University Press, Cambridge.

Swierstra, T. E. and Jelsma, J. (2006) 'Responsibility without moralism in techno-scientific design practice', in: *Science, Technology, & Human Values* 31 (3): 309-332.

Swierstra, T. E. and Rip, A. (2007) 'Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology', in: *NanoEthics* 1 (1): 3-20.

Tan, K. C. (1998) 'Liberal toleration in Rawls's law of peoples', in: *Ethics* 108 (2): 276-295.

Thompson, D. F. (1980) 'Moral responsibility and public officials', in: *American Political Science Review* 74: 905-916.

Tobin, T.-W. (2005) 'The non-modularity of moral knowledge: Implications for the universality of human rights ', in: Rowan, J. R. (ed.), *Social Philosophy Today: Human Rights, Religion, and Democracy*, Philosophy Documentation Center, Charlottesville: 33-50.

Twiss, S. B. (2004) 'History, human rights, and globalization', in: *Journal of Religious Ethics* 32 (1): 39-70.

Van Asselt, M. B. A. and Rijkens-Klomp, N. (2002) 'A look in the mirror: Reflection on participation in Integrated Assessment from a methodological perspective', in: *Global Environmental Change* 12: 167-184.

Van de Kerkhof, M. (2006) 'Making a difference: On the constraints of consensus building and the relevance of deliberation in stakeholder dialogues', in: *Policy Sciences* 39 (3): 279-299.

Van de Poel, I. R. (2008) 'How should we do nanoethics? A network approach for discerning ethical issues in nanotechnology', in: *NanoEthics* 2 (1): 25-38.

Van de Poel, I. R. and Royakkers, L. M. M. (2007) 'The ethical cycle', in: *Journal of Business Ethics* 71 (1): 1-13.

Van de Poel, I. R. and Van Gorp, A. C. (2006) 'The need for ethical reflection in engineering design: The relevance of type of design and design hierarchy', in: *Science, Technology, & Human Values* 31 (3): 333-360.

Van de Poel, I. R. and Verbeek, P. P. (2006) 'Ethics and engineering design', in: *Science, Technology, & Human Values* 31 (3): 223-236.

Van de Poel, I. R. and Zwart, S. D. (2010) 'Reflective Equilibrium in R&D networks', in: *Science, Technology, & Human Values* 35 (2): 174-199.

Van de Scheer, L. and Widdershoven, G. A. M. (2004) 'Integrated empirical ethics: Loss of normativity', in: *Medicine, Health Care and Philosophy* 7 (1): 71-79.

Van den Hoven, M. J. (1997) 'Computer ethics and moral methodology', in: *Metaphilosophy* 28 (3): 234-248.

——— (1998) 'Moral Responsibility, Public Office and Information Technology', in: Snellen, I. T. M. and Van de Donk, W. B. H. J. (eds.), *Public Administration in an Information Age. A Handbook*, IOS Press, Amsterdam: 97-111.

Van der Burg, S. (2009) 'Imagining the future of photoacoustic mammography', in: *Science and Engineering Ethics* 15 (1): 97-110.

Van der Burg, W. and Van Willigenburg, T. (1998) 'Introduction', in: Van der Burg, W. and Van Willigenburg, T. (eds.), *Reflective Equilibrium: Essays in the Honour of Robert Heeger*, Kluwer Academic Publishers, Dordrecht: 1-25.

Van Hooft, S. (2006) *Understanding Virtue Ethics*, Acumen Publishing, Chesham.

Van Thiel, G. J. M. W. (2009) *Moral Wisdom in the Balance: Reflective Equilibrium as a Normative Empirical Model for Bioethics [PhD thesis]*, Utrecht University, Utrecht.

Van Thiel, G. J. M. W. and Van Delden, J. J. M. (2009) 'The justificatory power of moral experience', in: *Journal of Medical Ethics* 35 (4): 234-237.

——— (2010) 'Reflective Equilibrium as a normative empirical model', in: *Ethical Perspectives* 17 (2): 183-202.

Van Velsen, J. F. C. (2000) 'Relativity, universality, and peaceful coexistence', in: *Archiv für Rechts- und Sozialphilosophie* 86 (1): 88-108.

Vanderburg, W. H. (2000) *The Labyrinth of Technology*, University of Toronto Press, Toronto, Ontario.

————— (2006) 'Can the university escape from the labyrinth of technology? Part 3: A strategy for transforming the professions', in: *Bulletin of Science Technology Society* 26 (3): 189-203.

————— (2009) 'Rethinking engineering design and decision making in response to economic, social, and environmental crises', in: *Bulletin of Science Technology Society* 29 (5): 421-432.

