Practising Narrative Virtue Ethics of Technology in Research and Innovation

Wessel Reijers, MSc.

A dissertation submitted in fulfilment of the requirements for the award of

Doctor of Philosophy (Ph.D.)

To the

Dublin City University

School of Computing

Supervisors:

Dr. Bert Gordijn

&

Dr. Declan O'Sullivan

School of Computer Science and Statistics, Trinity College Dublin

January 2019

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

Wessel Reijers

(Candidate) ID No.: 15210725

Date: 1st of January 2019

Acknowledgements

To explain more, is to understand better (Ricoeur, 1983, p. x).

There are many whom I should explicitly thank for their support in bringing this dissertation to completion. First and foremost, I would like to thank Bert Gordijn for being the best supervisor one could wish for. He provided me with the freedom to pursue my own course and at the same time gave me incredibly useful and structured feedback, making sure that my constant diversions would not end up in chaos. He gave me the opportunity to work with amazing people in Dublin and supported me whenever I had crazy ideas to go abroad or write papers on topics that at first sight were unrelated to my work for this dissertation. I would also like to thank Declan O'Sullivan, my second supervisor, for providing great structural support and involving me in the wonderful work done in the ADAPT Centre, especially in its Ethics and Privacy Working Group. I am also very grateful for the enthusiastic and helpful feedback I gained from Renaat Verbruggen, who as a member of the supervisory panel has provided me with very valuable guidance in the writing process.

I would also like to thank all the people at the ADAPT Centre who made my PhD project such a wonderful experience. In particular, I want to thank David Lewis for his incredible liveliness, enthusiasm and hard work in the Ethics Canvas project, and also for being a great interlocutor in the pub, having good discussions with some pints together with Joss Moorkens and Marguerite Barry. Furthermore, it has been an absolute pleasure to work with Killian Levacher, Arturo Calvo, and Harshvardhan Pandit in the Ethics Canvas project, leading to what may be the first functioning "app" for practising ethics in R&I. Next to the people in the ADAPT Centre, I also want to thank my colleagues in the Institute of Ethics, and in particular Fiachra O'Brolchain for strolling through the DCU bookstore, grabbing a coffee after lunch and discussing the intricacies of virtue ethics.

In the research I did for my dissertation, I really have to say a big thank you to Mark Coeckelbergh and Marinus Ossewaarde. Mark has been an invaluable friend and fellow researcher who provided me with crucial assistance in developing the narrative technologies framework, which is a translation of the love we share for the work of Ricoeur. I am very grateful for the opportunity he gave me to have a research stay in Vienna, where I had the pleasure of working together with Janina Loh and Michael Funk. I have to deeply thank

Table of Contents

List of Papers

The papers displayed below have been accepted or published in peer-reviewed academic journals or conference proceedings, and have been integrated in the chapter(s) in this dissertation under which they are mentioned. Permission has been retrieved from co-authors to integrate the papers in this dissertation, and integration is in line with copyright policies of the relevant publishing companies.

Chapters 2&3

Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., & Gordijn, B. (2017). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. *Science and Engineering Ethics,* 24(5), 1437-1481.

Chapters 3&5

Reijers, W., & Coeckelbergh, M. (2016). The Blockchain as a Narrative Technology: Investigating the Social Ontology and Normative Configurations of Cryptocurrencies. *Philosophy & Technology*, 31(1), 103-130..

Chapter 5

Coeckelbergh, M., & Reijers, W. (2016). Narrative Technologies: a Philosophical Investigation of the Narrative Capacities of Technologies. *Human Studies*, 39(3), 325–346.

Chapter 6

Reijers, W. (2019). Beyond Postphenomenology: Ihde's Heidegger and the Problem of Authenticity. *Human Studies (Forthcoming).*

Chapter 7

Reijers, W., Gordijn, B., & O'Sullivan, D. (2016). Narrative Ethics of Personalisation Technologies. In D. Kreps, G. Fletcher, & M. Griffiths (Eds.), *IFIP Advances in Information and Communication Technology* (Vol. 474, pp. 130–140). Cham: Springer International Publishing.

Reijers, W., Koidl, K., Lewis, D., Harshvardhan, J. P., & Gordijn, B. (2018). Discussing Ethical Impacts in Research and Innovation: The Ethics Canvas. In D. Kreps, G. Fletcher, & M. Griffiths (Eds.), *IFIP Advances in Information and Communication Technology*. Cham: Springer International Publishing.

List of Tables

List of Figures

Abstract

Wessel Reijers. Practising Narrative Virtue Ethics of Technology in Research and Innovation

This dissertation develops a novel approach for practising ethics in research and innovation, called narrative virtue ethics of technology.

The increasing speed of technological developments in fields such as AI, robotics, biomedicine, and nanotechnologies, calls for proactive ethical reflection on the impacts of technologies. As a response, myriad methods for practising ethics in research and innovation have been developed. The first part of this research reviews existing methods and analyses how they deal with the anticipation of ethical issues of emerging technologies, the shaping of ethical design practices, and the evaluation of ethical impacts of existing technologies. This is followed by a critical discussion of the existing methods, outlining their shortcomings and offering recommendations for improvements.

Subsequently, a novel philosophical approach is developed that offers responses to the recommendations. This approach synthesises insights from work on virtue ethics by MacIntyre and Vallor with Ricoeur's philosophical hermeneutics. Ricoeur's narrative theory is used to construct an account of how technologies mediate people's experience and understanding of the social world, centring on the hermeneutic concepts of textuality, literacy, temporality, and distancing. Accordingly, virtue ethics and the approach of narrative technologies are integrated into an account of technical practice that is linked with Ricoeur's notion of the ethical aim as the good life, with and for others, in just institutions.

Finally, the philosophical approach is translated into a concrete method consisting of three phases that offer ways for analysing technical practices, evaluating them, and developing a broad set of prescriptions, including the use of an ethical oath, civic education and democratic decision-making in research and innovation practices. To demonstrate how this method could be practically used, a concrete tool is developed and evaluated in practice: the Ethics Canvas. This is a collaborative tool that enables researchers to discuss the ethical impacts of their work.

1 Introduction

1.1 Raison d'Être

When considering the grand theme of ethics, we often fix our gaze upon questions that concern the great problems our time. How can humankind address the great challenge of climate change, and is our generation morally indebted to future generations if we do not address it now, acknowledging its undeniable urgency (cf. Gardiner & Hartzell-Nicols, 2012)? How can we address the baffling challenges of human enhancement: is it ethically justifiable to enhance the capacities of our bodies and minds indefinitely, and should we be allowed to cheat on our own mortality (cf. Bostrom & Sandberg, 2008)? How can we deal with existential risks in a responsible way? Do we have an obligation to mitigate the dangers of an all-powerful artificial intelligence and should we treat the colonisation of other planets with a sense of urgency, acknowledging the risk of imminent destruction of the human species (cf. Yudowsky, 2008)?

Fortunately, most people do not have to deal with these all-encompassing and overly complex quandaries in their everyday existence. We go about caring for people and things that concern us in our proximate interactions: our family, friends, jobs, finances and sometimes our national or local elections. However, this does not mean that we are free from being confronted with moral difficulties. On the contrary, in our increasingly connected, global, technological society, moral problems – if only minute and seemingly insignificant – lie behind many of the practices we engage in. A banker, dealing with the approval or rejection of loan requests, might ask herself the question whether the algorithm providing her with suggestions could have a slight racial bias built into it because it surprisingly often suggests denying loans for a certain racial minority. A schoolchild might wonder whether she should react to a person on Twitter, not being sure whether that person is a "troll" or a genuine user. A doctor might wonder whether she should invest in hiring people or in buying care robots, to take care of her growing group of patients in need of daily care. A pilot navigating an unmanned aircraft to take out a battalion of enemy soldiers might wonder whether the effect of her tool of destruction is a new type of warfare that is utterly unfair.

The amount and the diversity of these moral problems that we are confronted with in our everyday existence are astounding. It seems therefore difficult to make sense of them in a

more general way. However, we can say that many of these quandaries have to do with our practices and more specifically with our *technical* practices. For instance, the question of hiring caregivers or care robots is linked to the technological innovations coming out of robotics applied to healthcare settings. Similarly, the question of conversations with "trolls" and genuine users on social media is linked to the way in which technological innovations enabling social interaction online condition or rather mediate online conversations. In other words, the place where technological innovation happens is the place where many moral concerns (first) arise in our contemporary world. It would be wise, therefore, to look more carefully at the way in which technological innovations are designed and implemented in society.

This dissertation answers to the increasing need for considering and addressing ethics of technological innovations. It does so by focusing on the context in which technological innovation happens, namely the context of research and innovation (R&I) practices. Examples of such practices are the development of robotic prostheses in universities by an international consortium of researchers, the development of machine translation processes based on neural networks in companies such as Google and the development of cryptocurrencies by small start-ups around the world. These processes of R&I, taking place in public and commercial institutions, contribute to a growing number of technological devices, systems and architectures that are integrated in societies around the globe. Contemporary paradigmatic developments in this context are captured under headings such as "the big data society" (Mantelero & Vaciago, 2015), "the Internet of Things" (French, 2016) and "the nanotechnology revolution" (Drexler & Peterson, 1991). Many of these developments intersect whenever we encounter them in our everyday lives, sedimented in artefacts, devices and systems. For instance, care robots in healthcare settings that provide input into processes of big data analytics are, by virtue of being connected to the Internet, part of the Internet of Things and depend on hardware that functions due to Nano-scale components. This intersection entails that processes of R&I become increasingly multi- and interdisciplinary and that innovations in one discipline can spill over to other disciplines, leading to shared opportunities but also to shared ethical challenges.

Thus, R&I processes seem to have multiple, significant impacts on the ways people live their daily lives and on the ways our societies are constructed and organised. At the same time, R&I processes become increasingly complex by being placed in the context of merging disciplinary fields and multiple modes of governance (public, private and mixed). Such considerations lead to a growing need for what is referred to as practises of "responsible

research and innovation" (Owen, Macnaghten, & Stilgoe, 2012). This notion entails an organised and reasoned reflection on R&I practices, in order to make sure that they are engaged with in a responsible way and that they aim at being beneficial for society while reducing the risks of harm they might bring about. As a whole, the field of responsible R&I can be said to focus on four dimensions: (1) the anticipation of future impacts of technology design, (2) reflection on the human values that are implied in technology design, (3) inclusion of stakeholders in the design process and (4) responsiveness of technology design to societal changes (Stilgoe, Owen, & Macnaghten, 2013). These dimensions show that understanding the role of technologies is crucial in reflecting on the societal impacts of R&I processes.

When considering the "future impacts" of technology on society we invariably discuss impacts that are normatively significant, such as the impact surveillance technologies have on the *privacy* of citizens or the impact of novel financial instruments have on the *just* distribution of means. The extent to which technology design can be said to embed certain human principles and respond to societal changes invariably leads to ethical questions such as: what human principles, values and virtues should be embedded in design? How can we deal with conflicts between principles, for instance between the security and privacy in designing social networks applications? What kinds of responses to societal changes are desirable and which are undesirable? Finally, the inclusion of stakeholders in the process of R&I also responds to an ethical requirement of justification. Considering that ethical issues are deliberated upon in public discourse, ensuring participation in this discourse is an important requirement for "practising" ethics in concrete settings. In other words, when debates on ethics leave the walls of the philosophy department and enter the realm of public discourse, justified stakeholder participation becomes an ethical requirement of any effort towards responsible R&I.

To respond to the abovementioned challenges, this dissertation develops a novel approach to practising ethics in R&I. It approaches the thematic of ethics in R&I from the distinct perspective of the field of ethics of technology. Since the "empirical turn" (Achterhuis, 2014), this field has become increasingly interested in the analysis and evaluation of concrete technologies and in processes of technological development. It has thereby come closer to pre-existing considerations in discourses on the ethical, legal and social impacts of technological innovation (ELSA) and RRI, which have originally been more focused on the aspect of governance of R&I rather than on the analysis and evaluation of the ethical impacts of R&I. On the one hand, this dissertation tries to bring the fields of RRI and ethics of technology closer together by introducing consideration of normative

ethics and philosophy of technology in the discourse on the organisation of R&I processes. On the other hand, however, its scope will remain tied to the field of ethics of technology, which implies that it will only partially engage with the broad range of considerations present in the RRI discourse.

As will become apparent, a different way of practising ethics is needed because of shortcomings in existing methods, ranging from difficulties of anticipating ethical impacts of emerging technologies, to problems regarding the embedding of values in technology design, to inadequate concrete ways to have stakeholders participate in the process. For the most part, the research in this dissertation will be of a rigorous philosophical character, targeting the conceptual grounding of methods for practising ethics in R&I by drawing from the traditions of virtue ethics and philosophical hermeneutics. Additionally, parts of this dissertation will have a descriptive and analytic character, notably the ones concerned with reviewing and analysing existing methods. Moreover, it will also engage in more pragmatic research because the purpose of the philosophical endeavour lies in its applicability to concrete R&I practices. This research will mostly be concerned with the formulation of practical procedural steps and one corresponding tool that R&I practitioners can use in their everyday activities.

1.2   Aims and Methodology

This dissertation focuses on the ethical impacts of R&I processes and particularly on how ethical reflection can be made an integral part of practices that constitute these processes. As such, we place our investigation under the heading of applied ethics, which is characterised by Singer as "practical ethics" or ethics dealing with "practical issues" (Singer, 1986). Beauchamp proposes another, broad definition of applied ethics, as "any use of philosophical methods to treat moral problems, practices, and policies in the professions, technology, government and the like" (Beauchamp, 2003, p. 3). In line with these characterisations, our general aim will be to use philosophical methods to address certain ethical problems pertaining to R&I practices. This general aim can be made more specific by dividing it up into three particular aims. First, we aim to get to know what methods exist to facilitate practising ethics in R&I practices. Second, we aim to critically review these methods in order to identify deficiencies in need of improvement. Third, we accordingly aim at formulating a philosophical approach that could underlie a novel method that could be applied to concrete R&I practices.

In line with these aims, the research presented in this dissertation focuses on answering three interrelated research questions:

I. What methods have been developed for practising ethics in research and innovation?
II. What are the main shortcomings of these methods and what recommendations for a novel method follow from these shortcomings?
III. To what extent can insights from virtue ethics and philosophical hermeneutics be synthesised to construct the theoretical foundation for a new method for practising ethics in research and innovation that incorporates the recommendations advanced?

We will address each research question from a distinct methodological angle. The *first* research question will be tackled by a literature review of methods for practising ethics in R&I. The approach that will be used for conducting the literature review is based on the "systematic" review method (Cronin, Ryan, & Coughlan, 2008), which implies that it will (1) be based on a research question, involve (2) inclusion and exclusion criteria, (3) a selection and accessing of literature (4) an assessment of the quality of the literature and (5) a systematic analysis and synthesis of the findings of the review. We will apply inclusion and exclusion criteria through different selection rounds: (1) a search of literature in three major databases for academic literature according to specified queries, (2) a selection of that literature based on titles and abstracts, (3) and a selection of additional literature by means of the "backwards snowballing" procedure (Wohlin, 2014), using criteria for the titles of sources referred to by the sources selected in the second round. We will subsequently read the selected literature, analyse the methods it contains and systematically categorised those.

Our *second* research question will be dealt with by means of a structured approach to criticise established methods for practising ethics in R&I, and to use to the criticism for the formulation of recommendations for improvements. The criticism will be structured in a two-fold manner: (1) we will consider criticisms that are raised in the literature we reviewed and (2) we will incorporate more general concerns raised in auxiliary applied ethics literature. Each particular criticism will lead to a set of short, concrete recommendations that can inform efforts for improvements. To explain how the recommendation will inform our work towards a novel method, a framework is presented based on three choices: for a theory in ethics, a theory of technology, and an approach to stakeholder participation. Taken together, these

steps should create a transparent overview of the connection between our literature review, the related academic debates and the criticism and recommendations we formulate. Thereby, it should provide for the rationale leading up to our third research question.

Our *third* research question will be addressed by engaging with both virtue ethics and with philosophical hermeneutics, as well as with methods for participatory design. Our approach in ethics will be placed within the tradition of virtue ethics, and more specifically in the novel specialisation in this tradition called virtue ethics of technology (cf. Vallor, 2016). We will predominantly focus on how virtue ethics approaches address the cultivation of the virtues through practice. The understanding of practice will therefore be crucial, and we will initially follow MacIntyre's (2007) model of practice, as being made intelligible through narratives that relate it to the narrative unity of a human life and a moral community. To complement our virtue ethics approach, we will use narrative theory to explicate what technical practices are and how they are mediated by technologies. Narrative theory is placed in the tradition of philosophical hermeneutics, which has Ricoeur as one of its main proponents. Departing from Ricoeur's work, we characterise hermeneutics initially as a *system of interpretation*. Kakkori explains that hermeneutics "is the theory of the rules that govern an exegesis, an interpretation of a particular text. The text is understood very widely" (Kakkori 2015, p. 25). Thus, hermeneutics aims at explicating a method for the interpretation of text, but text taken widely, as a model for human existence in general - as Ricoeur for instance explains when he claims that we can model the phenomenon of meaningful action by considering it along the lines of the model of a text (Ricoeur 1971)[1].

We examine how our approach can be applied to concrete R&I practices by drawing from methods outside of philosophy that are concerned with facilitating and supporting practical collaborative efforts, notably methods in the field of business development. We will draw from such methods to construct a collaborative tool that can be used by R&I practitioners in their day-to-day work. Furthermore, we will conduct an empirical study of the perceived usefulness of the tool we will design, which will be based on the use of a questionnaire. The questionnaire results will be used as the basis for our evaluation of the tool that will be designed. For this part of our work, we will adhere to a pragmatic method that is constrained by practical concerns (e.g. the institutional setting in which the work is done) rather than by theoretical or formal concerns. Even though we apply a certain procedural sequence to the construction of a collaborative tool (see section 7.2), the work on

---

[1] Taken together, we characterise philosophical hermeneutics as the theory of method for interpretation of

the tool will be done iteratively and in parallel with the development of the literature review and our theoretical framework. The conceptualisation and design of the collaborative tool will be organised within the context of R&I practices in the ADAPT Centre[2], and will take place in a multidisciplinary team setting.

1.3 Terms and Definitions

To make the reader acquainted with the central terms in this dissertation, we present these terms, contextualise them briefly and formulate working definitions.

Ethics

Ethics (or more specifically, *Western* Ethics in our investigation) as derived from the Greek word *ethos*, which means habit, custom or habitual character, is a branch of philosophy that concerns itself with questions about moral issues; for instance, what is a right action, what counts as a virtuous character and what has moral value (Copp, 2005, p. 4)? Normative ethics aims at formulating direct answers to those questions, by constructing ethical theories of which the most prevalent are consequentialism, Kantian, deontological ethics and virtue ethics. Meta-ethics is concerned with the meaning of moral concepts and the justification of moral propositions. Applied ethics is normative ethics as applied to specific contexts and real life domains. In the course of our investigation, we will mostly argue within the discipline of applied ethics.