Vaughan, D. (1996) *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*, University of Chicago Press, Chicago.

Vedder, A. (2001) 'Accountability of Internet access and service providers: Strict liability entering ethics?', in: *Ethics of Information Technology* 3 (1): 67-74.

Vincenti, W. G. (1990) *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History*, The Johns Hopkins University Press, Baltimore.

Wallace, R. J. (1994) *Responsibility and the Moral Sentiments*, Harvard University Press Cambridge, MA.

Walzer, M. (1983) *Spheres of Justice: A Defense of Pluralism and Equality.* , Basic Books Inc., New York.

————— (1987) *Interpretation and Social Criticism*, Harvard University Press, Cambridge, MA.

————— (2002) *The Company of Critics: Social Criticism and Political Commitment in the Twentieth Century*, Basic Books, New York.

Watson, G. (1996) 'Two faces of responsibility', in: *Philosophical Topics* 24: 227-248.

Weaver, G. R. and Trevino, L. K. (1994) 'Normative and Empirical Business Ethics: Seperation, Marriage of Convenience, or Marriage of Necessity?', in: *Business Ethics Quarterly* 4 (2): 129-143.

Whitbeck, C. (1998) *Ethics in Engineering Practice and Research*, Cambridge University Press Cambridge.

Williams, B. A. O. (1985) *Ethics and the Limits of Philosophy*, Harvard University Press, Cambridge, MA.

Williams, G. (2008) 'Responsibility as a Virtue', in: *Ethical Theory and Moral Practice* 11 (4): 455-470.

Wilsdon, J. (2005) 'Paddling upstream: New currents in European technology assessment', in: Rodemeyer, M., Sarewitz, D. and Wilsdon, J. (eds.), *The future of technology assessment*, Woodrow Wilson International Center for Scholars, Washington, DC: 22-29.

Yin, R. K. (2003) *Case Study Research: Design and Methods [Third Edition]*, SAGE Publications, Thousand Oaks.

Young, I. M. (1996) 'Communication and the Other: Beyond Deliberative Democracy', in: Benhabib, S. (ed.), *Democracy and Difference: Contesting the Boundaries of the Political,* Princeton University Press, Princeton.

——— (2000) *Inclusion and Democracy,* Oxfort University Press, Oxford.

——— (2006) 'Responsibility and global justice: A social connection model', in: *Social Philosophy and Policy* 23 (1): 102-130.

Zandvoort, H. (2005a) 'Knowledge, risk, and liability: Analysis of a discussion continuing within science and technology', in: *Cognitive Structures in Scientific Inquiry: Essays in Debate with Theo Kuipers, Vol 2* 84: 469-501.

——— (2005b) 'Globalisation, environmental harm and progress: The role of consensus and liability', in: *Water Science & Technology* 52 (6): 43-50.

——— (2008) 'Risk zoning and risk decision making', in: *International Journal of Risk Assessment and Management* 8 (1-2): 3-18.

Zarkada-Fraser, A. (2000) 'A classification of factors influencing participating in collusive tendering agreements', in: *Journal of Business Ethics* 23 (3): 269-282.

Zimmerman, M. (1988) *An Essay on Moral Responsibility,* Roman and Littlefield Totowa, NJ.

Zwart, S. D., Van de Poel, I. R., Van Mil, H. and Brumsen, M. (2006) 'A network approach for distinguishing ethical issues in Research and Development', in: *Science and Engineering Ethics* 12 (4): 663-684.

Zweigert, K. and Kötz, H. (1998[1977]) *Introduction to Comparative Law,* Clarendon Press, Oxford.

Index

Summary

Moral responsibility in R&D networks

This thesis is about moral responsibility in the context of Research and Development (R&D). Moral responsibility is one of the core notions in ethics and it is also widely discussed in both the general and the more applied philosophy and ethics literature. However, few articles explicitly discuss moral responsibility in the context of R&D. This is surprising given the way R&D is organized. First, technology is often developed in complex network organizations, involving many different actors and institutions. These networks often lack a strict hierarchy and a clear task division. This increases the chance that certain aspects are not correctly looked after because people expect someone else to do or have done it. Secondly, technology development involves a long causal chain from fundamental research to experimentation and (often) commercial exploitation. The negative consequences (risks and unforeseen side-effects) often only materialize during use, which makes it difficult to predict them with any accuracy beforehand. It is therefore desirable to pay special attention to potential risks and uncertainties and to look beyond the obvious and known consequences. The complex organization and the long causal chain together make R&D especially vulnerable for what is in the literature known as the problem of many hands; that is, the difficulty of identifying the person responsible if a large number of people are involved.

The occurrence of the problem of many hands in R&D is especially undesirable since the impact of technology, including its negative consequences, is often high (e.g., the use of asbestos, CFCs, DDT, nuclear waste and the greenhouse effect). Moreover, the pace of technological development is increasing, often running ahead of adequate legislation. In other words, while the impact of technology and the pace of its development increase, the contexts in which technology is developed increasingly blur the responsibility for societal consequences. These developments indicate that the problem of many hands is a very relevant and urgent problem in R&D networks.

Formulated very generally, the aim of this research is to alleviate the problem of many hands in R&D networks. There are different ways to interpret the problem of many hands; in this thesis, I conceive of it as the problem that stems from a pluralism of responsibility conceptions. My particular research

objective is therefore to develop a framework for resolving responsibility issues in R&D in a situation of pluralism. This pluralism of responsibility conceptions leads to different, sometimes conflicting, requirements for responsibility distributions. A procedural approach for distributing responsibilities is proposed to alleviate the tension between these different requirements.

Hence, the main argument put forward in this thesis hinges on three ideas:

1. the problem of many hands is partly a problem stemming from different conceptions of responsibility. People have different opinions on when it is fair to ascribe responsibility to someone (*fairness requirements*);

2. these different conceptions are all legitimate (under certain conditions) and cannot be reduced to one overarching view on responsibility (*pluralism of responsibility conceptions*);

3. if a procedure for distributing responsibilities is accepted by all as fair, the outcome can be considered fair as well (*procedural fairness*).

The central research question is how a procedural approach, based on Rawls' idea of Wide Reflective Equilibrium can be (further) developed and used to resolve the tension between competing conceptions of responsibility. This question is partly an empirical one, including qualitative subquestions: How do researchers distribute responsibilities? Why do they judge some responsibility ascriptions as fair and others not? What are the different rationales for distributing responsibilities? To answer these kinds questions – or at least tentatively explore them – I performed an ethical case study on a real technical research project: the ALwEN project. The ALwEN project, which is an acronym for Ambient Living with Embedded Networks, concerns the development of an in-house monitoring system based on Ambient Intelligence technology. This so-called ethical parallel research allowed me to gain insight in the distributing rationales of engineers and technological researchers involved in R&D.

The thesis can be roughly divided into three parts. The first part is based on the existing literature on responsibility (Chapter 2) and applications of Rawlsian procedural theory (Chapter 3). The second part forms the empirical body of the thesis. In this part, the engineers' rationales for distributing responsibilities are explored (Chapter 4) and analyzed in terms of a procedural framework (Chapter 5). The third part (Chapter 6) serves as an epilogue to the research: Given the

multidisciplinary nature of this research, what role could an engineering ethicist adopt when it comes to responsibility distributions in R&D networks?

In Chapter 2, I present an overview of different conceptions of responsibility in engineering. I explain how these different conceptions each have a different aim for ascribing responsibility, resulting in different criteria to judge the appropriateness of the ascription: fairness to potential wrongdoers, informed consent for imposed risks, and efficacy (prevention of risks and negative consequences). I argue that these different criteria for a distribution of responsibilities cannot be reduced to one overarching criterion; we have to take them all into account. This chapter concludes with the recommendation to apply a procedural criterion for judging the appropriateness of responsibility distributions; instead of focusing on single substantive fairness criteria it could be beneficial to focus on the distributing procedure and judge the resulting distribution in terms of procedural fairness.

Chapter 3 contains a discussion of a procedural approach to justification and fairness, based on Rawls' political theory. I explain the Rawlsian concepts of wide reflective equilibrium and overlapping consensus, which can be used to justify outcomes. According to the wide reflective equilibrium approach, a procedure can be justified as fair if it fits within one's individual set of background theories, moral principles and considered judgments. Working back and forth between the considered judgments, moral principles and background theories, every agent should strive for equilibrium between these three "layers of morality" by revising those elements that do not fit well with the other layers. If – after reflection – the three layers cohere, we speak of wide reflective equilibrium. Different people each have their own wide reflective equilibrium but there may be an overlap, for example, on the fair terms of cooperation. This shared module in each individual wide reflective equilibrium is by Rawls referred to as the overlapping consensus. The relation with responsibility distributions is that a particular distribution of responsibilities can be deemed fair if all actors involved can fit it in their own individual wide reflective equilibrium. In Chapter 3, I make an inventory of other applications of Rawls' concepts to actual problems and classify these according to a taxonomy of Rawlsian applications in order to gain insight how to apply Rawls' concepts in the context of responsibility distributions.