Method

With method, originating from the Greek word methodos, which means a way of doing something (Partridge, 2006, p. 2191), we more specifically mean a "detailed, content-specific […] problem solving procedure" (Nickles, 1987, p. 104 - emphasis added) that aims at structuring processes of practising ethics in R&I. A method is therefore not about the "what" of practising ethics (e.g. what values should one adhere to?), but about the "how" of practising ethics (e.g. how should one come to an identification of ethical issues?). For instance, a method can be a description of procedural steps an ethicist can take in order to evaluate a technological artefact, such as "gather the relevant facts", "describe possible consequences of the use of technology x, for different stakeholders", "evaluate these

---

[2] The ADAPT Centre is an Irish R&I centre that focuses on digital content technologies and their development throughout the R&I pipeline. This dissertation has been developed in the ADAPT Centre and has therefore used it as the context for constructing the collaborative tool for practising ethics in R&I.

consequences", and so forth. In this dissertation, an important difference is maintained between the concepts of method and of methodology.

Methodology

In contrast with a method, a methodology is not to be understood as a coherent set of procedural steps but as the *logos,* or the discourse supporting and justifying the method. As such, methodology is concerned with the philosophical assumptions and arguments that are constructed to bring forth and support the method that is aimed at. At certain points in this dissertation, depending on contextual suitability, we use the terms "philosophical approach" and methodology interchangeably. In line with Carr, we define methodology as "the theoretical rationale or"…"'principles' that justify the research methods appropriate to a field of study" (Carr, 2006, p. 422). The field of study, in this dissertation, is demarcated by the scope of "practising ethics in R&I".

Narrative

Bruner characterises "narrative" as "a conventional form, transmitted culturally and constrained by each individual's level of mastery and by his conglomerate of prosthetic devices, colleagues, and mentors" (Bruner, 1991, p. 5). For our purposes, in line with Ricoeur (1983), narrative is defined by its mediating role – between a human being and her experiences and understanding of the world. Plainly said, a narrative like a story in a book is defined by the way in which it mediates the experience and understanding of the reader's social world. This mediating role can be captured by the concept of *emplotment*, which designates the capacity of a narrative to organise heterogeneous elements (characters and events) in a meaningful synthesis that we commonly designate as the plot. As such, we define narrative as a mediation of human experience and understanding of the social world through the process of emplotment, which is the organisation of heterogeneous elements in a meaningful synthesis. The paradigmatic model for a narrative (its empirical correlate) is the model of the text, which is used as the basis for understanding how narratives mediate our experience and understanding of the world. However, narrative is not exclusively bounded to actual texts because it also mediates utterances (e.g. telling a story), performances (e.g. a tragedy, a comedy) and – as we will argue – technical practices (e.g. using a computer interface, driving a car).

Research and innovation (R&I)

Different characterisations of the process "R&I" are proposed in the literature (see, e.g. Hauser et al., 2006; Kajikawa et al., 2008; Roberts, 2007). Commonly, it is conceptualised as a set of practices (1) based on conceptual (scientific) knowledge that (2) translates a certain conceptual idea into a technological design that is tested, prototyped and translated into an actual application, which (3) can subsequently be put into production and introduced to society at large. This notion of R&I mostly corresponds to the "linear model of innovation" (Godin, 2014), which is dominant but far from the only model of R&I. In relation to ethics, it is argued that most conceptualisations of R&I are insufficient because of their focus on technological, economically motivated innovation rather than on innovation for the common good, and because of the underlying assumption in the literature of innovation being inherently good (Blok & Lemmens, 2015). To overcome these issues, it is suggested that the concept of innovation should include non-technological novelties and alternative strategies of innovation that offer radically different types of stakeholder relations and responsibilities, such as peer-to-peer production. Even though such considerations lie outside of the scope of this dissertation, we propose a broad definition of R&I that could encompass them. Hence, we characterise an R&I as a process consisting of a set of practices aimed at (1) the invention and design of new material-cultural phenomena through (2) the use of techno-scientific knowledge and (3) the introduction and diffusion of these phenomena in society.

Technological mediation

Unfortunately, "technological mediation" is not accurately defined in the literature that inaugurated and developed it as a concept, in which it is loosely conceptualised as technologies establishing specific *relationships* between humans and their reality (Tripathi, 2010, p. 5), as human existence being always *influenced* by technologies (Dorrestijn, 2011, p. 226) and technologies actively *co-shaping* "people's being in the world" (Verbeek, 2005, p. 356). Etymologically, the term mediation originates from the Latin term "medialis", which means being "situated in the middle" (Partridge 2006, p. 1942). As such, it has a double signification. The more fundamental one, which Heidegger designates as "being-in" (Heidegger, 1996, p. 51) signifies the human experience of the world as being surrounded by this world. In other words, figuratively speaking, human beings are surrounded by their technological world as fish are surrounded by their water-world (our world does not merely shape but also conditions what it means to be human). The second signification implies that technological devices, systems and architectures are "in between" our experiences and the

world, in the way that a pair of glasses can be between the world and the observer. As a working definition, our use of the verb "to mediate" should be understood along the lines of two significations proposed by the Oxford Dictionary, namely: "to bring about" in general, which can be a "means of conveying" (e.g. "glasses mediate the perception of the world of their bearer") and "forming a link between" two phenomena (e.g. "the introduction of credit cards mediated the practice of payment") (Oxford Living Dictionaries, 2017).

1.4 Expected Contribution

The reader can expect to find three significant contributions to the academic literature in this dissertation. The first contribution is one of overview and critique. Thus far, no attempt has been made to comprehensively map and analyse the great variety of methods for practising ethics in R&I that have been developed in recent years. By taking on this task, this dissertation aims to provide a much-needed general overview of these methods, complemented by a general critique that covers most of their central aspects. Separate critiques of these methods have already been developed, such as ones focusing on the weaknesses of value sensitive design (cf. Manders-huits, 2011) or ones focusing on the problem of speculation in future-oriented ethics assessment (cf. Nordmann, 2007). However, this dissertation aims to incorporate these separate critiques into a general critique of the field, pointing at particular ways in which methods can be improved while taking the entire R&I process into account, from the early stages of conceptual work to the stage of the introduction and distribution of technologies across society.

The second contribution is a theoretical one, impacting two traditions in the fields of ethics and philosophy of technology. First, our theoretical work is expected to contribute to the nascent theory of virtue ethics of technology (cf. Vallor, 2016). It does so by showing how virtue ethics can benefit from Ricoeur's theory of practice (Ricoeur, 1992). Also, it does so by drawing the problematic of technological mediation into the consideration of practices in virtue ethics, which has not been done before. Second, our theoretical work is expected to contribute to theoretical approaches in philosophy of technology, in particular those concerned with technological mediation (e.g. Ihde, 1990; Verbeek, 2006). It does so by offering a new theory that responds to a number of contemporary criticisms of theories of technological mediation, focusing on the appropriation of Ricoeur's (1983) narrative theory and model of emplotment. By doing so, our dissertation will develop a new hermeneutic approach of technological mediation that can contribute to existing approaches dealing with this thematic such as postphenomenology.

The third contribution is a practical one, which is to some extent experimental as well. That is, it can be understood as an instance of "experimental ethics" (cf. Verbeek, 2014), considered as the use of empirical methods to accompany the use of ethics in R&I practices. So far, the field of applied ethics has done fairly little to produce tools to proactively practise ethics in R&I. Most practically useful outcomes of applied ethics take the form of codes of conduct or checklists for R&I practitioners. This dissertation aims to move beyond this "passive" or top-down application of ethics, to a more "active" or bottom-up application that involves the construction of participatory tools that R&I practitioners can work with in their day-to-day activities. As such, this dissertation will contain one of the first attempts to use an approach in applied ethics to design a practical tool that can be used in R&I settings. This attempt will be experimental in nature, because it will draw from a field that is removed from applied ethics, namely the field of business development, and because it will evaluate the perceived usefulness of the tool by means of survey amongst R&I practitioners who will use the tool in their projects. This evaluation will be of an explorative nature, and will be used to tentatively indicate whether the tool could be a valuable way to practice ethics in R&I.

1.5 Structure

The search for a novel method for practising ethics has a three-fold structure that can be quickly summarised as follows: *first*, we investigate what methods already exist; *second*, we investigate what could be done better and what recommendations we could formulate to construct a novel method; *third*, we investigate what philosophical approach could be developed to incorporate the recommendations and how it could be applied as a method to concrete R&I practices. This three-fold structure translates into eight different chapters, of which this introduction is the **first**.

The second chapter presents the systematic literature review and provides an analysis of the established methods for practising ethics in R&I. The literature review shows that a great number of methods for practising ethics in R&I has been developed in the course of the last decades, amongst which are listed methods such as "value sensitive design" (cf. Friedman & Kahn, 2002), "ethical impact assessment" (cf. Wright, 2014) and "anticipatory ethics" (cf. Brey, 2012). To elucidate the origins and backgrounds of the methods, an overview is provided of the disciplinary fields to which they belong, such as the fields of health technologies and computer science. Subsequently, the different methods are analysed by considering a typology based on their position vis-à-vis the R&I process, which is expressed by categorising methods as "ex ante" (focusing on the start of the R&I process),

"intra" (focusing on the design phase) and "ex post" (focusing on technologies at the point of introduction in society).

The third chapter provides a criticism of the existing methods and formulates recommendations to improve them. The criticism revolves around an evaluation of the extent to which the existing methods address three main aims and one common challenge of practising ethics in R&I: dealing with uncertainty of technological change, achieving ethical technology design, identifying, analysing and resolving ethical impacts, and organising appropriate participation of stakeholders. For each of these core aims and the challenge, we discuss what shortcomings in the established methods prevent them from being adequately addressed. Subsequently, recommendations are formulated for solving the shortcomings we identify. For instance, to address the lack of participation of researchers and designers in the process of practicing ethics, it is recommended that future efforts should be aimed at including these stakeholders in the process. Subsequently, a framework for a novel approach is constructed, which makes explicit what theory in ethics will be used, what theory of technology could be used, and what means for stakeholder participation will be considered.

The fourth chapter presents an exploration of virtue ethics as the basis for our methodology. It provides a brief overview of the general historical development of virtue ethics, as well as the contributions by its central contemporary scholars such as Anscombe, Foot and MacIntyre. This overview culminates in a discussion of Vallor's recent work that establishes virtue ethics as an approach in ethics of technology and offers a heuristic of virtues that are particularly responsive to current times in which technological innovations ubiquitously impact our everyday practices. Subsequently, a transition is made to a methodological discussion, in which the work of MacIntyre occupies a central position. MacIntyre's conceptual scheme is presented, which captures the process of the cultivation of virtues, revolving around his notions of practice, narrative unity of a human life and moral community. Subsequently, some significant limits and shortcoming of Vallor's and MacIntyre's approaches are discussed, which leads to an argument for complementing the virtue ethics of technology approach with an account of technological mediation.

The fifth chapter provides a hermeneutic account of technological mediation that is based on the narrative theory of Ricoeur. Initially, it discusses why narrative theory can be used to understand technological mediation, focusing on the way it encompasses considerations of language and of intersubjectivity in technical practice. Subsequently, it presents and discusses Ricoeur's conceptual framework that revolves around the methodological triad of prefigured, configured and refigured narrative time. Focusing on the

moment of configured time, Ricoeur's framework is appropriated to provide an account of technological mediation that makes explicit how technologies can bring about a process of emplotment, understood as the organisation of characters and events in a meaningful synthesis. The resulting framework of "narrative technologies" provides a systematic understanding of technological mediation according to four central concepts of textuality, literacy, temporality and distancing.

The sixth chapter deals with the integration of the framework of narrative technologies with the virtue ethics approach, which revolves around the notion of technical practice that cultivates the virtues. To start, it presents an argument to link the notion of technical practice with virtue and narrative, which involves the work of Heidegger, Arendt, and Ricoeur. This argument explains what it means for a technical practice to cultivate the virtues, and why narrative is necessary for this. Accordingly, an account of technical practices is presented, which integrates Ricoeur's theory of practice with the framework of narrative technologies. This integration conceptualises the move from basic actions to practices, to life plans and narrative unity of a human life. The third section completes the transition to an understanding of technical practice that cultivates the virtues, by considering the ethical aim of the good life, with and for others, in just institutions. For each aspect of the ethical aim, suggestions are made regarding the ways in which technical practices could be cultivated to accord with the virtues.

The seventh chapter presents the procedural steps that make up our novel method, as well as a practical tool, the "Ethics Canvas", that enables us to apply one procedural step to concrete R&I practices. To illustrate how each procedural step relates to R&I practices, a particular type of technology, namely personalisation technology, is presented as well as a concrete application of this technology to smart border controls. Subsequently, nine stages are formulated that together constitute our novel method, each containing distinct procedural steps. These relate to three overarching phases, of (1) gathering and interpretation narratives (2) interpreting and evaluating technical practices, and (3) evaluating these practices according to the ethical aim, leading to different forms of prescription. Finally, a transition is made from one of the procedural steps to a concrete tool, the Ethics Canvas, which can be used to apply the method to R&I practices. The background of this tool in business development literature is explained, leading to an elucidation of the design process and the evaluation of the perceived usefulness in actual R&I practices.

The eight chapter formulates a comprehensive conclusion, starting with a summary of the work done which revolves around a description of the main findings in the dissertation

that answer to the three research questions. It presents the main achievements of this dissertation and some of its main limitations. Based on the limitations, some suggestions for future research are presented.

1.6 References

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

Beauchamp, T. L. (2003). The Nature of Applied Ethics. In R. . Frey & C. H. Wellman (Eds.), *A Companion to Applied Ethics*. Malden: Blackwell Publishing.

Blok, V. & Lemmens, P. (2015) The Emerging Concept of Responsible Innovation. Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation. In Koops, B. J., Oosterlaken, I., Romijn, H., Swierstra, T., and van den Hoven, J. (eds) *Responsible Innovation 2: Concepts, Approaches, and Applications. Heidelberg*: Springer, pp. 1–303. doi: 10.1007/978-3-319-17308-5.

Borning, A., Friedman, B., & Kahn, P. H. (2004). Designing for Human Values in an Urban Simulation System : Value Sensitive Design and Participatory Design. In *PDC-04 Proceedings of the Participatory Design Conference, Vol 2, Toronto, Canada* (pp. 68–71). Palo Alto.

Bostrom, N., & Sandberg, A. (2008). Ch.18 The Wisdom of Nature: An Evolutionary Heuristic for Human Enhancement. In Savulescu, J. & Bostrom, N. (Eds.), *Human Enhancement*. Oxford: Oxford University Press. (pp. 375–416).

Brey, P. (2000). Disclosive Computer Ethics: The Exposure and Evaluation of Embedded Normativity in Computer Technology. *Computers and Society*, *30*(4), 10–16.

Brey, P. (2012). Anticipatory Ethics for Emerging Technologies. *NanoEthics*, *6*(1), 1–13. http://doi.org/10.1007/s11569-012-0141-7

Bruner, J. (1991). The Narrative Construction of Reality. *Critical Inquiry*, *18*(1), 1. http://doi.org/10.1086/448619

Carr, W. (2006). Philosophy, Methodology and Action Research. *Journal of Philosophy of Education*, *40*(4), 421–435. http://doi.org/10.1111/j.1467-9752.2006.00517.x

Copp, D. (2005). *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press. http://doi.org/10.1093/0195147790.001.0001

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a Literature Review: a step-by-step approach. *British Journal of Nursing*, *17*(1), 38–43.

Dorrestijn, S. (2011). Technical Mediation and Subjectivation: Tracing and Extending Foucault's Philosophy of Technology. *Philosophy & Technology*, *25*(2), 221–241. http://doi.org/10.1007/s13347-011-0057-0

Drexler, E., & Peterson, C. (1991). The Nanotechnology Revolution. *The New Atlantis: A Journal of Technology & Society*, 17–34.

French, A. M. (2016). The Digital Revolution: Internet of Things, 5G, and Beyond. *Communications of the Association for Information Systems*, *38*(40), 840–850.

Friedman, B., Borning, A., Davis, J. L., Gill, B. T., Kahn, P. H., Kriplean, T., & Lin, P. (2008). Laying the Foundations for Public Participation and Value Advocacy : Interaction Design for a Large Scale Urban

Simulation. *Design*, *1*(2), 305–314. http://doi.org/10.3233/IP-2009-0172

Friedman, B., & Kahn, P. (2002). Value Sensitive Design: Theory and Methods. *University of Washington Technical*, (December), 1–8. http://doi.org/10.1016/j.neuropharm.2007.08.009

Gardiner, S. M., & Hartzell-Nicols, L. (2012). Ethics and Global Climate Change. *Nature Education Knowledge*, *3*(10), 555–600.

Godin, B. (2014) Invention, Diffusion and Linear Models of Innovation: the contribution of anthropology to a conceptual framework. *Journal of Innovation Economics*, *15*(3), 11-37. doi: 10.3917/jie.015.0011.

Hauser, J., Tellis, G. J., & Griffin, A. (2006). Research on Innovation: A Review and Agenda for Marketing Science. *Marketing Science*, *25*(6), 687–717. http://doi.org/10.1287/mksc.1050.0144

Heidegger, M. (1996). *Being and Time*. (J. Stambaugh, Ed.). New York: State University of New York.

Ihde, D. (1990). *Technology and the Lifeworld*. Bloomington: Indiana University Press.

Kajikawa, Y., Usui, O., Hakata, K., Yasunaga, Y., & Matsushima, K. (2008). Structure of knowledge in the Science and Technology Roadmaps. *Technological Forecasting and Social Change*, *75*(1), 1–11. http://doi.org/10.1016/j.techfore.2007.02.011

Manders-Huits, N. (2011). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*, *17*(2), 271–287. http://doi.org/10.1007/s11948-010-9198-2

MacIntyre, A. (2007). *After Virtue: A Study in Moral Theory* (Third Edit). Notre Dame, Indiana: University of Notre Dame Press. http://doi.org/10.1017/CBO9781107415324.004

Mantelero, A., & Vaciago, G. (2015). Data Protection in a Big Data Society. Ideas for a future regulation. *Digital Investigation*, *15*, 104–109. http://doi.org/10.1016/j.diin.2015.09.006

Nickles, T. (1987). Methodology, Heuristics, and Rationality. In J. C. Pitt & M. Pera (Eds.), *Rational Changes in Science*. Dordrecht: D. Reidel Publishing Company.

Nordmann, A. (2007). If and Then: A Critique of Speculative Nanoethics. *NanoEthics*, *1*(1), 31–46. http://doi.org/10.1007/s11569-007-0007-6

Osterwalder, A., & Pigneur, Y. (2004). An Ontology for E-business Models. In Curry, W. (Ed), *Value Creation from E-Business Models*. Oxford: Butterworth-Heinemann. (pp. 65–97). http://doi.org/10.1016/B978-075066140-9/50006-0

Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: a handbook for visionaries, game changers, and challengers*. No publisher. Retrieved from: BusinessModelgeneration.com.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible Research and Innovation: From science in society to science for society, with society. *Science and Public Policy*, *39*(6), 751–760. http://doi.org/10.1093/scipol/scs093

Oxford Living Dictionaries. (2017). Mediate. English. Retrieved October 12, 2017, from https://en.oxforddictionaries.com/definition/mediate

Partridge, E. (2006). *Origins: A Short Etymological Dictionary of Modern English*. London: Routledge.

Ricoeur, P. (1983). *Time and Narrative - volume 1*. (K. McLaughlin & D. Pellauer, Eds.) (Vol. 91). Chicago: The University of Chicago. http://doi.org/10.2307/1864383

Ricoeur, P. (1992). *Oneself as Another*. (K. Blamey, Ed.). Chicago: University of Chicago Press.

Roberts, E. B. (2007). Managing Invention and Innovation. *Research Technology Management*, *50*(1), 35–54. http://doi.org/10.1038/427679a

Singer, P. (1986). *Applied Ethics*. Oxford: Oxford University Press. http://doi.org/10.1017/CBO9781107415324.004

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a Framework for Responsible Innovation. *Research Policy*, *42*(9), 1568–1580. http://doi.org/http://dx.doi.org/10.1016/j.respol.2013.05.008

Tripathi, A. K. (2010). Ethics and Aesthetics of Technologies. *AI and Society*, *25*(1), 5–9. http://doi.org/10.1007/s00146-010-0265-7

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.

Verbeek, P.-P. (2005). *What Things Do; philosophical reflections on technology, agency, and design*. Pennsylvania: Pennsylvania University Press.

Verbeek, P.-P. (2014). Technology Design as Experimental Ethics. In S. van der Burg & T. Swierstra (Eds.), *Ethics on the Laboratory Floor* (pp. 79–80). Heidelberg: Springer. http://doi.org/10.1057/9781137409805

Wohlin, C. (2014). Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014)*, 1–10. http://doi.org/10.1145/2601248.2601268

Wright, D. (2014). Ethical Impact Assessment. *Ethics, Science, Technology, and Engineering*, *163*(c), 163–167. http://doi.org/10.1016/j.clsr.2011.11.007.Wright

Yudowsky, E. (2008). Artificial Intelligence as a Positive and Negative Factor in Global Risk. *Global Catastrophic Risks*, *1*(303), 1–42.

2   A Literature Review and Analysis[3]

2.1   Introduction

This chapter provides a review and analysis of literature that deals with methods for practising ethics in research and innovation (R&I). Although a lot has been written on the topic of practising ethics in R&I, no attempt has yet been made to present a comprehensive overview of the myriad of methods that have been proposed to structure and develop it. The notion of *practising* ethics is used to capture the practices of "doing" ethics in R&I under consideration in their broadest possible sense. However, notions of *including, integrating* or *incorporating* ethics in R&I are occasionally referred to as well, when such notions are used in the particular literature that is included in this review.

Dealing with ethics in R&I is increasingly urgent because of the transformative potential and complexity of contemporary advancements in science and technology (Brey, 2012, p. 2). Ethical issues in the context of R&I are often multifaceted and pervasive, because they result from complex changes in people's behaviours, socio-economic relations, power relations between people and institutions, and changes in the environment. For instance, technological innovations grouped under the heading of "ubiquitous computing", which relate to the growing presence of computational devices connected to the Internet in many aspects of everyday life, are said to be capable of causing harm to inter-human relations and create severe social shifts in economic and political power (Bohn, Coroama, Langheinrich and Mattern, 2005). Another example, in the field of biomedical research, concerns technologies that prolong human life, which evoke ethical questions of the justification of life-prolongation and of fairness in the resulting societal transformations (Kaufman, Shim and Russ, 2004). The complexity and ambivalence of ethical issues emerging from the design and outcomes of contemporary R&I call for the development of comprehensive methods that can be used by ethicists, researchers, policy-makers and various other stakeholders (technology users, companies, etc.). These methods can assist in the anticipation or foresight

---

[3] This chapter has been adopted from the paper: Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., & Gordijn, B. (2017). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. Science and Engineering Ethics. http://doi.org/10.1007/s11948-017-9961-8

and identification of ethical issues; normative assessment of ethical impacts and participation of the people involved in or affected by R&I processes.

In order to analyse proposed methods for practising ethics in R&I, we conducted a systematic review of the academic literature to identify methods for practising ethics in R&I. First, the scope and the methodology used to execute our literature searches are discussed. Second, the results of the literature review are presented and an analysis of these results is provided.

2.2   Practising Ethics in R&I

Practising ethics in R&I may be manifested in formulations of R&I project-specific codes of conduct or checklists of ethical issues and principles. Also, it can take the shape of ethicists joining in the design process of new technologies and innovations, and of researchers and other stakeholders[4] engaging with ethical challenges in a collaborative setting. By "methods" for practising ethics in R&I, we mean "detailed, content-specific […] problem solving procedures" (Nickles, 1987, p. 104) that aim at structuring the process of practising ethics in R&I. Thus, the methods that are discussed in this chapter will commonly present different procedural steps that should be part of practising ethics in R&I. "Methods for practising ethics in R&I" need to be distinguished from methods in conventional research ethics, which belong to a traditional branch of applied ethics that focuses on professional ethics of researchers, including for instance considerations of scientific integrity and treatment of human subjects in experiments. Though the subject matter of traditional research ethics and ethics in R&I show some family resemblance, the fields can be distinguished by their different foci. Whilst traditional research ethics is focused on normative aspects of doing professional scientific research, this chapter deals with methods that mostly focus on the ethics of *technological innovation and its ethical impacts*. Compared to conventional research ethics, which gained momentum after the Second World War with the Nuremberg Code, the academic discussion on methods for practising ethics in R&I gained traction only in the 1990s.

In Europe, in particular, the trend towards the inclusion of practising ethics in R&I projects funded by the European Commission is strongly related to the emergence of the "responsible research and innovation" (RRI) discourse. According to Owen et al., this relatively recent discourse revolves around three features: (1) an emphasis on science for

---

[4] In line with Achterkamp and Vos (2008), a stakeholder is conceptualised as either a group or an individual who potentially affects or is affected by an ethical impact and/or has a vested interest in the R&I context to which the ethical impact is ascribed.

society, discussing the impacts of science, (2) an emphasis on science with society, making R&I responsive to society, and (3) a re-evaluation of the concept of "responsible" as a moral ascription applied to the future-oriented, complex and collective phenomenon of R&I (Owen et al., 2012 p. 757). This is reflected by the notion of RRI as it is used in EU institutions, namely as an approach that "implies that societal actors (researchers, citizens, policy makers, businesses, third sector organisations, etc.), work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society" (European Commission, 2014)[5].

Notwithstanding the increasing interest of academics and governmental and non-governmental actors in ethics in R&I, philosophers and ethicists have only recently started to discuss this thematic and to develop methods to suggest how to incorporate ethical concerns in the R&I process (see also figure 2). This has had two notable consequences. First, it has led to a schism between academic debates on RRI on the one hand and ethics of technology with a focus on R&I on the other. Most ethicists of technology do not thoroughly engage with the RRI literature and vice-versa RRI scholars do not thoroughly engage with debates in ethics of technology. Recently, some notable exceptions emerged (cf. Blok and Lemmens 2015; Zwart et al. 2014) of scholars who actively connect the two fields. Second, it has led to an unreflective attitude in the field of ethics of technology regarding the notion of "research and innovation". R&I is often reduced to the process of technology development and design, and only little attention is paid to the diffusion of technological innovation through commercial processes, for instance through the process of "creative destruction" in industrial capitalism (Blok and Lemmens 2015, p. 30).

Our sample of the literature will be limited with regards to the abovementioned two consequences of the schism between RRI and ethics of technology. Concerning the first, our point of departure from the field of ethics of technology and therefore from explicit discussion of practising ethics in R&I will skew our sample towards ethics of technology and might thereby lead to the omission of potentially valuable insights from the broader and more comprehensive body of literature discussing RRI. We will therefore take some of this literature into account at a later point in chapter 3, notably when considering stakeholder participation (see section 3.3.3). Concerning the second, it will depart from a narrow and unreflective notion of R&I. The conception of R&I in ethics of technology as predominantly

---

[5] However, the RRI framework of the EU is not any direct adoption of academic frameworks such as the one developed by Owen et al. (2012). It focuses on pragmatic policy aims such as multi-actor and public engagement, gender equality, and science education.

associated with (1) publicly funded (2) technological innovation and (3) design[6] seems to lie close to the one criticised by Blok and Lemmens (2015), who reflect on the concept of innovation in the responsible innovation literature. They argue that innovation is uncritically seen as (1) always concerned with technology, (2) embedded within the economic paradigm, in that innovation always primarily ought to increase economic utility, and (3) inherently good in principle because they deliver on economic promises (e.g. jobs and prosperity) (Blok & Lemmens, 2015). Considered in this context ethics is more an afterthought, a reflection on a process of innovation that itself is inherently good and produces useful economic outcomes, but might also produce some negative externalities. Instead, Blok and Lemmens (2015) argue, the R&I process should be considered as not necessarily dealing with technologies (it might also deal with novel organisation practices), not necessarily having an economic purpose (it might also contribute to the public sphere), and even as being inherently conflictual, for instance through acknowledgment that economic profit and ecological goals cannot but clash, or that the creative destruction of innovation processes is not inherently good but inherently brings forth negativities (e.g. truck drivers massively losing their jobs because of the introduction of driverless trucks).

Even though the scope of our literature review will incorporate the unreflective notion of R&I, we will address especially the second and third point of criticism of this notion in the development of our method, by approaching the R&I process from a conception of the common good instead of from the economic paradigm and by acknowledging the inherent contractions in the R&I process between its aim and its negative externalities (see section 6.4.1). However, we will stick to the focus on technology and technological development in our conception of R&I because of the embedding of our study in the work on ethics of *technology*. Other innovations, in the areas of management, politics, arts, sports, education and so forth, will therefore be left out.

---

[6] For instance, Brey frames the context of his method as one of "technology development" (Brey, 2012b, p. 1), Borning et al. as one of "design of technology" (Borning et al., 2004) and Wright as – more broadly – one of "developing new technologies, services, projects, policies or programmes (Wright, 2011). Clearly, the development of technologies or of innovations is central, while their diffusion isn't taken into account. Scholars discussing the "Ethical Matrix" that is more focused on the introduction of innovations in society do focus somewhat on the diffusion aspect of innovations, but generally stick to publicly funded research and technological innovations (see e.g. Forsberg 2004; Jensen et al. 2011).

2.3 Scope and Methodology

This dissertation discusses practising ethics in R&I from the distinct perspective of ethics of technology. Coming from this perspective, it will predominantly focus on literature in applied ethics that deals with the themes of R&I and technological innovation. Therefore, even though the literature covered by this chapter overlaps with the discourse on RRI and by extension the earlier discourses on "ELSA" (ethical, legal and social aspects) and "ELSI" (ethical, legal and social implications/impacts) (Zwart et al. 2014), its scope is substantially narrower through focusing on literature that *explicitly* and *predominantly* deals with ethics and not with the broad spectrum of potential impacts of R&I (which might be labelled "social" and "legal" in addition to "ethical") in general. Moreover, it focuses in particular on technological innovations and their role in the R&I process and focuses less on ethical issues in scientific research (e.g. the use of human subjects in genomics research) or on the governance of science and technology[7]. As Grunwald (2011) argues, RRI brings together applied ethics, technology assessment (TA) as well as science, technology and society studies (STS) research. This literature review overlaps with these fields but always focuses on ethics and excludes foci on for instance policy-advice or descriptive accounts of sociotechnical systems. Considered in relation to STS, this literature review mostly captures what Mitcham characterises as one out of four approaches, which concerns "philosophical and ethical reflections on the essence and meaning of science and technology" (Mitcham, 1999, p. 130).

---

[7] For literature reviews that focus on RRI in general, or on the governance of science and technology, see for instance Burget et al., 2017, Forsberg et al., 2014, Landeweerd et al., 2015 and Stilgoe et al., 2013.

Figure 1: Schematic representation of the scope of the literature review. The literature on methods for practising ethics in R&I falls within the ELSA/ELSI/RRI discourse and is situated on the intersection of the fields of applied ethics as a sub branch of philosophy, STS, foresight studies and technology assessment.
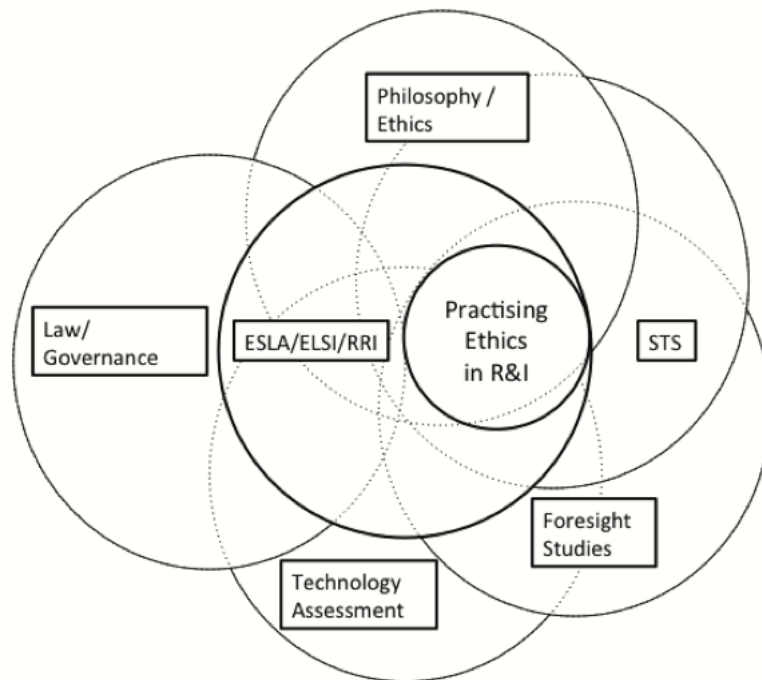
In order to obtain a broad overview of existing methods for practising ethics in the context of R&I, three major databases for scientific literature were consulted: Web of Knowledge[8] (containing major scientific databases such as Web of Science, BIOSIS, MEDLINE and SciELO Citation Index), Scopus[9] (including major scientific databases such as Elsevier, Wiley-Blackwell and IEEE) and the Springerlink database[10]. Seven search terms were selected by considering the most common concepts that are used in the literature to refer to practising ethics in R&I. Since the wordings that are closer to this chapter's characterisation of this thematic, namely "practising ethics" and "incorporating ethics" did not deliver many valuable results, the term "assessment" in combination with "ethics", "ethical", "technology" and "impacts" was used, as the practice of ethics in the fields of applied ethics and ethics of technology is mostly referred to as "assessment" and sometimes as "evaluation". Additionally, three search terms were formulated to ensure that the literature selection is properly embedded in the ELSA, ELSI and RRI discourses, using these three terms explicitly in conjunction with the wordings "method" "research" and "innovation". Only literature written in English has been included in the literature selection, which probably will have led

---

[8] Accessed through: http://apps.webofknowledge.com/
[9] Accessed through: http://www.scopus.com/
[10] Accessed through: http://link.springer.com/

to the exclusion of a number of interesting sources on methods for practicing ethics in R&I in other languages. Table 1 shows the total number of sources identified in each database for each particular query. Ten different searches in each database resulted in a total of 2626 hits (including duplicates).

The second selection of the literature was made manually according to three necessary conditions: the titles and abstracts of the available sources were scanned for the presence of the aspects of (1) ethics in an (2) R&I context, clearly mentioning *or* discussing (3) methods or approaches to practice ethics in R&I. The results of this selection are displayed in Table 1 under the headings "useful results". Certain sources, such as those explicitly focusing on methods for practising ethics in science *education,* on practising ethics in the context of *clinical practice* or discussing the practice of ethics in *management* or *organisational* contexts, however interesting, were omitted from the secondary selection results because of their lack of connection with R&I. Eventually, a total number of 107 useful sources was arrived at, which after a check for duplicates was reduced to 73 unique useful sources.

Table 1: Overview of search queries in different academic databases, the total number of search results per database query and the useful search results.

Search query:	Web of knowledge: total results	Web of knowledge: useful results	Scopus: total results	Scopus: useful results	Springerlink: total results	Springerlink: useful results
1. "Ethical assessment"	213	5	224	5	438	16
2. "Ethics assessment"	30	1	42	0	86	2
3. "Ethical impact assessment"	7	4	10	6	16	6
4. "Ethical technology assessment"	9	2	7	2	29	4
5. "Ethics methodology"	24	4	15[11]	4	89	5
6. "Practising	14	0	5	0	170[12]	2

[11] In Scopus, this search was conducted using the query "{ethics methodology}" in order to limit the initial unworkable number of search results (3781) by searching for occurrences of the specific combination of these two terms.

ethics"						
7. "Incorporating ethics" AND "research"	11	3	16	1	273	4
8. "Research and innovation" AND "method"	8	0	134	9	86	7
9. "ELSA" AND "research" AND "innovation" AND "method"	1	0	206	3	56[13]	5
10. "ELSI" AND "research" AND "innovation" AND "method"	1	0	120	0	281	7
Total sources:	**318**	**19**	**784**	**30**	**1524**	**58**

A third selection of the literature was made according to the backwards-snowballing method (Wohlin, 2014), looking at the reference lists of all useful sources found in the literature searches. During this snowballing process, sources were selected according to two criteria concerning the references: (1) their titles should explicitly mention the terms "ethics" or closely related terms (e.g., "ethical", "moral") *and* at least one of the terms "research", "innovation" or "technology" *or* (2) their titles should mention a specific methodology found in the earlier selection[14]. After this snowballing process, 63 additional sources[15] were selected, which gave a total of 136 useful sources.

---

[12] In Springerlink, this search was adjusted by filtering for "Applied Ethics" as the "subdiscipline" to which the results of the query were assigned, in order to limit the initial unworkable number of search results (482).

[13] In Springerlink, this search was adjusted by filtering for "Ethics" as the "subdiscipline" to which the results of the query were assigned, in order to limit the initial unworkable number of search results (567)

[14] This includes "the ethical Matrix", "ETHICS", "anticipatory ethics", "ethical technology assessment", "ethical impact assessment", "ethical dilemma scenarios", "value sensitive design", "the SBU approach", "the walkshop approach", "ethical parallel research", "just war theory", "human practices approach" and "interactive technology assessment".

[15] Three sources were left out of the final selection because they were not available (they were searched for by using two different institutional subscription systems).