Chapter 4 forms the empirical counterpart of Chapter 2. In this chapter, I explore the opinions of technical researchers and engineers in the ALwEN project on social and moral issues related to technology. On the basis of a series

of interviews and a workshop with the researchers working on the ALwEN project, I investigate how actors in a research network distribute responsibilities for these issues. The empirical data confirmed the conceptual analysis in Chapter 2; there are different responsibility conceptions or rationales. It was found that there is a correlation between people's normative background theories and their responsibility conception, which confirms the *legitimacy* of the different conceptions. After all, the different normative background theories reflect particular visions on the good life and it is seen as the cornerstone of democracy that these different views should be recognized as legitimate. This means that they cannot be reduced to one overarching view. Similarly, because of the correlation between the normative background theories and the different responsibility conceptions, the different responsibility conceptions are also legitimate. We therefore need a neutral framework for deciding on responsibility issues that is impartial towards the different responsibility conceptions.

The descriptive analysis presented in Chapter 4 is elaborated in Chapter 5, where I apply Rawls' procedural approach to the empirical case. This chapter serves to further develop and demonstrate the use of Rawls' procedural approach. I operationalize Rawls' procedural approach in terms of two procedural norms based on literature on policy and innovation networks: reflective learning and inclusiveness. Analysis of the case shows that, in a pluralist setting, a procedural approach can be useful for encouraging discussion on the legitimacy of different responsibility conceptions and the question what a fair responsibility distribution amounts to. Although it is difficult to achieve a full overlapping consensus regarding all responsibility issues, the case indicates that a workshop structured along the lines of Rawls' wide reflective equilibrium approach can help focusing the technical work and make certain moral issues that were until then not recognized become part of the work. The ethical parallel research, and the workshop in particular, prompted sufficient discussion such that, in the end, most issues that were at first not addressed were later taken up by one of the partners in the research project. In that sense, the procedural approach did indeed alleviate the problem of many hands.

Chapter 6 serves as an epilogue to the discussions on responsibility in R&D. Given the discussions in the previous chapters, what role could engineering ethicists take upon themselves in distributing responsibilities? On the basis of a brief history of technology management, where the focus was traditionally put on the management of technology before or after the phase of development, I advocate the idea of technology management as "midstream modulation" or

"real-time Technology Assessment." This new approach is based on the idea that technology management or assessment should be integrated in the practice of R&D and become a research practice internal to the process of R&D itself. Rather than managing technology before R&D ("upstream"), when not enough can be known about its harmful social consequences to warrant controlling its development, or after the fact ("downstream"), when control has become costly and slow, technology should be managed or modulated during R&D itself ("midstream"). The results of these reflective studies can then feed back into the ongoing technological research. This requires technology managers and engineering ethicists to become more involved in the technological research rather than assessing it from an outsider perspective. Instead of focusing on the question whether or not to authorize, approve, or adopt a certain technology or on the question of who is to blame for potential mistakes, the guiding question in this new approach is how research is to be carried out. Engineering ethicists could help distributing responsibilities and ensure that no important responsibilities are overlooked. In their new role, engineering ethicists can facilitate a discussion of potential risks and other ethically relevant issues at a time that changes in the technological design can still be made. Subsequently, they can guide the discussion on how to assign and distribute the responsibilities for addressing these issues.

With this thesis, I hope to contribute to the discussion of responsibility in the context of R&D, and more in particular, to help alleviating the problem of many hands in R&D. I hope to have shown how a procedural can be used to resolve the tension between competing responsibility conceptions. Concerning the role of engineering ethicists, I argue for a more involved role in R&D. Based on my experience with ethical parallel research, I see a positive role for ethicists involved in a technological project. I think that the future for applied research is multidisciplinary in nature, involving both natural scientists, and social scientists and researchers with a background in the humanities. The combination of the empirical social sciences and the more conceptual humanities promises to provide a valuable surplus to (technological) research. An ethicist or social scientist can provide a different perspective, represent the "general public," and bring to the fore issues that are otherwise overlooked.