2.4 Results

In this section, the results of the literature review are presented by considering (1) the distribution of the sources across different fields, and by (2) classifying and analysing the main traits of these methods.

The literature selection shows that the academic discussion on practising ethics in R&I commenced relatively recently, in the 1990s, and has gained considerable momentum in the last 10 years (see Figure 2). The literature also shows that ethics in the context of R&I is gaining in importance.

Figure 2: Number of sources published in five-year intervals, between 1990 and 2015.[16][17]
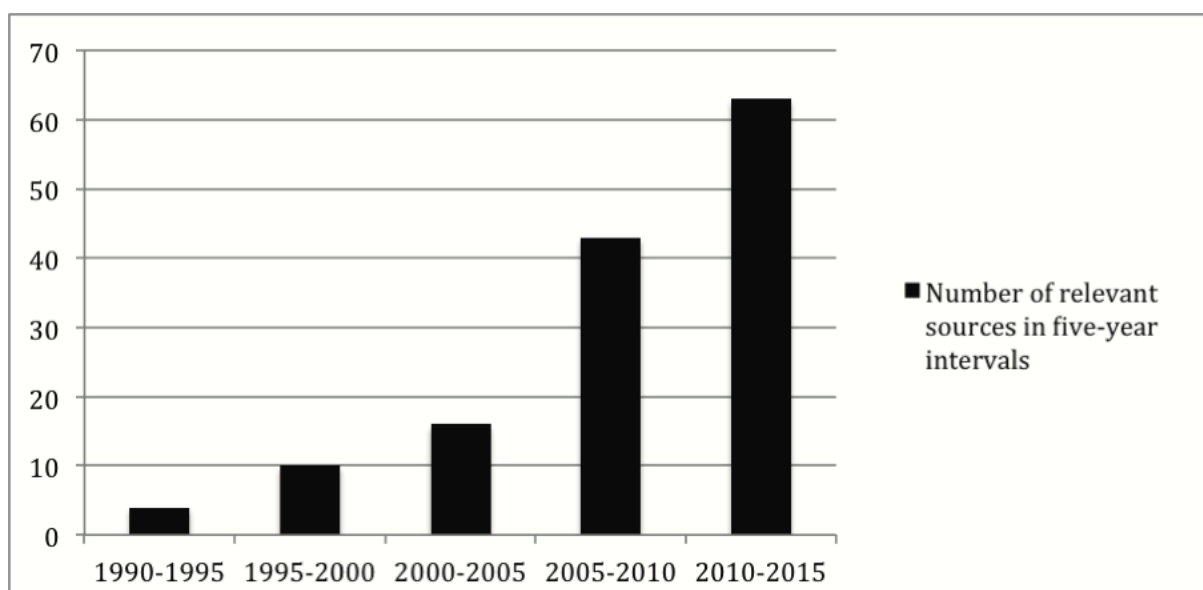
2.4.1   Practising ethics in different fields of R&I

Most of the sources that were selected do not discuss general methods for practising ethics in R&I, but focus on specific fields in which R&I processes take place. The following table provides an overview of the distribution of the different fields discussed in the selected literature:

Table 2: Classification of the sources according to the fields of R&I they discuss. If no specific field is discussed, classification is based on the thematic focus of the sources.

Field of R&I	Number of relevant sources

---

[16] The last interval is not fully representative, because the literature searches were conducted in October 2015. It could therefore be the case that more relevant sources were published after the literature searches in 2015 that are not taken into account in this overview.

[17] One search was conducted in 2017 (using the query "research and innovation" AND "method"). However, for the sake of consistency the results were filtered for sources that were published before November 2015.

Field of R&I	Number of relevant sources
Sources with a field-specific focus:	**93**
Health technologies	24
Information systems and/or systems development	13
ICT development and/or computer science	13
Nanotechnology	9
Agricultural and environmental research	9
Engineering sciences	8
Business development and innovation	5
Healthcare and medicine	4
Biomedical sciences	3
Operational research	2
Military research	2
Organisation studies	1
Sources with a general focus:	**43**
Emerging technologies	10
Responsible research and innovation	9
Technology design	9
Discussion of method(s)	7
Technology assessment	7
Ethics of technology	1
Total:	**136**

It can be observed that of the disciplinary fields of R&I, the field of health technologies is the most represented in the literature, followed by the fields of information systems research and computer science. The considerable number of sources that belong to these fields in comparison to, for example, organisational studies can be explained by the fact that technological innovation is the core focus of these fields.

R&I in the context of health technologies has traditionally been ethically sensitive because it deals with technologies that aim at assisting patients and vulnerable groups such as disabled and elderly people. This would explain its prominent position in the selected literature. Most discussions about ethics in the field of health technologies focus on the perceived need to increase the role of ethics in the broader practice of health technology assessment (HTA) (Braunack-Mayer, 2006; DeJean et al., 2009; Ten Have, 2013; Ten Have, 1995; Hofmann, 2008; Lehoux and Williams-Jones, 2007) and on the articulation of criteria that should inform methods for practising ethics in HTA (Arellano, Willett and Borry, 2011; Burls et al., 2011; Duthie and Bond, 2011; Grunwald, 2004; Potter et al., 2008; Saarni et al.,

2011; Sandman and Heintz, 2014). Importantly, though, discussions also focus on the perceived lack of clear methodological guidance for practising ethics in the field of HTA as it currently stands (Autti-Ramo and Makela, 2007; Burls et al., 2011; Hofmann, 2014). Ashcroft even questions the ability of the field of HTA to address ethical issues altogether, because of its focus on technical questions rather than on evaluative ones (Ashcroft, 1999).

As yet (2015), the selected sources discussing ethics in the field of health technologies do not show signs of widely used methods. The methods that were found for practising ethics proposed in the health technology literature are "checklist"-based methods (Heintz et al., 2015; Hofmann, 2005b), a "value analysis" that seems to be closely related to the value sensitive design approach (Hofmann, 2005a), a "rapid ethical assessment" method (Addissie et al., 2014), an approach according to which stakeholder participation can be organised (Autti-Ramo and Makela, 2007), a bibliometric method for conducting desk research of ethical issues in HTA (Droste et al., 2010), an "interactive technology assessment" approach that combines TA approaches with stakeholder engagement (Van der Wilt et al., 2015), a revised version of the "Socratic approach" (Hofmann et al., 2014) and an "triangular" approach that proposes to organise the practising ethics according to concrete steps, based on the principlism method that originates from biomedical ethics (Sacchini et al., 2009). None of these proposed methods seems to be (as of 2015) adopted by the health technologies field in a broad fashion.

In contrast to the field of health technologies, the fields of information systems research and computer science seem to have produced and seem to use more generally established methods for practising ethics, since several methods are discussed and developed in multiple sources by different authors. Notable examples of these are the ETHICS method (Adman and Warren, 2000; Arellano et al., 2011; Hirschheim and Klein, 1994; Leitch and Warren, 2010; Mumford, 1995; Singh et al., 2007; Wong and Tate, 1994), the Value Sensitive Design (VSD) method (Friedman, 1996; Friedman et al., 2006; Manders-Huits and Van den Hoven, 2009; Shilton, 2014; Van den Hoven, 2008; Van den Hoven, 2007; Van den Hoven and Manders-Huits, 2009; Le Dantec et al., 2009), the ETICA approach (Stahl, 2013; Stahl, 2011; Stahl et al., 2010; Wakunuma and Stahl 2014), the discourse ethics method (Rehg, 2015; Mingers and Walsham, 2010; Mittelstadt et al., 2015), the disclosive ethics approach (Brey, 2000; Light and McGrath, 2010), the ethical impact assessment approach that focuses on stakeholder consultation (Wright, 2010; Wright 2011; Bailey et al., 2013), human-driven design (Ikonen et al., 2012; Niemela et al., 2014) and a checklist approach (Van Gorp, 2009). Some sources in these fields discuss the general need for practising ethics,

and criteria for designing methods to do so (Brey, 2012a; Carew and Stapleton, 2013; Carpenter and Dittrich, 2013; Gorp, 2009; Tavani, 2013), and how ethical analyses could add to the general success of ICT projects (Stapleton, 2008). Markus and Mentzer (2014) discuss particular methodologies for conducting foresight studies, which could be integrated in an approach for practising ethics (such as the Delphi method, anticipatory technology ethics and sociotechnical transition analysis). Sassaman (2009) discusses specific ethical issues for computer security research, but without applying a distinct methodology.

Methods discussed in the field of nanotechnologies are the network approach and the impact and acceptability analysis method (Patenaude et al., 2015; van de Poel, 2008). Also, the organisation of interviews with nanotechnology researchers is discussed to show how ethical issues are incorporated in their work (Viseu and Maguire, 2012). The ethical matrix is widely used in agricultural and environmental R&I (Boucher and Gouch, 2012; Bruijnis et al., 2015; Heleski and Anthony, 2012; Kaiser et al., 2007; Whiting, 2004). Sources that consider ethics in the engineering sciences mostly focus on professional ethics for engineers: on how engineers should work in order to foster a practice of responsible engineering in R&I (Grunwald, 2001; Herkert, 2001; Riley, 2013; Verharen et al., 2013; Whitbeck, 2011). Methods used in the engineering sciences are the VSD method (van Wynsberghe and Robbins, 2013) and a scenario approach dealing with science-fiction narratives of technological prototypes (Stahl et al., 2014). In other, less well-represented fields, sources discuss the perceived need for practising ethics (for operations research) and criteria for doing so (Brans, 2004) as well as ways for practising professional ethics, notably in business settings (Bose, 2012; Fassin, 2000; Polonsky, 1998; Schumacher and Wasieleski, 2013). An embedded researcher approach (Reiter-Theil, 2004) and checklist approaches are used in the biomedical sciences (Winkler et al., 2011), an application of just war theory (Malsch, 2013) and a "metric of evil" method in military R&I (Reed and Jones, 2013), care ethics integrated in VSD in R&I in healthcare settings (Van Wynsberghe, 2013) and a stakeholder framework in operations research (Drake et al., 2009).

A considerable share of our selected sources does not deal with any particular field of R&I, but rather with technology design and development, or ethics of technology in general (see e.g. Bitay et al., 2005). Some sources especially focus on explicating the need to have methods for practising ethics in R&I and formulating concrete steps and criteria that should be part of such methods (Decker, 2004; Palm and Hansson, 2006; Skorupinski and Ott, 2002; Swierstra and Rip, 2007). Some of these sources especially focus on dealing with uncertainty when dealing with emerging technologies (Lucivero et al., 2011; Rommetveit et al., 2013;

Sollie, 2007) and the effects of practising ethics on R&I practices (Graffigna et al., 2010). Sources also present and discuss specific methods for practising ethics in R&I, such as codes of ethics or checklist approaches (Verharen and Tharakan, 2010), VSD, the ethical impact assessment (EIA) method (Wright, 2011, 2014; Wright & Friedewald, 2013), ethical scenario methods (Boenink et al., 2010; Ikonen et al., 2012; Ikonen and Kaasinen, 2008; Wright et al., 2014), the network approach (Zwart et al., 2006), the ethical matrix approach (Forsberg, 2004; Forsberg, 2007; Mepham and Tomkins, 2006), the human practices (HP) approach (Balmer and Bulpin, 2013), the anticipatory technology ethics (ATE) method (Brey, 2012a, 2012b), pro-ethical design (Floridi, 2015), the walkshop approach (Wickson et al., 2015) and the technological mediation approach (Verbeek, 2006). Some sources also critically assess the general role of ethics in R&I processes (Grunwald, 2000), focusing for instance on the difficulty of making epistemic claims about future impacts of R&I processes (Mittelstadt et al., 2015) and the difficulty of reconciling the assessment of a certain vision of future outputs of R&I processes with the idea of a person's future (Karafyllis, 2009).

A number of sources explicitly criticise methods for practising ethics, providing criticisms of the principlism approach or principled approaches in general (Groves, 2015; Page, 2012; Ten Have, 2014), the ethical matrix (Cotton, 2009; Jensen et al., 2011; Mepham, 2000; Schroeder and Palmer, 2003), VSD (Borning and Muller, 2012) checklist approaches (Masclet and Goujon 2012; Roberts, 1999), the ETHICS method (Stahl, 2007), the ETICA method (Rainey and Goujon, 2011) and the methodologies of specific R&I projects focussing on ethics (Thorstensen, 2014). Finally, sources contain discussions comparing different methods (Beekman and Brom, 2007; Doorn, 2012; Ferrari, 2010; Flipse et al., 2013; Forsberg et al., 2014; Gamborg, 2002; Hummels and de Leede, 2000; Lindfelt and Tornroos, 2006; Markus and Mentzer, 2014; Wickson and Forsberg 2014) and discussions about the effects of practising ethics on research professionals and on the R&I process (Foley et al., 2012; Schummer, 2004).

2.4.2 Analysis of methods for practising ethics in R&I

35 different methods for practising ethics in R&I were found in the selected sources, which were categorised according to their position with regard to the R&I process. We refer to these positions as "ex ante", which means that a method aims at practising ethics at the early stage of the R&I process, "intra", which means that a method aims at practising ethics during the design and testing stage of the R&I process and "ex post", which means that methods aim at practising ethics at the stage when an R&I process is already finished and has resulted in

concrete applications[18]. The table below (table 3) provides an overview of the methods, mapped according to their characterisation as ex ante, intra and ex post and according to their sub-aims (which are explained below). It must be noted that not each method fits perfectly in one of the three categories (e.g. discourse ethics can also be used for ex post analyses), but that when considering their general foci and their novelty the categories provide a useful heuristic for typifying the methods. In this section, the methods will be analysed by elucidating certain procedural steps that are common for each stage (ex ante, intra and ex post) of the R&I process. For each general type of method (1) the targeted users[19], (2) the over-arching aim and (3) typical procedural steps proposed (relating to the "who", "why" and "what" aspects of the methods) for the sub-aims of each main method-type will be discussed. The procedural steps will be illustrated by highlighting some of the methods.

Table 3: Categorisation of the methods for practising ethics in R&I according to their characterisation as "ex ante", "intra" and "ex post" and the sub-aims belonging to each of these main method-types. The number of sources using each method is indicated in brackets.

Method types (number of methods per type; sources using a method):		
Ex ante (8 methods; 16 sources):	**Intra (14 methods; 31 sources):**	**Ex post (13 methods; 27 sources)**
Identify emerging technologies	***Integrate ethicists in R&I contexts***	***Identify ethical impacts of existing technologies***
• Ethical Issues of Emerging ICT Applications (ETICA) (2)	• Ethical parallel research (1) • Embedded researcher (1) • Walkshop approach (1)	• Ethical checklist approaches (5) • Swedish Council of Health Technology Assessment (SBU) approach (1) • Survival ethics (1)
Understand ethical impacts of emerging technologies	***Disclose ethical issues in technology design***	***Evaluate ethical impacts of existing technologies***
• Scenario approaches (4)	• Disclosive ethics (2)	• Ethical Matrix (10) • Network approach (2) • Eclectic approach (1) • Human practices (HP)

---

[18] Van den Hoven (2008) also typifies methods in computer ethics as "ex ante" (emphasis on design) and "ex post" (emphasis on evaluation of existing technologies). However, we introduce the "intra" type to distinguish between methods that focus on the design process in which the conceptual steps have already been taken (a general idea of the type of technology is already present) vis-à-vis "ex ante" methods that focus on technological systems, artefacts and applications that might be designed at some point but have not entered the design process yet.

[19] By targeted user is meant the type of person who would use the method when engaging with ethics in R&I. The user could also be termed an "assessor", i.e., the person who is responsible for conducting an ethics assessment or review.

Method types (number of methods per type; sources using a method):		
Ex ante (8 methods; 16 sources):	**Intra (14 methods; 31 sources):**	**Ex post (13 methods; 27 sources)**
		approach (1)
Ethically evaluate emerging technologies	*Embed values in technology design*	*Organise the governance of ethical analyses*
• Anticipatory Technology Ethics (2) • Ethical technology assessment (eTA) (1) • Impact and acceptability analysis (1) • Ethics of uncertainty (1)	• Value sensitive design (VSD) (9) • Human-driven design (2) • Pro-ethical design (1) • Value analysis (1) • Triangular model (1) • Interactive technology assessment (iTA) (1) • Technological mediation approach (1)	• Socially responsible modelling framework (SRM) (1) • The Socratic approach (revisited) (1) • Ethic-innovation model (1) • Just war theory (JWT) (1)
Assess claims concerning the impacts of emerging technologies	*Organise practising ethics in R&I*	*Support ethical decision-making with technology*
• Discourse ethics (4) • Pragmatist NEST-Ethics (1)	• Effective, Technical and Human Implementation of Computer based Systems ETHICS (6) • Ethical Impact Assessment (EIA) (3) • Rapid Ethical Assessment (REA) (1)	• Metric of evil (1) • Ethical algorithm (1)

Ex ante methods

Ex ante methods position themselves in such a way that they try to organise practising ethics at the start of an R&I process, when the R&I activities have not yet been translated into any concrete design or application. These methods are therefore much concerned with "emerging technologies", the ethical impacts of which still lie in the future and can therefore only be "anticipated" or "foreseen". For this reason, ex ante methods often propose or include foresight approaches or other methods that construct scenarios to delineate ethical issues and/or desirable or undesirable futures that have been impacted by the ethical issues or their resolution. Most of the ex ante methods belong to a unified body of literature; except for the methods that are grouped under "scenario approaches". Scenario approaches are proposed under different headings (e.g. techno-ethical scenarios, ethical dilemma scenarios) by

different authors, but because they roughly share a common aim and structure they are grouped under a single heading. Ex ante methods mostly target **ethicists** and **RRI or foresight specialists** and thereby are typically expert-focused. Exceptions are the eTA method, which includes technology developers (Palm and Hansson, 2006, p. 555), Rehg's version of discourse ethics, which includes policy makers and a variety of stakeholders (Rehg, 2015, p. 38)[20], Ikonen and Kaasinen's (2008) version of a scenario approach and the ethical impact and accessibility analysis, which included a variety of actors (Patenaude et al., 2015). However, mechanisms for involving different stakeholders are scarcely discussed in ex ante methods.

The overarching aim of ex ante methods is to deal with the issue of uncertainty of technological change that results from "new" and "emerging" technologies. Sources belonging to five out of the eight ex ante methods refer to the Collingridge control dilemma, which stipulates that controlling a technology is difficult due to (1) the lack of knowledge about harmful impacts of a technology during the early stages of its development and (2) the difficulty of changing a technology once it has been embedded and stabilised in a society (Sollie, 2007, p. 297). Responding to this dilemma therefore provides the overall rationale for ex ante methods. A related issue that ex ante methods typically address is the interconnection between technological change and morality: technological change implies a change in morality and vice-versa (Boenink et al., 2010, p. 9). The sub-aims of ex ante methods can be characterised as follows: (1) to identify emerging technologies (e.g. the ETICA approach), (2) to understand possible future ethical impacts (e.g. scenario approaches), (3) to ethically evaluate emerging technologies (e.g. the ATE method) and (4) to critically assess the status of normative claims concerning ethical issues of emerging technologies (e.g. the discourse ethics method).

Procedural steps of ex ante methods

- Ex ante methods propose steps **to identify potential emerging technologies.** The ETICA project provides a good illustration of such procedural steps (Stahl, 2011; Stahl et al., 2010). First, it performs a foresight analysis to identify emerging technologies by capturing the current discourse on future ICTs, which entails a discourse analysis of documents issued by governments, research institutes and companies about expected impacts of emerging technologies. This review is

---

[20] It should be noted that Rehg's version of discourse ethics could also be characterised as ex post, which would explain the greater focus on a variety of stakeholders.

condensed in "meta-vignettes", which contain data about the emerging technologies, applications and artefacts (their defining features, related ethical issues, etc.). Second, the approach includes a bibliometric analysis that comprises a large amount of scholarly work to show which ethical concepts are used in relation to which (emerging) technology.

- Ex ante methods propose steps **to construct scenarios about future ethical impacts.** Scenarios are considered powerful tools for thinking about the future help understanding future ethical impacts of emerging technologies. The techno-ethical scenarios approach offers a good illustration of related procedural steps (Boenink et al., 2010). It proposes three steps: (1) a thorough analysis of the point of departure of the relevant moral landscape, (2) the introduction of technological development and an imaginary sketch of its interaction with the moral landscape and (3) the closure of moral controversies that had arisen due to the interaction in (2).

- Ex ante methods propose steps **to evaluate potential ethical impacts.** Impacts can be evaluated by offering a heuristic of ethical principles, such as is proposed by the eTA method. The ATE method (Brey 2012a) is a good illustration of the embedding of such a heuristic in a conceptual understanding of emerging technologies. It distinguishes three levels of ethical analysis: (1) analysis of the technology (collection of techniques related to a common purpose or domain), (2) analysis of the artefact (functional systems, artefacts and procedures based on a technology) and (3) analysis of the application level (the specific way in which artefacts are configured to be used). The ATE postulates an identification stage at which ethical impacts are identified and descriptions of a technology (at the three levels mentioned above) are analysed by means of a list of ethical values and principles. Brey proposes an evaluation stage, during which the relative importance of ethical impacts is assessed along with their likelihood of occurring.

- Ex ante methods propose steps **to assess the status of uncertain normative claims.** This assessment deals with the problem that claims about the future have a different epistemic status than claims about the past or present. The discourse method offers a good illustration of relevant steps that can be taken. The method of discourse ethics draws strongly from the work of Habermas (1990) and Apel (1980) and was firstly introduced to practise ethics in ICT R&I by Rehg (2015). It is based on two basic principles: the discourse and the universality principle (Mittelstadt et al., 2015, p.

1037). The discourse principle states that only those norms that could meet with the approval of all affected in their capacity as participants in a practical discourse could have a claim to validity. The universality principle states that the (unforeseen) consequences of adherence to a norm should be acceptable to all involved stakeholders (ibid.). The application of discourse ethics to ethical analysis of emerging technologies stipulates the following steps: (1) claims are broken down into constitutive parts, (2) uncertain claims about the future are assessed according to empirical evidence and general indicators about the future and (3) the normative position that relies on the uncertain claim is assessed for acceptability amongst relevant stakeholders.

Intra methods

Intra methods position themselves to practise ethics at the stage of an R&I in which conceptual ideas are being translated into a concrete technology design, and in which prototypes are made and tested. Methods focusing on this stage of the R&I process commonly deal with the translation of ethical values into design requirements and with the formulation of concrete design steps. Intra methods focus on three main groups of targeted users: (1) **ethicists**, mainly for disclosing ethical issues in design, determining how values can be embedded in design or understanding how they can partake in the R&I process, (2) **policy makers**, mainly for being able to organise the process of practising ethics in an R&I context and (3) **researchers and designers,** mainly for being able to integrally address ethical issues in the process of technology design.

The over-arching aim of intra methods is to enable, organise and ensure ethical technology design. Particular examples of ethical issues related to technology design are commonly referred to, such as care robots affecting a patient's autonomy (Van Wynsberghe, 2013, p. 409), social network services affecting the privacy of their users (Wright, 2011, p. 199) and the impact of cell phones on human communication and interaction (Verbeek, 2006, p. 2), to provide a rationale for their existence. The sub-aims of intra methods can be characterised as follows: (1) to disclose ethical issues in design (e.g. disclosive ethics), (2) to stipulate how values can be embedded in technology design (e.g. VSD), (3) to organise the process of practising ethics in the technology-development pipeline (e.g. EIA) and to make sure ethicists can work integrally in R&I projects, to be as "close" to the R&I process as possible (e.g. parallel researcher approach).

Procedural steps of intra methods

- Intra methods propose steps to **integrate ethicists in the R&I context.** This should bring ethicists closer to the everyday reality of R&I practitioners. The parallel researcher approach offers a good illustration of related procedural steps (Van Gorp and Van der Molen, 2011). It proposes five steps: (1) gathering information about an R&I project, (2) reflecting on resulting ethical issues and search for relevant literature, (3) prepare a discussion with R&I practitioners about the ethical issues and possible ways to mitigate these issues, (4) have the discussion and take decisions and (5) report about the ethical issues and the decisions made during the discussion.

- Intra methods propose steps to **disclose ethical issues in technology design.** The main issue that is addressed is at what point during the design process ethical issues are disclosed, and how they are disclosed. The disclosive ethics approach provides a pertinent illustration of related procedural steps. Brey stipulates two stages of ethical analysis: (1) the analysis of a technology on the basis of a moral value and (2) the use of a theory to formulate guidelines for the design process. Tavani discusses the three levels in the R&I process at which ethical analysis takes place: the disclosure level, at which ethical issues are identified, the theoretical level at which moral theory is developed and the application level, at which findings from moral theory are applied to the issues identified in the disclosure level (Tavani, 2013, p. 26). The disclosive ethics approach stipulates that practising ethics in R&I should be a multidisciplinary exercise, because the knowledge of researchers is needed as input at the disclosure level while the knowledge of ethicists is explicitly needed at the theoretical level.

- Intra methods propose steps to **embed values in technology design.** This is proposed to ensure that human values are taken into account during the design process, instead of post hoc when reflecting on a completed technology design. VSD offers a good illustration of related procedural steps (Friedman et al., 2006). At the centre of VSD lies a "tripartite" methodology that sets out three stages of investigations that aim at stipulating how human values can be embedded in technology design. The first (1) of these stages is the conceptual stage in which working conceptualisations of relevant human values are proposed and basic questions are answered, e.g., about the relevant stakeholders. The second (2) stage is the empirical one, in which social science methods (qualitative and quantitative) are used to gather empirical data about embedding values in technology design, for instance, by looking at how stakeholders

consider certain values in a use-context. The third stage (3) is the technical stage during which researchers investigate how technical properties of technologies both hinder and promote human values.

- Intra methods propose steps to **organise "practising ethics" in the design process.** This effort focuses on the governance of ethics in an R&I context, prescribing what it should contribute at which stage of the R&I process. The EIA method provides a fitting illustration of procedural steps (Wright, 2011). It presents a list of 15 steps that should be part of an EIA, namely, (1) to determine whether an EIA is needed, (2) to identify the team of assessors and its terms of reference, (3) to prepare an EIA plan, (4) to describe the process to be assessed, (5) to identify the stakeholders, (6) to analyse the ethical impacts, (7) to consult with stakeholders, (8) to check whether the R&I project complies with regulations, (9) to identify risks and possible solutions, (10) to formulate recommendations, (11) to prepare and publish an EIA report, (12) to implement the recommendations, (13) to organise a (third-party) review and audit of the EIA, (14) to update the EIA if changes in the R&I project occur and (15) to ensure ethical awareness throughout the organisation conducting the R&I project (Wright & Friedewald, 2013, p. 763).

Ex post methods

Ex post methods are concerned with what is often characterised as "analysing ethical issues" of technologies that already exist, as outcomes of R&I processes. These methods therefore mostly engage in retrospective ethical reflections that take known ethical issues of known technologies as their subject (such as privacy issues arising from the use of social media). Ex post methods predominantly target **experts**, which can be ethicists, institutional bodies (e.g. government agencies focusing on ethics), expert groups or specific R&I professionals. Some methods focus on the **general public** (any type of relevant stakeholder), such as the ethical matrix.

The over-arching aim of ex post methods is to identify, analyse and resolve ethical impacts of technologies. These methods commonly refer to problems faced by R&I practitioners to consider ethical issues in their work, for instance by members of the "network on appropriate technology" (Verharen and Tharaken, 2010, p. 36), modellers in operation research (Drake et al., 2009) and corporate governance practitioners (Schumacher and Wasieleski, 2013), as a rationale for their existence. Additionally, they refer to the integral role of ethics in a respective R&I context (Heintz et al., 2015) and to the need to democratise

the ethical assessment of technologies (Mepham, 2000). The sub-aims of ex post methods can be characterised as follows: (1) to identify ethical impacts related to technologies (e.g. the human practices approach), (2) to analyse ethical impacts (e.g. the ethical matrix), (3) to organise the ethical analysis of technologies from a governance or compliance perspective (e.g. the ethic-innovation approach) and (4) to support ethical decision-making with technology (concerning technologies in use) (e.g. the metric of evil).

Procedural steps of ex post methods

- Ex post methods propose steps to **identify ethical issues raised by existing technologies.** The focus here lies on providing heuristics for practitioners to explore ethical issues in their work. Checklist approaches are perhaps the best illustration in this context, because they are widely used in practical settings. Even though they don't form a singular, coherent method, they can be grouped under one heading because of their common structure and purpose. A checklist approach can consist of a list of "items", such as "flourishing" or "freedom" (Verharen and Tharakan, 2010, p. 39), a list of questions such as "are their moral challenges related to components of the technology?" (Burls et al., 2011, p. 233), or a list of ethical issues such as "safety" and "sustainability" (Gorp, 2009, p. 36).

- Ex post methods propose steps to **analyse ethical issues raised by existing technologies.** Analysis of ethical issues is commonly based on a heuristic of ethical principles, in combination for instance with a list of stakeholders. The ethical matrix is a congruous illustration of this. It makes use of the principlism approach of Beauchamp and Childress (2001) by providing principles that are guidelines for the evaluation of R&I outcomes. It depicts the principles in a visual matrix on one axis and different relevant stakeholders on the second axis. In each cell of the matrix, assessors can reason how a certain principle might be either infringed upon or promoted by a certain R&I outcome, depending on the respective stakeholder (Mepham, 2000, p. 171).

- Ex post methods propose steps to **organise the governance of ethical analyses.** In this context, the focus lies to some extend on "compliance", on determining whether R&I processes comply with certain ethical standards. A demonstrative illustration of this is the ethic-innovation method, which stipulates four stages in the governance of ethical analyses in commercial contexts (Schumacher and Wasieleski, 2013, p. 27).

The first stage revolves around determining the time horizon of the relevant decision makers. The second stage aims to determine the degree of ethical sensitivity. The third stage provides steps to integrate ethics in a company's value system, goals-set, strategic decisions and business models. The fourth step characterises innovation decisions that take ethical values into account.

- Ex post methods propose steps to **support ethical decision-making with existing technology.** Currently, this seems to be an increasing area of interest with for instance the advent of self-driving cars that need "ethical" decision-making systems. The "metric of evil" proposed in the context of military research provides a pertinent illustration of related procedural steps (Reed and Jones, 2013). It revolves around designing an "equation" for determining the potential evil for an action in a military context. This requires the identification of the potential consequences of actions, determining parameters that capture the feature of a baseline moral system, calibrate parameter values using expert consultation and incremental adjustment of parameter weightings.

2.5 Chapter Summary

The previous decades have seen a great increase in discussions about methods for practising ethics in R&I. This chapter provided an overview and analysis of these methods. As we have seen, methods focus on different phases of the R&I process, from the early conceptual phase in which emerging technologies can still take many forms and applications to the phase of actual introduction of novel technological innovations in society. Even though these methods have many merits, a number of which have been highlighted in the analysis, they also show a number of general shortcomings. In the next chapter, we will discuss these shortcomings and possible ways to overcome them.

2.6 References

Achterkamp, M. C., & Vos, J. F. J. (2008). Investigating the Use of the Stakeholder Notion in Project Management Literature, a meta-analysis. *International Journal of Project Management*, *26*(7), 749–757. http://doi.org/10.1016/j.ijproman.2007.10.001

Addissie, A., Davey, G., Newport, M. J., Addissie, T., MacGregor, H., Feleke, Y., & Farsides, B. (2014). A Mixed-methods Study on Perceptions Towards use of rapid ethical assessment to improve informed consent processes for health research in a low-income setting. *BMC Medical Ethics*, *15*(1), 35. http://doi.org/10.1186/1472-6939-15-35

Adman, P., & Warren, L. (2000). Participatory Sociotechnical Design of Organizations and Information Systems - an adaptation of ETHICS methodology. *Journal of Information Technology*, *15*(1), 39–51.

http://doi.org/10.1080/026839600344393

Apel, K.-O. (1980). *Towards a Transformation of Philosophy*. (G. Adey and D. Frisby, Trans.). Routledge and Kegan Paul: London.

Arellano, L. E., Willett, J. M., & Borry, P. (2011). International Survey on Attitudes toward Ethics in Health Technology Assessment: An exploratory study. *International Journal of Technology Assessment in Health Care*, *27*(1), 50–54. http://doi.org/http://dx.doi.org/10.1017/S0266462310001182

Ashcroft, R. (1999). Ethics and Health Technology Assessment. *Monash Bioethics Review*, *18*(2).

Autti-Ramo, I., & Makela, M. (2007). Ethical Evaluation in Health Technology Assessment Reports: an eclectic approach. *International Journal of Technology Assessment in Health Care*, *23*(1), 1–8. http://doi.org/10.1017/S0266462307051501

Balmer, A. S., & Bulpin, K. J. (2013). Left to their Own Devices: Post-ELSI, ethical equipment and the International Genetically Engineered Machine (iGEM) Competition. BioSocieties, *8*(3), 311–335. http://doi.org/10.1057/biosoc.2013.13

Boucher, P., & Gough, C. (2012). Mapping the Ethical Landscape of Carbon Capture and Storage. *Poiesis Und Praxis*, *9*(3–4), 249–270. http://doi.org/10.1007/s10202-012-0117-2

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of Biomedical Ethics*. Oxford: Oxford University Press.

Beekman, V., & Brom, F. W. A. (2007). Ethical Tools to Support Systematic Public Deliberations about the Ethical Aspects of Agricultural Biotechnologies. *Journal of Agricultural and Environmental Ethics*, *20*(1), 3–12. http://doi.org/http://dx.doi.org.ezproxy.usherbrooke.ca/10.1007/s10806-006-9024-7

Bitay, B., Brandtand, D., & Savelsberg, E. (2005). The Global Validity of Ethics: Applying ethics to engineering and technology development. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *16*, 19–24.

Blok, V. and Lemmens, P. (2015). The Emerging Concept of Responsible Innovation. Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation. In Koops, B. J., Oosterlaken, I., Romijn, H., Swierstra, T., and van den Hoven, J. (eds) *Responsible Innovation 2: Concepts, Approaches, and Applications*. Heidelberg: Springer, pp. 1–303. doi: 10.1007/978-3-319-17308-5.

Boenink, M., Swierstra, T., & Stemerding, D. (2010). Anticipating the Interaction between Technology and Morality: A Scenario Study of Experimenting with Humans in Bionanotechnology. *Studies in Ethics, Law, and Technology*, *4*(1). http://doi.org/10.2202/1941-6008.1098

Bohn, J., Coroama, V., Langheinrich, M., & Mattern, M. (2005). Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing. *Ambient Intelligence*, *10*(5), 5–29. http://doi.org/10.1007/3-540-27139-2_2

Borning, A., & Muller, M. (2012). Next steps for Value Sensitive Design. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems* - CHI '12, 1-10. http://doi.org/10.1145/2207676.2208560

Bose, U. (2012). An ethical framework in information systems decision making using normative theories of business ethics. *Ethics and Information Technology*, *14*(1), 17–26. http://doi.org/10.1007/s10676-011-9283-5

Brans, J. P. (2004). The management of the future Ethics in OR: Respect, multicriteria management, happiness. *European Journal of Operational Research*, *153*(2), 466–467. http://doi.org/10.1016/S0377-

2217(03)00166-8

Braunack-Mayer, A. J. (2006). Ethics and health technology assessment: handmaiden and/or critic? *International Journal of Technology Assessment in Health Care*, *22*(3), 307–312. http://doi.org/10.1017/S0266462306051191

Brey, P. (2000). Disclosive Computer Ethics: The Exposure and Evaluation of Embedded Normativity in Computer Technology. *Computers and Society*, *30*(4), 10–16.

Brey, P. (2012a). Anticipating ethical issues in emerging IT. *Ethics and Information Technology*, *14*, 305–317. http://doi.org/10.1007/s10676-012-9293-y

Brey, P. (2012b). Anticipatory Ethics for Emerging Technologies. *NanoEthics*, *6*(1), 1–13. http://doi.org/10.1007/s11569-012-0141-7

Bruijnis, M. R. N., Blok, V., Stassen, E. N., & Gremmen, H. G. J. (2015). Moral Lock-In in Responsible Innovation: The Ethical and Social Aspects of Killing Day-Old Chicks and Its Alternatives. *Journal of Agricultural and Environmental Ethics*, *28*(5), 939–960. http://doi.org/10.1007/s10806-015-9566-7

Burget, M., Bardone, E., & Pedaste, M. (2017). Definitions and Conceptual Dimensions of Responsible Research and Innovation: A Literature Review. *Science and Engineering Ethics*, *23*(1), 1-19. http://doi.org/10.1007/s11948-016-9782-1

Burls, A., Caron, L., Cleret de Langavant, G., Dondorp, W., Harstall, C., Pathak-Sen, E., & Hofmann, B. (2011). Tackling ethical issues in health technology assessment: a proposed framework. *International Journal of Technology Assessment in Health Care*, *27*(3), 230–237. http://doi.org/http://dx.doi.org/10.1017/S0266462311000250

Carew, P. J., & Stapleton, L. (2013). Towards empathy: a human-centred analysis of rationality, ethics and praxis in systems development. *Ai & Society*, *29*(2), 149–166. http://doi.org/10.1007/s00146-013-0472-0

Carpenter, K. J., & Dittrich, D. (2013). Bridging the Distance: Removing the Technology Buffer and Seeking Consistent Ethical Analysis in Computer Security Research. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. http://doi.org/10.1017/CBO9781107415324.004

Cotton, M. (2009). Evaluating the "ethical matrix" as a radioactive waste management deliberative decision-support tool. *Environmental Values*, *18*(2), 153–176. http://doi.org/10.3197/096327109X438044

Decker, M. (2004). The role of ethics in interdisciplinary technology assessment. *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science*, *2*(2–3), 139–156. http://doi.org/10.1007/s10202-003-0047-0

DeJean, D., Giacomini, M., Schwartz, L., & Miller, F. A. (2009). Ethics in Canadian health technology assessment: a descriptive review. *International Journal of Technology Assessment in Health Care*, *25*(4), 463–469. http://doi.org/http://dx.doi.org/10.1017/S0266462309990390

Doorn, N. (2012). Responsibility Ascriptions in Technology Development and Engineering: Three Perspectives. *Science and Engineering Ethics*, *18*(1), 69–90. http://doi.org/10.1007/s11948-009-9189-3

Drake, M. J., Gerde, V. W., & Wasieleski, D. M. (2009). Socially responsible modeling: a stakeholder approach to the implementation of ethical modeling in operations research. *OR Spectrum*, *33*(1), 1–26. http://doi.org/10.1007/s00291-009-0172-9

Droste, S., Dintsios, C. M., & Gerber, A. (2010). Information on ethical issues in health technology assessment: how and where to find them. *International Journal of Technology Assessment in Health Care, 26*(4), 441–

449. http://doi.org/http://dx.doi.org/10.1017/S0266462310000954

Duthie, K., & Bond, K. (2011). Improving ethics analysis in health technology assessment. *International Journal of Technology Assessment in Health Care*, *27*(1), 64–70. http://doi.org/http://dx.doi.org/10.1017/S0266462310001303

European Commission. (2014). *Responsible research & innovation.* Brussels: EU Publication Office.