Samenvatting

Morele verantwoordelijkheid in onderzoeksnetwerken

Dit proefschrift handelt over morele verantwoordelijkheid in de context van Onderzoek en Ontwikkeling, doorgaans aangeduid met de Engelse term *Research and Development* (R&D). Morele verantwoordelijkheid is een van de kernbegrippen in de ethiek en het komt ook uitgebreid aan bod in zowel de algemene als de meer toegepaste filosofie- en ethiekliteratuur. In weinig artikelen wordt echter gesproken over morele verantwoordelijkheid in de context van R&D. Gezien de wijze waarop R&D is georganiseerd is dit een opmerkelijke omissie. Ten eerste wordt technologie vaak ontwikkeld in complexe netwerkorganisaties, waar verschillende actoren en instituten bij betrokken zijn. Deze netwerkorganisaties worden gekenmerkt door niet-hiërarchische relaties en een onduidelijke taakverdeling, wat de kans vergroot dat bepaalde zaken over het hoofd worden gezien. De mensen verwachten dat iemand anders bepaalde taken al heeft gekregen of anderszins op zich zal nemen. Ten tweede gaat technologieontwikkeling gepaard met een lange causale keten; deze loopt van fundamenteel onderzoek tot een testfase en dikwijls ook tot commerciële exploitatie. De negatieve consequenties (risico's en onvoorziene neveneffecten) komen vaak pas naar voren tijdens het gebruik van die nieuwe technologie, wat het moeilijk maakt om deze effecten van tevoren nauwkeurig te voorspellen. Het is daarom wenselijk om speciaal aandacht te besteden aan potentiële risico's en onzekerheden en verder te kijken dan de voor de hand liggende en bekende gevolgen van de ontwikkelde technologie. De complexe organisatie en de lange causale keten tezamen maken R&D erg vatbaar voor het zogenaamde "probleem van de vele handen". Dit probleem wordt doorgaans omschreven als de moeilijkheid om de verantwoordelijke persoon aan te wijzen wanneer er een groot aantal mensen betrokken is.

Het probleem van de vele handen is juist in R&D zo onwenselijk omdat de impact van technologie (inclusief de negatieve gevolgen) dikwijls groot is. Hierbij valt te denken aan het gebruik van asbest, DDT, kernafval en het broeikaseffect. Bovendien ontwikkelt technologie zich zeer snel, waardoor het vaak vooruit loopt op adequate wetgeving. Met andere woorden, terwijl de impact van technologie en het tempo van technologieontwikkeling toenemen, maakt de context waarin technologie ontwikkeld wordt het verdelen van verantwoordelijk-

heden juist onduidelijk. Deze ontwikkelingen geven aan dat het probleem van de vele handen een actueel en dringend probleem is in de context van R&D.

Heel algemeen geformuleerd hoop ik met dit onderzoek bij te dragen aan een oplossing voor het probleem van de vele handen in R&D. Het probleem van de vele handen kan op verschillende manieren geïnterpreteerd worden; in dit proefschrift vat ik het op als het probleem dat voortkomt uit een pluralisme van verantwoordelijkheidsopvattingen. Mijn specifieke onderzoeksdoel is een methode te ontwikkelen die kan helpen verantwoordelijkheidsvragen op te lossen in R&D in een situatie van pluralisme van verantwoordelijkheidsopvattingen. Dit pluralisme leidt tot verschillende, soms strijdige, criteria die gesteld kunnen worden aan een verantwoordelijkheidsverdeling. Met behulp van een procedurele benadering hoop ik de spanning tussen deze criteria te verminderen.

Het hoofdargument in dit proefschrift steunt aldus op drie centrale gedachten:

1. het probleem van de vele handen is gedeeltelijk een probleem dat teruggevoerd kan worden op verschillende opvattingen van verantwoordelijkheid. Mensen hebben verschillende opvattingen ten aanzien van de vraag wanneer het *eerlijk* is om iemand verantwoordelijk te maken (*eerlijkheidscriteria*);

2. deze verschillende opvattingen zijn allemaal gelegitimeerd (onder bepaalde voorwaarden) en kunnen niet gereduceerd worden tot één allesomvattende opvatting van verantwoordelijkheid (*pluralisme van verantwoordelijkheidsopvattingen*);

3. indien een procedure voor het verdelen van verantwoordelijkheden door alle betrokkenen wordt geaccepteerd als eerlijk, kan de uitkomst van deze procedure ook als eerlijk beschouwd worden (*procedurele eerlijkheid*).

De centrale onderzoeksvraag is hoe een procedurele benadering, gebaseerd op Rawls' idee van breed reflectief evenwicht (doorgaans aangeduid met de Engelse term *Wide Reflective Equilibrium* of de afkorting WRE), verder ontwikkeld en gebruikt kan worden om de spanning tussen verschillende verantwoordelijkheidsopvattingen te verminderen. Dit is gedeeltelijk een empirische vraag, inclusief kwalitatieve deelvragen: Hoe verdelen ingenieurs en technische onderzoekers de verantwoordelijkheden? Waarom beschouwen ze bepaalde

verantwoordelijkheidstoeschrijvingen als eerlijk en andere niet? Wat zijn de verschillende principes voor het verdelen van verantwoordelijkheden? Om dergelijke vragen te beantwoorden – of tenminste op verkennende wijze te onderzoeken – heb ik een ethisch parallel onderzoek uitgevoerd bij een technisch onderzoeksproject: het ALwEN project (waarbij de naam ALwEN een acroniem is voor *Ambient Living with Embedded Networks*). Het ALwEN project is gericht op de ontwikkeling van een bewakingssysteem voor thuisgebruik gebaseerd op *Ambient Intelligence* technologie. Dit ethisch parallel onderzoek stelde mij in staat om inzicht te verkrijgen in de verdelingsprincipes van ingenieurs en technische onderzoekers die betrokken zijn bij R&D.