Fassin, Y. (2000). Innovation and ethics ethical considerations in the innovation business. *Journal of Business Ethics*, *27*(1/2), 193–203. http://doi.org/10.1023/A:1006427106307

Ferrari, A. (2010). Developments in the debate on nanoethics: Traditional approaches and the need for new kinds of analysis. *NanoEthics*, *4*(1), 27–52. http://doi.org/10.1007/s11569-009-0081-z

Flipse, S. M., van der Sanden, M. C. A., & Osseweijer, P. (2013). The Why and How of Enabling the Integration of Social and Ethical Aspects in Research and Development. *Science and Engineering Ethics*, *19*(3), 703–725. http://doi.org/10.1007/s11948-012-9423-2

Floridi, L. (2015). Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration. *Science and Engineering Ethics*. http://doi.org/10.1007/s11948-015-9733-2

Foley, R. W., Bennett, I., & Wetmore, J. M. (2012). Practitioners' Views on Responsibility: Applying Nanoethics. *NanoEthics*, *6*, 231–241. http://doi.org/10.1007/s11569-012-0154-2

Forsberg, E. (2004). The Ethical Matrix — A Tool for Ethical Assessments of Biotechnology Ellen-Marie Forsberg. *Global Bioethics*, *17*(1). http://doi.org/10.1080/11287462.2004.10800856

Forsberg, E. M. (2007). Pluralism, the ethical matrix, and coming to conclusions. *Journal of Agricultural and Environmental Ethics*, *20*, 455–468. http://doi.org/10.1007/s10806-007-9050-0

Forsberg, E. M., Thorstensen, E., Nielsen, R. Ø., & de Bakker, E. (2014). Assessments of emerging science and technologies: Mapping the landscape. *Science and Public Policy*, *41*(3), 306–316. http://doi.org/10.1093/scipol/scu025

Friedman, B. (1996). Value-sensitive design. *Interactions*, *3*(6), 16–23. http://doi.org/10.1145/242485.242493

Friedman, B., Kahn Jr., P. H., & Borning, A. (2006). Value Sensitive Design and Information Systems. In K. E. Himma & H. T. Tavani (Eds.), *Human-Computer Interaction and Management Information Systems: Foundations* (pp. 1–27). John Wiley & Sons, Inc. http://doi.org/10.1145/242485.242493

Gamborg, C. (2002). The acceptability of forest management practices: An analysis of ethical accounting and the ethical matrix. *Forest Policy and Economics*, *4*(3), 175–186. http://doi.org/10.1016/S1389-9341(02)00007-2

Graffigna, G., Bosio, a. C., & Olson, K. (2010). How do ethics assessments frame results of comparative qualitative research? A theory of technique approach. *International Journal of Social Research Methodology*, *13*(4), 341–355. http://doi.org/10.1080/13645570903209076

Groves, C. (2015). Logic of Choice or Logic of Care? Uncertainty, Technological Mediation and Responsible Innovation. *NanoEthics*, *9*(3), 321–333. http://doi.org/10.1007/s11569-015-0238-x

Grunwald, A. (2000). Against over-estimating the role of ethics in technology. *Science and Engineering Ethics*, *6*(2), 181–196. http://doi.org/10.1007/s11948-000-0046-7

Grunwald, a. (2001). The application of ethics to engineering and the engineer's moral responsibility: perspectives for a research agenda. *Science and Engineering Ethics*, *7*(3), 415–28. http://doi.org/10.1007/s11948-001-0063-1

Grunwald, A. (2004). The normative basis of (health) technology assessment and the role of ethical expertise. *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science*, *2*, 175–193. http://doi.org/10.1007/s10202-003-0050-5

Grunwald, A. (2011). Responsible Innovation: Bringing together Technology Assessment, Applied Ethics, and STS research. *Enterprise and Work Innovation Studies*, 7, IET, 9-31.

Habermas, J. (1990). *Moral consciousness and communicative action.* (C. Lenhardt and S. W. Nicholsen Trans.). Cambridge: MIT Press.

Hauser, J., Tellis, G. J., & Griffin, A. (2006). Research on Innovation: A Review and Agenda for Marketing Science. *Marketing Science*, *25*(6), 687–717. http://doi.org/10.1287/mksc.1050.0144

Heintz, E., Lintamo, L., Hultcrantz, M., Jacobson, S., Levi, R., Munthe, C., … Sandman, L. (2015). Framework for Systematic Identification of Ethical Aspects of Healthcare Technologies: the Sbu Approach. *International Journal of Technology Assessment in Health Care*, *31*(3), 124–130. http://doi.org/10.1017/S0266462315000264

Heleski, C. R., & Anthony, R. (2012). Science alone is not always enough: The importance of ethical assessment for a more comprehensive view of equine welfare. *Journal of Veterinary Behavior: Clinical Applications and Research*, *7*(3), 169–178. http://doi.org/10.1016/j.jveb.2011.08.003

Herkert, J. R. (2001). Future directions in engineering ethics research: microethics, macroethics and the role of professional societies. *Science and Engineering Ethics*, *7*(3), 403–14. http://doi.org/10.1007/s11948-001-0062-2

Hirschheim, R., & Klein, H. K. (1994). Realizing Emancipatory Principles in Information Systems Development: The Case for ETHICS. *MIS Quarterly*, *18*(1), 83–109. http://doi.org/10.2307/249611

Hofmann, B. (2005a). On value-judgements and ethics in health technology assessment. *Poiesis Und Praxis*, *3*(4), 277–295. http://doi.org/10.1007/s10202-005-0073-1

Hofmann, B. (2005b). Toward a procedure for integrating moral issues in health technology assessment. *International Journal of Technology Assessment in Health Care*, *21*(3), 312–318. http://doi.org/10.1017/S0266462305050415

Hofmann, B. M. (2008). Why ethics should be part of health technology assessment. *International Journal of Technology Assessment in Health Care*, *24*(4), 423–429. http://doi.org/http://dx.doi.org/10.1017/S0266462308080550

Hofmann, B. (2014). Why not integrate ethics in HTA : identification and assessment of the reasons. *GMS Health Technology Assessment*, *10*, 1–9. http://doi.org/10.3205/hta000120

Hofmann, B., Oortwijn, W., Cleemput, I., & Sacchini, D. (2014). Harmonization of Ethics in Health Technology Assessment: A Revision of the Socratic Approach. *International Journal of Technology Assessment in Health Care*, *30*(1), 3–9. http://doi.org/10.1017/S0266462313000688

Hummels, H., & de Leede, J. (2000). Teamwork and Morality: Comparing Lean Production and Sociotechnology. *Journal of Business Ethics*, *26*, 75–88. http://doi.org/10.1023/A:1006242516664

Ikonen, V., & Kaasinen, E. (2008). Ethical Assessment of Future-Oriented Design Scenarios. In T. W. Bynum, M. Calzarossa, I. de Lotto, & S. Rogerson (Eds.), *Ethicomp 2008* (pp. 433–445). Pavia.

Ikonen, V., Kaasinen, E., Heikkilä, P., & Niemelä, M. (2012). Human-driven design of micro- and nanotechnology based future sensor systems. *Journal of Information, Communication and Ethics in*

Society, 13(2), 110–129. http://doi.org/10.1108/JICES-07-2013-0021

Jensen, K. K., Forsberg, E. M., Gamborg, C., Millar, K., & Sandøe, P. (2011). Facilitating Ethical Reflection Among Scientists Using the Ethical Matrix. *Science and Engineering Ethics*, *17*(3), 425–445. http://doi.org/10.1007/s11948-010-9218-2

Kaiser, M., Millar, K., Thorstensen, E., & Tomkins, S. (2007). Developing the Ethical Matrix as a decision support framework: GM fish as a case study. *Journal of Agricultural and Environmental Ethics*, *20*(1), 65–80. http://doi.org/10.1007/s10806-006-9023-8

Kajikawa, Y., Usui, O., Hakata, K., Yasunaga, Y., & Matsushima, K. (2008). Structure of knowledge in the science and technology roadmaps. *Technological Forecasting and Social Change*, *75*(1), 1–11. http://doi.org/10.1016/j.techfore.2007.02.011

Karafyllis, N. C. (2009). Facts or Fiction? A Critique on Vision Assessment as a Tool for Technology Assessment. In P. Sollie & M. Duwell (Eds.), *Evaluating new technologies: Methodological problems for the ethical assessment of technology developments*. Heidelberg: Springer. http://doi.org/10.1007/s13398-014-0173-7.2

Kaufman, S. R., Shim, J. K., & Russ, A. J. (2004). Revisiting the biomedicalization of aging: clinical trends and ethical challenges. *The Gerontologist*, *44*(6), 731–738. http://doi.org/10.1093/geront/44.6.731

Landeweerd, L., Townend, D., Mesman, J., & Van Hoyweghen, I. (2015). Reflections on different governance styles in regulating science: a contribution to "Responsible Research and Innovation." *Life Sciences, Society and Policy*, *11*(8), 1–22. http://doi.org/10.1186/s40504-015-0026-y

Le Dantec, C. A., Poole, E. S. E. S., & Wyche, S. P. S. P. (2009). Values as lived experience: Evolving value sensitive design in support of value discovery. *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*, 1141–1150. http://doi.org/10.1145/1518701.1518875

Lehoux, P., & Williams-Jones, B. (2007). Mapping the integration of social and ethical issues in health technology assessment. *International Journal of Technology Assessment in Health Care, 1*, 9–16. http://doi.org/10.1017/s0266462307051513

Leitch, S., & Warren, M. J. (2010). ETHICS: The past, present and future of socio-technical systems design. In: A. Tatnall (Ed.), History of computing. Learning from the past (pp. 189–197). Heidelberg: Springer.

Light, B., & McGrath, K. (2010). Ethics and Social Networking Sites: a Disclosive Analysis of Facebook. *Information Technology and People*, 23(4), 290–311.

Lindfelt, L.-L., & Tornroos, J.-A. (2006). Ethics and value creation in business research: comparing two approaches. *European Journal of Marketing*, *40*(3/4), 328–351.

Lucivero, F., Swierstra, T., & Boenink, M. (2011). Assessing Expectations: Towards a Toolbox for an Ethics of Emerging Technologies. *NanoEthics*, *5*, 129–141. http://doi.org/10.1007/s11569-011-0119-x

Malsch, I. (2013). The Just War Theory and the Ethical Governance of Research. *Science and Engineering Ethics*, *19*(2), 461–486. http://doi.org/10.1007/s11948-012-9357-8

Manders-Huits, N., & Van den Hoven, J. (2009). The Need for a Value-Sensitive Design of Communication Infrastructures. In P. Sollie & M. Duwell (Eds.), *Evaluating new technologies: Methodological problems for the ethical assessment of technology developments*. Heidelberg: Springer. http://doi.org/10.1007/s13398-014-0173-7.2

Markus, M. L., & Mentzer, K. (2014). Foresight for a responsible future with ICT. *Inf Syst Front*, (January),

353–368. http://doi.org/10.1007/s10796-013-9479-9

Masclet, L., & Goujon, P. (2012). Implementing ethics in information systems, presuppositions and consequences in ethics and information systems. In: Magda David Hercheui; Diane Whitehouse; William McIver; Jackie Phahlamohlaka. *10th International Conference on Human Choice and Computers (HCC)*, Sep 2012, Amsterdam, Netherlands. Springer, IFIP Advances in Information and Communication Technology, AICT-386, pp.287-298, 2012, ICT Critical Infrastructures and Society. http://doi.org/10.1007/978-3-642-33332-3_26

Mepham, B. (2000). A Framework for the Ethical Analysis of Novel Foods: The Ethical Matrix. *Journal of Agricultural and Environmental Ethics*, *12*(2), 165–176. http://doi.org/10.1023/A:1009542714497

Mepham, B., Kaiser, M., Thorstensen, E., Tomkins, S., & Millar, K. (2006). *Ethical Matrix Manual*. The Hague.

Mingers, J., & Walsham, G. (2010). Toward Ethical Information Systems: the Contribution of Discourse Ethics. *MIS Quarterly*, *34*(4), 833–854. http://doi.org/Article

Mitcham, C. (1999). Why Science, Technology, and Society Studies? *Bulletin of Science and Technology in Society*, *19*(2), 128–134.

Mitcham, C., & Waks, L. J. (1996). Technology in Applied Ethics: Moving From the Margins To the Center. *Bulletin of Science, Technology & Society*, *16*(4), 217–226. http://doi.org/10.1177/027046769601600441

Mittelstadt, B. D., Stahl, B. C., & Fairweather, N. Ben. (2015). How to Shape a Better Future? Epistemic Difficulties for Ethical Assessment and Anticipatory Governance of Emerging Technologies. *Ethical Theory and Moral Practice*, 1027–1047. http://doi.org/10.1007/s10677-015-9582-8

Mumford, E. (1995). *Effective Systems Design and Requirements Analysis - The Ethics Approach*. London: Macmillan Press Ltd. http://doi.org/10.1017/CBO9781107415324.004

Nickles, T. (1987). Methodology, heuristics, and rationality. In J. C. Pitt & M. Pera (Eds.), *Rational Changes in Science*. Dordrecht: D. Reidel Publishing Company.

Niemela, M., Ikonen, V., Leikas, J., Kantola, K., Kulju, M., Tammela, A., & Ylikauppila, M. (2014). Human-Driven Design: A Human-Driven Approach to the Design of Technology. IFIP Advances in Information and Communication Technology, 431, 78–91. http://doi.org/10.1007/978-3-662-44208-1_8

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*, *39*(6), 751–760. http://doi.org/10.1093/scipol/scs093

Page, K. (2012). The four principles: can they be measured and do they predict ethical decision making? *BMC Medical Ethics*, *13*(1), 10. http://doi.org/10.1186/1472-6939-13-10

Palm, E., & Hansson, S. O. (2006). The case for ethical technology assessment (eTA). *Technological Forecasting and Social Change*, *73*(5), 543–558. http://doi.org/10.1016/j.techfore.2005.06.002

Patenaude, J., Legault, G.-A., Beauvais, J., Bernier, L., Béland, J.-P., Boissy, P., … Tapin, D. (2015). Framework for the analysis of nanotechnologies' impacts and ethical acceptability: basis of an interdisciplinary approach to assessing novel technologies. *Science and Engineering Ethics*, *21*(2), 293–315. http://doi.org/10.1007/s11948-014-9543-y

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artifacts: or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, *14*(3),

221–232. http://doi.org/10.1177/030631284014003004

Polonsky, M. J. (1998). Incorporating Ethics into Business Students' Research Projects: A Process Approach. *Journal of Business Ethics*, *17*(11), 1227–1241.

Potter, B. K., Avard, D., Graham, I. D., Entwistle, V. A., Caulfield, T. A., Chakraborty, P., … Wilson, B. J. (2008). Guidance for considering ethical, legal, and social issues in health technology assessment: application to genetic screening. *International Journal of Technology Assessment in Health Care*, *24*(4), 412–422. http://doi.org/http://dx.doi.org/10.1017/S0266462308080549

Rainey, S., & Goujon, P. (2011). Toward a normative ethics for technology development. *Journal of Information, Communication and Ethics in Society*, *9*(3), 157–179. http://doi.org/10.1108/JICES-07-2013-0021

Reed, G. S., & Jones, N. (2013). Toward Modeling and Automating Ethical Decision Making: Design, Implementation, Limitations, and Responsibilities. *Topoi*, *32*(2), 237–250. http://doi.org/10.1007/s11245-012-9127-x

Rehg, W. (2015). Discourse ethics for computer ethics: a heuristic for engaged dialogical reflection. *Ethics and Information Technology*, *17*(1), 27–39. http://doi.org/10.1007/s10676-014-9359-0

Reiter-Theil, S. (2004). Does empirical research make bioethics more relevant? "The embedded researcher" as a methodological approach. *Medicine, Health Care, and Philosophy*, *7*(1), 17–29. http://doi.org/10.1023/B:MHEP.0000021846.20508.c8

Riley, D. (2013). Hidden in Plain View: Feminists Doing Engineering Ethics, Engineers Doing Feminist Ethics. *Science and Engineering Ethics*, *19*, 189–206. http://doi.org/10.1007/s11948-011-9320-0

Roberts, E. B. (2007). Managing invention and innovation. *Research Technology Management*, *50*(1), 35–54. http://doi.org/10.1038/427679a

Roberts, L. W. (1999). Ethical dimensions of psychiatric research: a constructive, criterion-based approach to protocol preparation. The Research Protocol Ethics Assessment Tool (RePEAT). *Biological Psychiatry*, *46*(8), 1106–1119. http://doi.org/10.1016/S0006-3223(99)00146-8

Rommetveit, K., Gunnarsdóttir, K., Jepsen, K. S., Bertilsson, T. M., Verrax, F., & Strand, R. (2013). The Technolife project: an experimental approach to new ethical frameworks for emerging science and technology. *International Journal of Sustainable Development*, *16*(1/2), 23–45. http://doi.org/10.1504/IJSD.2013.053789

Saarni, S. I., Braunack-Mayer, A., Hofmann, B., & Van Der Wilt, G. J. (2011). Different methods for ethical analysis in health technology assessment: An empirical study. *International Journal of Technology Assessment in Health Care*, *27*(4), 305–312. http://doi.org/10.1017/S0266462311000444

Sacchini, D., Virdis, A., Refolo, P., Pennacchini, M., & de Paula, I. C. (2009). Health technology assessment (HTA): ethical aspects. *Medicine, Health Care, and Philosophy*, *12*(4), 453–7. http://doi.org/10.1007/s11019-009-9206-y

Sandman, L., & Heintz, E. (2014). Assessment vs. appraisal of ethical aspects of health technology assessment: can the distinction be upheld? *GMS Health Technology Assessment*, *10*. http://doi.org/http://dx.doi.org/10.3205/hta000121

Sassaman, L. (2009). Ethical Guidelines for Computer Security Researchers: "Be Reasonable." In *Financial Cryptography and Data Security* (Vol. 5628, pp. 230–237). http://doi.org/10.1007/978-3-642-03549-4

Schroeder, D., & Palmer, C. a. (2003). Technology assessment and the ethical matrix. *Poiesis & Praxis*, *1*, 295–307. http://doi.org/10.1007/s10202-003-0027-4

Schumacher, E. G., & Wasieleski, D. M. (2013). Institutionalizing Ethical Innovation in Organizations: An Integrated Causal Model of Moral Innovation Decision Processes. *Journal of Business Ethics*, *113*(1), 15–37. http://doi.org/10.1007/s10551-012-1277-7

Schummer, J. (2004). Societal and Ethical Impliations of Nanotechnology - Meanings, Interest Groups, and Social Dynamics. *Techne*, *8*, 56–87.