Het proefschrift kan ruwweg verdeeld worden in drie delen. Het eerste deel is gebaseerd op de bestaande literatuur over verantwoordelijkheid (Hoofdstuk 2) en toepassingen van Rawlsiaanse procedurele theorie (Hoofdstuk 3). Het tweede deel vormt de empirische kern van het proefschrift. In dit deel worden de verdelingsprincipes van de ingenieurs verkend (Hoofdstuk 4) en geanalyseerd in termen van een procedureel raamwerk (Hoofdstuk 5). Het derde deel (Hoofdstuk 6) vormt een epiloog bij het onderzoek: Gegeven het multi-disciplinaire karakter van dit onderzoek, welke rol kan een ingenieursethicus op zich nemen bij het verdelen van verantwoordelijkheden in R&D netwerken?

In Hoofdstuk 2 presenteer ik een overzicht van verschillende perspectieven voor het toekennen van verantwoordelijkheid in de techniek. Ik leg hier uit hoe deze verschillende perspectieven ieder een bepaald doel beogen, resulterend in verschillende criteria om de adequaatheid van de verantwoordelijkheids-toeschrijving te beoordelen: eerlijkheid naar potentiële overtreders, geïnformeerde toestemming (*informed consent*) voor opgelegde risico's, en doeltreffendheid (voorkomen van risico's en negatieve consequenties). Ik betoog dat deze verschillende criteria voor een verantwoordelijkheidsverdeling niet gereduceerd kunnen worden tot één alomvattend criterium; alle criteria moeten op een of andere manier meegenomen worden. Dit hoofdstuk besluit met de aanbeveling om een procedureel criterium toe te passen voor het beoordelen van de adequaatheid van verantwoordelijkheidsverdelingen. In plaats van te kijken naar één eerlijkheidscriterium dat behoort bij één specifiek verantwoordelijk-heidsperspectief, kan het wellicht beter zijn om te kijken naar een procedureel criterium en de resulterende verdeling te beoordelen in termen van procedurele eerlijkheid.

Hoofdstuk 3 bevat een discussie van een procedurele benadering van rechtvaardigheid en eerlijkheid, gebaseerd op Rawls' politieke theorie. In dit

hoofdstuk ga ik nader in op de Rawlsiaanse concepten "breed reflectief evenwicht" en "overlappende consensus", die gebruikt kunnen worden om uitkomsten te rechtvaardigen. Volgens de benadering van het breed reflectief evenwicht kan een procedure als gerechtvaardigd of eerlijk worden beschouwd als het voor elk individu past binnen haar individuele verzameling achtergrondtheorieën, morele principes en weloverwogen opvattingen (*considered judgments*). Al heen en weer werkend tussen deze drie "lagen van de moraliteit" streeft iedere actor naar evenwicht tussen de verschillende lagen door díe elementen aan te passen die niet passen bij de andere lagen. Indien er – na reflectie – coherentie bestaat tussen de drie lagen spreken we van een breed reflectief evenwicht. Verschillende mensen hebben elk hun eigen breed reflectief evenwicht maar er kan wel sprake zijn van overlap, bijvoorbeeld over de eerlijke voorwaarden voor samenwerking. Deze gedeelde module binnen ieders breed reflectieve evenwicht wordt door Rawls de overlappende consensus genoemd. De relatie met verantwoordelijkheidsverdelingen is als volgt. Een bepaalde verantwoordelijkheidsverdeling kan als eerlijk worden beschouwd als alle betrokkenen deze verdeling kunnen inpassen in hun eigen individuele breed reflectieve evenwicht. In Hoofdstuk 3 maak ik een inventarisatie van verschillende actuele toepassingen van de concepten "breed reflectief evenwicht" en "overlappende consensus" en classificeer ik deze volgens een taxonomie van het gebruik van Rawlsiaanse concepten. Hiermee hoop ik inzicht te krijgen hoe Rawlsiaanse concepten toegepast kunnen worden in de context van verantwoordelijkheidsverdelingen.

Hoofdstuk 4 vormt de empirische tegenhanger van Hoofdstuk 2. In dit hoofdstuk verken ik de opvattingen van technische onderzoekers en ingenieurs in het ALwEN project over maatschappelijke en morele kwesties betreffende technologie. Op basis van een serie interviews en een workshop met de betrokkenen onderzoek ik hoe actoren in een onderzoeksnetwerk de verantwoordelijkheden voor deze kwesties verdelen. De verdelingsprincipes die in Hoofdstuk 2 onderscheiden waren, komen terug in de empirische gegevens. Er blijkt een correlatie te zijn tussen de morele achtergrondtheorieën die mensen aanhangen en hun verantwoordelijkheidsopvatting, wat de *legitimiteit* van de verschillende opvattingen bevestigt. De verschillende morele achtergrondtheorieën weerspiegelen immers een bepaalde visie op het goede leven en het erkennen van de legitimiteit van deze verschillende opvattingen wordt gezien als de hoeksteen van onze democratie. Dit betekent dat de verschillende opvattingen niet gereduceerd kunnen worden tot één alomvattende

opvatting van het goede leven. Evenzo houdt de correlatie tussen de morele achtergrondtheorieën en de verschillende opvattingen van verantwoordelijkheid in dat deze verantwoordelijkheidsopvattingen ook legitiem zijn. Er is daarom een neutraal kader nodig op basis waarvan beslissingen kunnen worden genomen over verantwoordelijkheidsvragen; dit kader moet onpartijdig zijn ten opzichte van de verschillende verantwoordelijkheidsopvattingen.

De beschrijvende analyse die ik in Hoofdstuk 4 presenteer werk ik verder uit in Hoofdstuk 5, waar ik Rawls' procedurele benadering toepas op de empirische casus. In dit hoofdstuk beoog ik Rawls' procedurele benadering verder te ontwikkelen en het gebruik ervan te demonstreren. Ik operationaliseer Rawls' procedurele benadering in termen van twee procedurele normen gebaseerd op de literatuur over beleids- en innovatienetwerken: reflectief leren en inclusie. Analyse van de casus toont aan dat, in een pluralistische omgeving, een procedurele benadering discussie kan stimuleren over de legitimiteit van verschillende verantwoordelijkheidsopvattingen en de vraag wat een eerlijke verantwoordelijkheidverdeling inhoudt.

Hoewel het moeilijk is om een volledige overlappende consensus te krijgen over alle verantwoordelijkheidsvragen laat de casus zien dat een workshop die georganiseerd is langs de lijnen van Rawls' idee van breed reflectief evenwicht richting kan aanbrengen in het technische werk en bepaalde morele kwesties, die tot dan toe niet meegenomen werden, onderdeel kan laten uitmaken van het werk. Het ethisch parallel onderzoek, en de workshop in het bijzonder, riep voldoende discussie op waardoor de meeste kwesties die in eerste instantie niet meegenomen werden uiteindelijk toch opgepakt werden door een van de partners in het project. In die zin leidde de procedurele benadering inderdaad tot een vermindering van het probleem van de vele handen.

Hoofdstuk 6 dient als een epiloog op de discussie over verantwoordelijkheid in R&D. Gegeven de discussies in de voorgaande hoofdstukken, welke rol kan een ingenieursethicus spelen in het verdelen van verantwoordelijkheden? Op basis van een beknopte geschiedenis van technologiemanagement, waar de focus traditioneel ligt op de fase voor of juist na de eigenlijke technologie-ontwikkeling, propageer ik het idee van technologiemanagement als *midstream modulation* of *real-time Technology Assessment (TA)*, dat wil zeggen, management gelijktijdig aan de fase van technologieontwikkeling. Deze nieuwe benadering is gebaseerd op het idee dat de beoordeling van technologie (TA) geïntegreerd zou moeten worden in de praktijk van R&D zelf. In plaats van sturing voorafgaand aan R&D ("bovenstrooms"), wanneer er nog onvoldoende

bekend is van de eventuele schadelijke sociale effecten, of na afloop ("beneden-stroomms"), wanneer de controle duur en traag is geworden, kan technologie beter beheerst of gestuurd worden tijdens R&D zelf ("middenstrooms"; in het Engels *midstream*). De resultaten van dergelijke reflecterende studies kunnen dan geïntegreerd worden in het lopende technische onderzoek. Dit vereist van technologiemanagers en ingenieursethici dat ze meer betrokken raken bij het technische onderzoek zelf in plaats van het te beoordelen als buitenstaander. Zij dienen zich daarbij dan niet zozeer te laten leiden door de vraag of een bepaalde technologie al dan niet geïmplementeerd, toegestaan of ingevoerd moet worden, of door de vraag wie er schuldig is voor bepaalde fouten. Bij deze nieuwe opzet zou de vraag *hoe* onderzoek uitgevoerd moet worden leidend moeten zijn. Ingenieursethici zouden kunnen helpen verantwoordelijkheden te verdelen en ervoor te zorgen dat belangrijke maatschappelijke en morele kwesties niet over het hoofd worden gezien. In deze nieuwe rol kunnen ingenieursethici een discussie over potentiële risico's en andere ethisch-relevante kwesties faciliteren op een moment dat veranderingen in het ontwerp nog steeds mogelijk zijn. Vervolgens kunnen ze de discussie leiden hoe de verantwoordelijkheden om deze kwesties aan te pakken verdeeld kunnen worden.