Shilton, K. (2014). This is an Intervention: Foregrounding and Operationalizing Ethics During Technology Design. In K. D. Pimple (Ed.), Emerging Pervasive Information and Communication Technologies (PICT) (pp. 176–192). Heidelberg: Springer. http://doi.org/10.1007/978-94-007-6833-8

Singh, R., Wood, B., & Wood-Harper, T. (2007). Socio-technical design of the 21st century: A vision. *IFIP International Federation for Information Processing*, *235*, 503–506. http://doi.org/10.1007/978-0-387-72804-9_39

Skorupinski, B., & Ott, K. (2002). Technology assessment and ethics. *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science*, *1*, 95–122. http://doi.org/10.1007/s102020100010

Sollie, P. (2007). Ethics, technology development and uncertainty: an outline for any future ethics of technology. *Journal of Information, Communication and Ethics in Society*, *5*(4), 293–306. http://doi.org/10.1108/14779960710846155

Stahl, B. C. (2007). ETHICS, Morality and Critique: An Essay on Mumford´s Socio-Technical Approach. *Journal of the Association for Information Systems*, *8*(9), 479–490. http://doi.org/Article

Stahl, B. C. (2011). IT for a better future: how to integrate ethics, politics and innovation. *Journal of Information, Communication and Ethics in Society*, *9*(3), 140–156. http://doi.org/10.1108/14779961111167630

Stahl, B. C. (2013). Virtual suicide and other ethical issues of emerging information technologies. *Futures, 50*, 35–43. http://doi.org/10.1016/j.futures.2013.03.004

Stahl, B. C., Heersmink, R., Goujon, P., Flick, C., van den Hoven, J., Wakunuma, K., … Rader, M. (2010). Identifying the Ethics of Emerging Information and Communication Technologies. *International Journal of Technoethics*, *1*(4), 20–38. http://doi.org/10.4018/jte.2010100102

Stahl, B. C., McBride, N., Wakunuma, K., & Flick, C. (2014). The empathic care robot: A prototype of responsible research and innovation. *Technological Forecasting and Social Change, 84*, 74–85. http://doi.org/10.1016/j.techfore.2013.08.001

Stapleton, L. (2008). Ethical decision making in technology development: a case study of participation in a large-scale information systems development project. *Ai & Society*, *22*(3), 405–429. http://doi.org/10.1007/s00146-007-0150-1

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. http://doi.org/http://dx.doi.org/10.1016/j.respol.2013.05.008

Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. *NanoEthics*, *1*(1), 3–20. http://doi.org/10.1007/s11569-007-0005-8

Tavani, H. T. (2013). *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing*. New York: John Wiley & Sons, Inc. http://doi.org/10.1017/CBO9781107415324.004

Ten Have, H. (2014). Theoretical models and approaches to ethics. In H. ten Have & B. Gordijn (Eds.), *Bioethics in a European Perspective* (pp. 1–5). Dordrecht: Springer. http://doi.org/10.1007/s13398-014-0173-7.2

Ten Have, H. (2004). Ethical perspectives on health technology assessment. *International Journal of Technology Assessment in Health Care, 1*, 71–76.

Ten Have, H. (1995). Medical technology assessment and ethics. Ambivalent relations. *Hastings Center Report*, *25*(5), 13–19.

Thorstensen, E. (2014). Public Involvement and Narrative Fallacies of Nanotechnologies. *NanoEthics*, 8(3), 227–240. http://doi.org/10.1007/s11569-014-0202-1

Van de Poel, I. (2008). How should we do nanoethics? A network approach for discerning ethical issues in nanotechnology. *NanoEthics*, *2*, 25–38. http://doi.org/10.1007/s11569-008-0026-y

Van der Wilt, G. J., Reuzel, R., & Grin, J. (2015). Design for Values in Healthcare Technology. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains,* Heidelberg: Springer. (pp. 1–871). http://doi.org/10.1007/978-94-007-6970-0

Van den Hoven, J. (2008). Moral methodology and information technology. In K. E. Himma & H. T. Tavani (Eds.), *The Handbook of Information Systems Research*. London: John Wiley & Sons, Inc. http://doi.org/10.4018/978-1-59140-144-5

Van den Hoven, J. (2007). ICT and Value Sensitive Design. In P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa, & V. Laurent (Eds.), *IFIP International Federation for Information Processing* (Vol. 233, pp. 75–83). Boston: Springer. http://doi.org/10.1002/9781118551424.ch4

Van den Hoven, J. & Manders-Huits, N. (2009). Value-sensitive Design. In J. Kyrre, B. Olsen, & V. F. Hendricks (Eds.), *A Companion to the Philosophy of Technology*. Malden: Blackwell Publishing. http://doi.org/10.1002/9781444310795.ch1

Van Gorp, A. (2009). Ethics in and During Technological Research; An Addition to IT Ethics and Science Ethics. In *Evaluating New Technologies* (Vol. 3). Heidelberg: Springer. http://doi.org/10.1007/978-90-481-2229-5_6

van Gorp, A., & van der Molen, S. (2011). Parallel, Embedded or Just Part of the Team: Ethicists Cooperating Within a European Security Research Project. *Science and Engineering Ethics, 17*(1), 31–43. http://doi.org/10.1007/s11948-009-9187-5

Van Wynsberghe, A. (2013). Designing robots for care: care centered value-sensitive design. *Science and Engineering Ethics*, *19*(2), 407–33. http://doi.org/10.1007/s11948-011-9343-6

Van Wynsberghe, A., & Robbins, S. (2013). Ethicist as Designer: A Pragmatic Approach to Ethics in the Lab. *Science and Engineering Ethics*, 947–961. http://doi.org/10.1007/s11948-013-9498-4

Verbeek, P.-P. (2006). Materializing Morality. *Science, Technology & Human Values*, *31*(3), 361–380. http://doi.org/10.1177/0162243905285847

Verharen, C. C., & Tharakan, J. (2010). Barefoot ethics: Social justice through an appropriate technology checklist. In V. Dzikzeniyo (Ed.), *Proceedings of the 4th international conference on appropriate Technology* (pp. 1–179).

Verharen, C., Tharakan, J., Middendorf, G., Castro-Sitiriche, M., & Kadoda, G. (2013). Introducing Survival

Ethics into Engineering Education and Practice. *Science and Engineering Ethics*, *19*(2), 599–623. http://doi.org/10.1007/s11948-011-9332-9

Viseu, A., & Maguire, H. (2012). Integrating and Enacting "Social and Ethical Issues" in Nanotechnology Practices. *NanoEthics*, *6*(3), 195–209. http://doi.org/10.1007/s11569-012-0162-2

Wakunuma, K. J., & Stahl, B. C. (2014). Tomorrow's ethics and today's response: An investigation into the ways information systems professionals perceive and address emerging ethical issues. *Information Systems Frontiers, 16*(3), 383–397. http://doi.org/10.1007/s10796-014-9490-9

Whitbeck, C. (2011). *Ethics in Engineering Practice and Research*. Cambridge: Cambridge University Press.

Whiting, T. L. (2004). Application of the ethical matrix in evaluation of the question of downer cattletransport. In: *Proceedings CanWest Veterinary Conference*, October 2–5, 2004. Banff: Alberta and British Columbia Veterinary Medical Associations.

Wickson, F., & Forsberg, E. M. (2014). Standardising Responsibility? The Significance of Interstitial Spaces. *Science and Engineering Ethics*, *21*(5), 1159–1180. http://doi.org/10.1007/s11948-014-9602-4

Wickson, F., Strand, R., & Kjølberg, K. L. (2015). The walkshop approach to science and technology ethics. *Science and Engineering Ethics, 21*(1), 241–264. http://doi.org/10.1007/s11948-014-9526-z

Winkler, E. C., Hiddemann, W., & Marckmann, G. (2011). Ethical assessment of life-prolonging treatment. *The Lancet Oncology*, *12*(8), 720–2. http://doi.org/10.1016/S1470-2045(11)70148-6

Wohlin, C. (2014). Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. *18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014)*, 1–10. http://doi.org/10.1145/2601248.2601268

Wong, E., & Tate, G. (1994). A study of user participation in information systems development. *Journal of Information Technology*.

Wright, D. (2011). A framework for the ethical impact assessment of information technology. *Ethics and Information Technology*, *13*, 199–226. http://doi.org/10.1007/s10676-010-9242-6

Wright, D. (2014). Ethical impact assessment. *Ethics, Science, Technology, and Engineering*, *163*(c), 163–167. http://doi.org/10.1016/j.clsr.2011.11.007

Wright, D., Finn, R., Gellert, R., Gutwirth, S., Schütz, P., Friedewald, M., Venier, S., Mordini, E. (2014). Ethical dilemma scenarios and emerging technologies. *Technological Forecasting and Social Change*, *87*, 325–336. http://doi.org/10.1016/j.techfore.2013.12.008

Wright, D., & Friedewald, M. (2013). Integrating privacy and ethical impact assessments. *Science and Public Policy*, *40*(6), 740–754. http://doi.org/10.1093/scipol/sct083

Zwart, H., Landeweerd, L., & van Rooij, A. (2014). Adapt or perish? Assessing the recent shift in the European research funding arena from "ELSA" to "RRI." *Life Sciences, Society and Policy, 10*(1), 1-19. http://doi.org/10.1186/s40504-014-0011-x

Zwart, S. D., van de Poel, I., van Mil, H., & Brumsen, M. (2006). A network approach for distinguishing ethical issues in research and development. *Science & Engineering Ethics*, *12*(4), 663–684.

3 Criticism of Existing Methods and Recommendations

3.1 Introduction

In this chapter, a number of shortcomings of the methods for practising ethics in R&I are discussed and recommendations for improvements are presented. The discussion is based on the analysis of methods in the previous chapter, which is graphically represented in the figure below (figure 3).

Figure 3: A map for characterising methods for practising ethics in R&I. The centre contains the three main method types: ex ante, intra and ex post. The second circle provides a sub-categorisation of methods according to the type of procedural steps they focus on. The third circle displays the overarching aims of the three main method types and the fourth circle displays the issue common to methods across the main three method types of appropriate stakeholder participation.
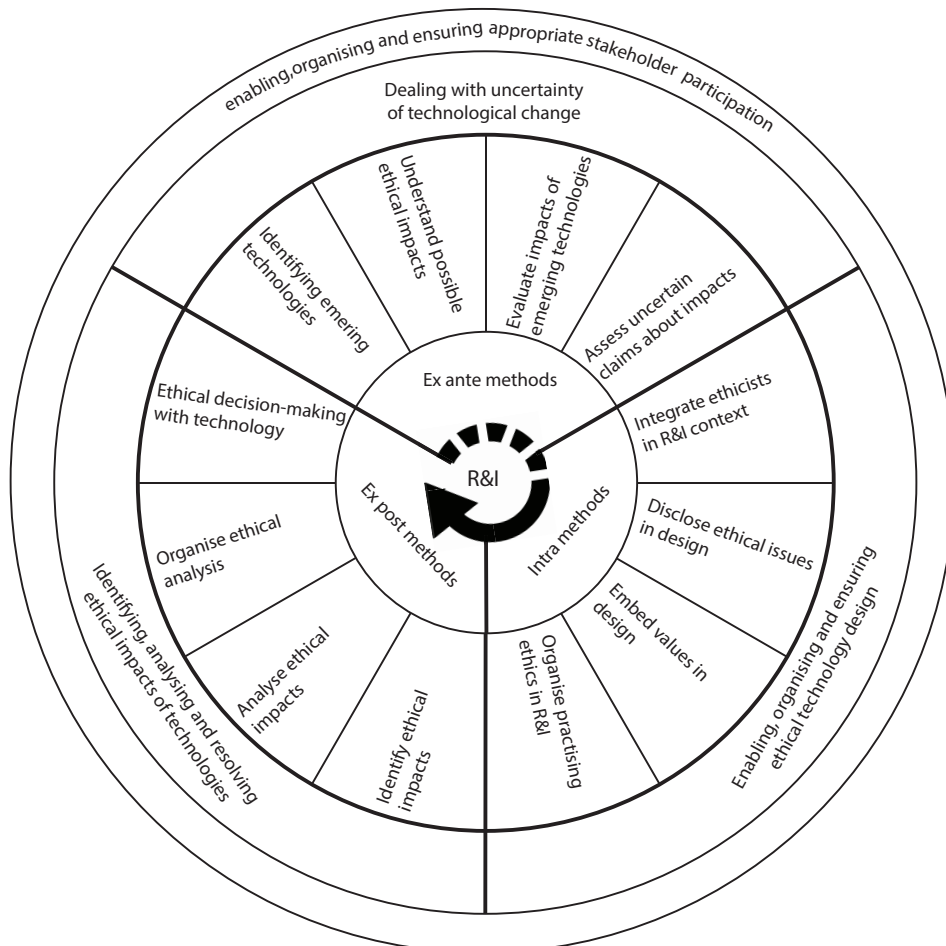
The criticism of existing methods is structured by considering the over-arching aims of the three types of methods that were analysed (ex ante, intra, ex post) and one common

challenge that all methods have to deal with, concerning the appropriate participation of stakeholders. For each aim, problems in the methods to deal with it are discussed and recommendations towards the solving of these problems are put forward. The discussion is based on two types of input: (1) sources from the literature selection that critically discuss a specific challenge (or a related method), (2) auxiliary sources from the literature on philosophy, applied ethics, RRI, STS, foresight studies and technology assessment that offer additional relevant insights. An important limitation of the discussion is that it does not provide a comprehensive overview of all the problems in the methods for practising ethics in R&I and all possible ways to deal with these problems. Rather, it presents a tentative overview of some of the key problems that were encountered in the literature and some suggestions for dealing with these problems.

From this tentative overview, we proceed towards the formulation of general recommendations, which are integrated in the text containing the criticism. These general recommendations leave room for many different avenues for working on the shortcomings in the existing methods and could therefore be taken up by any researcher wishing to work on advancing methods for practising ethics in R&I. Second, we discuss more specific, strategic choices based on the general recommendations, leading us to a framework for the construction of a novel method for practising ethics in R&I that will be developed in the following chapters. These strategic choices narrow down the effort towards constructing a novel method, by considering choosing between different bodies of literature and methods. At the end of this chapter, we therefore end up with a reasoned proposal for constructing a novel method for practising ethics.

3.2 Criticism and General Recommendations[21]

In this section, we formulate a general critique of the methods that were analysed in the previous chapter. We do so by discussing the over-arching aims of (1) dealing with uncertainty of technological change for ex ante methods, (2) dealing with ethical technology design for intra methods and (3) identifying, analysing and resolving ethical impacts for ex post methods, and the common challenge of appropriate stakeholder participation in the process of practising ethics. Based on this critique, we will formulate specific recommendations that will guide the finding of solutions to shortcomings in the methods.

---

[21] This section has been adopted from: Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., & Gordijn, B. (2017). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. Science and Engineering Ethics. http://doi.org/10.1007/s11948-017-9961-8

3.2.1 Uncertainty of technological change

The over-arching aim of ex ante methods is dealing with the uncertainty in R&I processes, especially pertaining to the ethical reflection on emerging technologies. The methods that were analysed deal with uncertainty differently: by means of scenario building, foresight methods, bibliographical analysis and customised heuristics. Two main problems related to these approaches were identified.

First, the aspect of *speculation* in future-oriented ethical frameworks constitutes a problem. As Nordmann argues, speculation about the future or about the "if and then" should be rejected from a philosophical point of view, amongst other alternatives for the reason that it is impossible for us "to imagine ourselves as something other than we are" (Nordmann, 2007, p. 41). Procedural approaches such as discourse ethics and the ethics of uncertainty to some extent deal with this issue by offering ways to assess uncertain claims, but they do not stipulate what type of claims one should be looking for in a discourse about emerging technologies. Lucivero et al. argue along similar lines as does Nordmann, and they contend that one ought to look into the structure of present expectations or promises (a reflection on the present) rather than to speculate about future impacts (Lucivero et al., 2011). As Borup et al. explain, "expectations play a central role in science and technology not least because they mediate across boundaries between different scales, levels, times and communities" (Borup et al., 2006, p. 289). Expectations could for instance be understood along the lines of "future imaginaries" constructed by actors dealing with emerging technologies (Groves, 2013).

Second, a problem arises out of the implicit assumption in the methods that were analysed: that one *can* arrive at a situation in which one has gained sufficient knowledge about the future to stipulate procedures for action or guidance in R&I processes. As Markus and Mentzer argue, "the attempt to anticipate future conditions is frequently frustrated by *unpredictable* technological and social discontinuities" (Markus and Mentzer, 2014, p. 363). Similarly, Vallor argues that emerging technologies bring about "acute technosocial opacity", meaning an uncertainty that comes with the growing complexity and ubiquity of technologies that play a role in our everyday lives (Vallor, 2016, p. 6). She argues against the idea that the consequences of emerging technologies can always be meaningfully predicted, and instead claims that ethics of emerging technologies should not merely focus on action-guidance based on foresight but also on the cultivation of virtuous character. According to this approach, one should also aim to cultivate ones abilities to predict and cope with unforeseen consequences, rather than only trying to formulate predictions that offer ground for action-guidance. Pandza and Ellwood offer a good illustration for the relevance of virtue ethics

when considering the issue of uncertainty, by showing how cultivation of virtues of R&I practitioners assists them in dealing with responsibility in situations of high uncertainty (Pandza and Ellwood, 2013).