Met dit proefschrift hoop ik bij te dragen aan de discussie over verantwoordelijkheid in de context van R&D, en in het bijzonder aan de vermindering van het probleem van de vele handen. Ik hoop getoond te hebben hoe een procedurele benadering gebruikt kan worden om de spanning tussen concurrerende verantwoordelijkheidsopvattingen te verminderen. Wat de rol van ingenieursethici betreft pleit ik ervoor dat het werk van ethici meer wordt ingebed in R&D. Gebaseerd op mijn ethisch parallel onderzoek zie ik een positieve rol voor dergelijke "ingebedde" ethici. Ik denk dat de toekomst van toegepast onderzoek multidisciplinair van aard is, waarbij zowel natuur-wetenschappelijke onderzoekers, als onderzoekers met een achtergrond in de sociale of geesteswetenschappen betrokken zijn. De combinatie van de empirische sociale wetenschappen en de meer conceptuele geestes-wetenschappen kunnen het technische onderzoek een toegevoegde waarde geven. Een ethicus of sociaal wetenschapper kan een ander perspectief toevoegen, het algemene publiek vertegenwoordigen en kwesties voor het voetlicht brengen die anders over het hoofd worden gezien.

About the author

Neelke Doorn (1973) has a degree in civil engineering and philosophy, which she obtained at Delft University of Technology (MSc, 1997) and Leiden University (MA, 2005), both in the Netherlands. Before starting her PhD research at the Philosophy Department of Delft University of Technology in 2007, Neelke worked 10 years as a research engineer at Delft Hydraulics (now Deltares). Between 2006 and 2010, she also had a part-time research position at the Centre of Ethics of the Radboud University Nijmegen. Her work concentrates on moral issues in engineering, design, and the regulation of technology. During her PhD research, Neelke completed the PhD training program of the Netherlands Research School for Practical Philosophy (OZSE). She also started the trajectory to become a certified university teacher (BKO), which she is expected to finish in June 2011. Whilst doing her PhD research, Neelke spent three months as a visiting researcher at the Royal Institute of Technology in Stockholm (Sweden). She currently works as a postdoctoral researcher on the NanoNextNL project at Delft University of Technology. As of the academic year 2011/12, Neelke will be employed as assistant professor at Delft University of Technology and start a research project on the ethics of water governance.

Simon Stevin Series in Ethics of Technology
Delft University of Technology, Eindhoven University of Technology &
University of Twente
Editors: Philip Brey, Peter Kroes and Anthonie Meijers

Books and Dissertations

Volume 1: Lotte Asveld, *Respect for Autonomy and Technological Risks*, 2008

Volume 2: Mechteld-Hanna Derksen, *Engineering Flesh, Towards Professional Responsibility for 'Lived Bodies' in Tissue Engineering*, 2008

Volume 3: Govert Valkenburg, *Politics by All Means. An Enquiry into Technological Liberalism*, 2009

Volume 4: Noëmi Manders-Huits, *Designing for Moral Identity in Information Technology*, 2010

Volume 5: Behnam Taebi, *Nuclear Power and Justice between Generations. A Moral Analysis of Fuel Cycles*, 2010

Volume 6: Daan Schuurbiers, *Social Responsibility in Research Practice. Engaging Applied Scientists with the Socio-Ethical Context of their Work*, 2010

Volume 7: Neelke Doorn, *Moral Responsibility in R&D Networks. A Procedural Approach to Distributing Responsibilities*, 2011

Simon Stevin (1548-1620)

'Wonder en is gheen Wonder'

This series in the philosophy and ethics of technology is named after the Dutch / Flemish natural philosopher, scientist and engineer Simon Stevin. He was an extraordinarily versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbours, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview, which did not make him popular in religious circles. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically he was a pragmatic rationalist for whom every phenomenon, however mysterious, ultimately had a scientific explanation. Hence his dictum 'Wonder is no Wonder', which he used on the cover of several of his own books.