Recommendations:

- Methods for practising ethics in R&I that make use of methodological constructs to imagine or foresee possible futures pertaining to the development and use of emerging technologies should more thoroughly engage in an epistemological discussion of the limits of knowledge pertaining to such foresight.

- Whenever future development and use of emerging technologies cannot be meaningfully foreseen, methods for practising ethics in R&I should take appropriate approaches into account that divert from action-guidance based on speculative knowledge about the future, such as approaches for the analysis of present promises and expectations concerning emerging technologies and approaches in virtue ethics.

3.2.2 Ethical technology design

The over-arching aim of intra methods is enabling, organising and ensuring ethical technology design. The methods that were analysed offer different, complementary ways of dealing with this: by integrating ethicists in R&I practises, identifying ethical issues in design, embedding values in design and organising ethical design in the R&I process. Two key problems in these approaches were identified.

First, next to integrating ethicists in the R&I process, as is suggested by some of the approaches (e.g. the embedded researcher approach), researchers should be able to integrate ethics in their work. As Brey argues, the knowledge of researchers is pivotal for disclosing ethical issues in design for their knowledge about the design and its potential usually substantially surpasses that of the ethicist working in R&I projects (Brey, 2000). Borning and Muller acknowledge this issue when stressing that the VSD approach should make "clearer the voice of the researcher" (Borning and Muller, 2012, p. 1125). Along similar lines, Le Dantec et al. argue that VSD offers "inadequate guidance on empirical tools" (Le Dantec et al., 2009, p. 1141) and additionally suggest that the empirical investigation should be given priority in VSD and related approaches. VSD is primarily considered here, but this issue seems to persist in the other methods for ethical technology design for none of them offers a way for R&I practitioners to integrate ethics in their work. An exception might be the ETHICS approach, which seems more accessible to R&I practitioners and much more closely

relates to their day-to-day reality (at least in the context of information systems R&I). However, the ETHICS approach suffers from another weakness, namely that it lacks an ethical foundation, which makes it difficult to argue that the design choices resulting from the use of the ETHICS method can also be justified as *ethical* choices (Stahl, 2007).

Second, the notion of embedding values in design as it is pragmatically used in approaches dealing with ethical technology design is insufficiently theoretically supported. As Van de Poel (2013) for instance explains, the *translation* of values into design requirements – the "how" of values in design – is not adequately dealt with. In line with this critique, Manders-Huits argues that the concept of "values" in VSD and their realisation is left undetermined (Manders-Huits, 2011, p. 271). Also, Albrechtslund argues that VSD and related approaches insufficiently deal with the difference between designer's intentions and users' practice (Albrechtslund, 2007, p. 63). All these concerns seem to focus around a lack of theoretical grounding of *how* values can be embedded in technology design, or how technology design *mediates* certain values. Not only VSD has to deal with this problem, but also other methods dealing with embedding values in design such as the triangular model, value analysis and human-driven design. Possible ways to deal with this issue are the introduction of a value hierarchy, as argued by Van de Poel (2013), but more sources seem to point at a more thorough interaction between methods dealing with ethical technology design and theories in STS and philosophy of technology. For instance, Manders-Huits (2011) mentions Winner's account of a technological arrangement mediating political values (Winner, 1980), Verbeek (2008) mentions Latour's notion of technical mediation through programs of action (Latour, 1994) and Albrechtslund (2007) and Spahn (2015) mention the notion of technological mediation in Ihde (1990) and Verbeek (2005) as entry points for integrating the understanding of values embedded in design with technological mediation.

Recommendations:
- Approaches dealing with ethical technology design should focus more on the integration of ethics in the day-to-day work of R&I practitioners, especially with regard to the disclosure of ethical issues in design.
- Considerations of methodological aspects of ethical technology design should be based on a normative theoretical framework that explicates how certain technology design choices can be identified as ethical, or how "ethics" is mediated by technology design.

3.2.3   Identifying, analysing and resolving ethical impacts

The over-arching aim of ex post methods is to identify, analyse and resolve the ethical impacts of developed technologies. They deal with these aspects by offering heuristics such as checklists, analytic frameworks and ethical decision-making procedures. Two key problems in these approaches were identified.

First, a general problem that arises from many of the established methods is the problem of value conflicts (conflicts between different ethical principles that apply). For instance, the ethical principles of security and privacy might conflict in an ethical analysis of cyber security technologies. Principled or checklist-based approaches based on ethical principles can create such normative conflicts or at least questions since they do not include a hierarchy or lexical order that could help to decide which principle or moral value is to be prioritised. As Schroeder and Palmer argue, the widely used ethical matrix is inadequate in terms of "weighing the ethical problems that it uncovers" (Schroeder and Palmer, 2003, p. 295). Moreover, Cotton argues that the approach could instigate "potential conflict among stakeholders" (Cotton, 2009, p. 164). The problems should not be uniquely attributed to the ethical matrix, but apply to all methods offering heuristics for the identification and analysis of ethical impacts without the provision of methods for dealing with value conflicts. The pragmatic NEST-ethics (Swierstra and Rip, 2007) approach offers a way to deal with this problem, by focusing on the argumentative patterns around an ethical controversy instead of on the ethical-decision making process. However, this approach has the disadvantage that it focuses merely on description and prediction and insufficiently offers a normative, prescriptive perspective (Brey, 2012a). From this, one can concede that if a method for an ethical analysis of technology includes multiple values, a well-grounded and justified order of those values should be provided. For instance, decision criteria for resolving value conflicts could be provided (Wenstøp and Koppang, 2009).

Second, ex post methods offer inadequate guidance on how to choose between sociotechnical alternatives or courses of action based on an ethical analysis. Arguably, this issue is relevant for both intra and ex post methods, though technologies need to have been developed to a significant extend to meaningfully consider choices of sociotechnical alternatives. As Markus and Mentzer argue, an important limitation of literature dealing with the analysis of ethical impacts of technologies is that it "offers little guidance on how to identify the sociotechnical alternatives that should be compared for their ethical consequences" (Markus and Mentzer, 2014, p. 359). Related to this, Page argues that the heuristics used for ethical analyses, such as lists of ethical principles, do not by themselves

lead to adequate decision-making when R&I practitioners are confronted with different alternatives (Page, 2012). A way to deal with this problem is by engaging with multiple criteria design analysis (Van de Poel, 2009). As Van Gorp indicates, ethical issues can be related to actions in the design process by considering the "operationalization of technical requirements and the making of trade-offs" (Van Gorp, 2005, p. 154). Another approach suggested by Cotton is a deliberative ethical procedure that focuses on problem framing, option scoping, criteria elicitation, and option appraisal (Cotton, 2009, p. 167). When considering the concrete reality of practising ethics on R&I practise, Shilton proposes an interesting model for determining when an intervention in the R&I process can be considered successful (Shilton, 2014).

Recommendation:

- Researchers and assessors should use a convincing methodological solution for the problem of value conflicts, when they occur. This could be done by including procedures for reasoned balancing of ethics principles whenever no fixed and justified ranking of principles is provided.
- Methods that analyse ethical impacts of technologies should offer procedural guidance that would allow for using the analysis to choose between certain sociotechnical alternatives.

3.2.4   A common challenge: appropriate participation of stakeholders

A challenge that presents itself across the main method-types concerns the appropriate identification and participation of stakeholders in the process of practising ethics in R&I. Participation is seen to be important because there is a plurality of value systems (different ethical perspectives), which means that different values can only be considered when relevant stakeholders participate in the process of practising ethics. Two problems concerning this challenge were identified that cut across the methods.

First, a problem that is inadequately addressed is the appropriate identification of *relevant* stakeholders: who should be included in the process. Even though many methods formulate procedural steps entailing that stakeholders should be identified, how this should be done remains generally unclear. As Schroeder and Palmer argue when discussing the ethical matrix, "difficulties arise in this analysis both with respect to what constitutes stakeholders […] and also concerning how to deal with those one might definitely want to include as stakeholders […] but who are unable to enter deliberative discourse themselves"

(Schroeder and Palmer, 2003, p. 301). In relation to VSD, Borning and Muller discuss this as "the problem of speaking for others", of ethicists speaking for those without a voice by, for instance, determining which stakeholders can be considered relevant (Borning and Muller, 2012, p. 1130). Additionally, as Ferrari argues, the identification of different stakeholders is based on the problematic assumption that different people with different particular and isolated interests exist (Ferrari, 2010, p. 36). Considering the methods that were analysed, these problems do not limit themselves to the ethical matrix and VSD, but extend to all the methods that discuss "stakeholder identification" but do not offer ways how to justifiably do so. The network approach seems to be an exception, for it provides an account of relevant stakeholders by offering several criteria for the selection of stakeholders that pertain to the position of a stakeholder in relation to the decision-making process in an R&I project (Zwart et al., 2006, p. 672). This relates to the suggestion of Cotton, who argues that a "meta-ethically justified process for the selection of stakeholders is necessary – a mapping device for identifying actors and the relationships between them" (Cotton, 2009, p. 166). Other possible ways to deal with the above-mentioned problem is, instead of having a method for stakeholder participation, having a method for collaborative research (Flipse et al., 2013), or by focusing on practising ethics in R&I as a democratic and reflexive discourse (Genus, 2006), which would imply that democratic principles should inform the process of participation.

Second, stakeholder participation in general suffers from the problem of a top-down approach, according to which stakeholders are confronted a-priori with a framework of principles and concerns. When discussing EIA, Markus and Mentzer criticise the list of values it proposes of being of "only a certain kind", which means that stakeholders might not be able to engage with values they would deem important (Markus and Mentzer, 2014, p. 359). In line with this critique, Borning and Muller question whether VSD should "single out certain values as particularly worthy of consideration" and if so, by whom they should be chosen and how (Borning and Muller, 2012, p. 1129). Again, it can be argued that methods dealing with stakeholder participation face this problem across the board. Ways to deal with this problem are for instance to make value identification dependent on a well-organised participatory process (Bombard et al., 2011) and by taking participatory design "commitments to co-design and power sharing" (Borning and Muller, 2012, p. 1131) into account.

As a final note regarding the above-mentioned problems, it should be stressed that while participations should be taken into account, its limitations when applied to ethical

reflection should also be taken into account as for instance Felt et al. (2008) forcefully show. Through the application of procedures to ensure participation, knowledge might certainly be gained about ethical impacts but one might doubt whether this entails some kind of moral judgement. From a philosophical point of view moral judgements are more than fact-finding and moral conclusions seeking to move from "is to ought" are often misguided. Participation can therefore not in and by itself guide the process of practising ethics in R&I.

Recommendations:

- For methods that deal with stakeholder identification, we recommend that they should include considerations of justified stakeholder selection. These considerations could be based on reasoned criteria or a mapping-framework for stakeholder identification, or could gain from collaborative approaches or approaches guided by democratic principles.
- For methods that facilitate stakeholder participation in the process of practising ethics in R&I, we recommend that they should include considerations that negate a top-down approach. They could do so by shaping the framework for ethical analysis according to a participatory process or by integrating insights from participatory design in process of practising ethics in R&I.

3.3   Framework for a Novel Method

This section outlines a framework for constructing a novel method that can adequately deal with the four over-arching aims and the common challenge in practising ethics in R&I as discussed above. We relate this framework to three major choices that we will need to make: (1) the choice for an ethical framework, (2) the choice for a corresponding framework to consider the role of technological mediation and (3) the choice for a framework for including people in the process of practising ethics. For each of these choices, we will reflect on the way in which they answer to the general recommendations presented above.

3.3.1   Choice for an ethical approach

The first choice we need to make concerning the framework for our novel method for practising ethics in R&I, is the choice for an approach in ethics that is most likely to be responsive to the general recommendations presented above. For this choice, we engage in two considerations. The first pertains to the question of whether we should base our theory on a heuristic, such as a list of principles or values, or instead on a coherent theory in normative

57

ethics, such as deontology. We will argue for the latter route, and offer justifications for this choice. The second consideration pertains to the subsequent choice for an ethical theory, for which we will consider consequentialism, deontology, virtue ethics and care ethics. Eventually, we will choose virtue ethics and justify why we do so.

Heuristic or systematic theory?

Starting with our first consideration, we have to take into account that most of the methods for practising ethics we analysed use some form of a heuristic rather than an ethical theory as the basis for their approach. The use of heuristics is dominant across the ex ante, intra, and ex post methods. For instance, ATE (ex ante) offers an "ethics checklist", which includes values such as "freedom", "privacy" and "human dignity" (Brey, 2012a, p.12), the EIA framework (intra) offers a list of "ethical and social issues" to be considered such as "autonomy" and "informed consent" (Wright, 2014, p.166) and the ethical matrix (ex post) offers a heuristic based on the principlism method for analysing ethical issues, which includes "wellbeing", "autonomy" and "justice" (Mepham, 2000, p.170). Using a heuristic as the basis for our ethical framework would therefore seems to be a starting point that deserves our consideration. However, we will argue that instead we should depart from a systematic theory in normative ethics.

Because there is a lack of critical discussions regarding the use of heuristics in literature on methods for practising ethics in R&I, we turn towards an established discussion in the more mature field of bioethics on the use of "principlism" as a heuristic for ethical reflection. Beauchamp and Childress (2001) developed this approach, which revolves around the formulation and use of four "prima-facie", mid-level ethical principles that should guide ethical reflections in biomedical contexts: the principles of autonomy, beneficence, non-maleficence and justice. Several criticisms have been raised against this approach, such as that the four principles are not used for actual decision-making in clinical contexts (Page, 2012), that the approach lacks grounding in moral philosophy and has no theoretical coherence (Clouser and Gert, 1990), and that principlism is "thick" in terms of its status when it is applied to practical contexts but "thin" in content (Lee, 2010, p. 527). Schöne-seifert considers the different objections to the principlism method and argues that a main concern is that the principles lack "justificatory ground through systematic connectedness" (Schöne-seifert, 2006, p. 115). Especially this concern is one that we want to raise against the current use of heuristics in methods for practising ethics in R&I. The lists mentioned above offer very unsystematic groupings of ethical "values", "principles" and "issues", without

analytically distinguishing these. For instance, what the EIA approach labels an "ethical issue" (e.g. privacy), is labelled a "value" by the ATE approach, whereby the distinction between "issue" and "value" is obscured. Moreover, it is unclear how a "value" such as "autonomy" is related to "human dignity" or "freedom", as part of the same heuristic.

Furthermore, the uses of heuristics such as "checklists" have a number of additional shortcomings when being applied to R&I processes. As Kiran et al. (2015) argue, uses of heuristics can be highly problematic because a conventional list of values does not reflect the context of technological innovation in which values are shaped by technology design and perhaps even "new" (or newly relevant) values might arise out of emerging technological contexts. More importantly, we argue that most of the approaches using heuristics lack a major feature of principlism by not being unified. In principlism, each principle implies the others (e.g. justice can only be realised through non-maleficence). However, this is certainly not the case for many heuristics we discussed, such as the one used by the EIA approach in which e.g. "profiling and social sorting" is not implied in "dignity". This creates the serious issue of heuristics often having an arbitrary character, of them being collections of ethical values, principles or issues that were deemed relevant by the author(s) of the collections themselves. No systematic theory or approach can consequently be used to justify why a certain heuristic is included in a list and another left out. We need to stress at this point that we do not aim to disqualify the use of heuristics altogether, but only that we argue against the use of heuristics as the sole basis for a method for practising ethics in R&I. As we will show in the later sections, heuristics can co-exist with an ethical theory when they are grounded in such a theory[22]. From this decision, we are led to a second strategic choice: of choosing a systematic normative theory.

Which ethical theory?

Moving to our second consideration, we are to discuss which systematic theories in normative ethics might be chosen as ethical frameworks for our novel method, and which theory would be most suitable as our point of departure. We will consider the theories of consequentialism, deontology, virtue ethics and care ethics. Eventually, we will choose virtue ethics as the ethical framework as point of departure for our novel method and justify this choice.

---

[22] See also Clarke (2009) for an argument why systematic theories in normative ethics are still essential to be considered in conjunction with a heuristic such as the one presented by principlism and Polansky and Cimakasky (2015) for an argument that shows how the four principles, when considered as virtues, can be compatible with virtue ethics.

First, we consider **consequentialism**. Consequentialism can be understood as "the set of moral theories that make the good explanatorily primary, explaining other moral notions, such as duty or virtue, in terms of promoting value" (Brink, 2005, p. 381). It is therefore the *good consequences* of an action, for instance an action out of duty or out of virtue, on which we should base our moral assessment. Within the framework of consequentialism different approaches exist, which express different considerations of what is valuable. We can distinguish between ethical egoism, which looks at whether actions are beneficial for the agent performing them; ethical altruism, which considers whether actions are beneficial for everyone but the agent performing them; and utilitarianism, which considers whether actions are beneficial for everyone. Notwithstanding the different types of consequentialism, Parfit characterises the central claim of consequentialism (with which all its theories agree) as follows: that (1) there is one ultimate moral aim, that outcomes be as good as possible; (2) that applied to acts (i) what one ought to do is whatever would make the outcome best, and (ii) if one does what one believes will make the outcome worse, one is acting wrongly (Parfit, 1984, p. 29). Considered in the context of R&I, a consequentialist assessment could for instance be as follows: the developments of human enhancement technologies such as cognitive prosthetics make the outcomes best in any case (in terms of maximising pleasure for a greater amount of time, i.e. in terms of extended lifespan). Each of us should support the development of human enhancement technologies, and those who act against the development of these technologies are acting wrongly.

Second, we consider **deontology**. Moral theories grouped under deontology are generally characterised by their non-teleological character: they deny the primacy of goodness or badness of an action's consequences in determining whether an action is right or wrong, but instead posit other criteria (Gaus, 2001). Such a criterion can for instance be the divine will of God in case of the Ten Commandments, which stipulate that killing another human being is wrong. Based on certain criteria, deontological theories aim to make explicit what types of actions ought to be required, forbidden or permitted. Kantian ethics is considered the most prominent version of a deontological theory. Kant formulated his renowned categorical imperative as the formulation of a command of reason, with the contents: "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (Kant, 2002, p. 37). Acts that fail to comply with the categorical imperative should be deemed morally wrong (e.g. one cannot lie and simultaneously will that to lie becomes a universal law, which makes lying morally wrong). Kant has strongly influenced other deontological theories. Rawls formulated the hypothetical

initial situation as a constraint on the consideration of just actions (Rawls, 1971, p. 11) and Habermas formulated a dialogical principle for moral impartiality, which constrains the validity of a moral norm (Habermas, 1998, p. 42). In an assessment of R&I outcomes, a deontological argument could have the following form. Big data systems used for policing civilians have certain inherent biases (reflecting existing biases in society). Placing oneself behind the veil of ignorance, in which one doesn't know about ones place in society, leads one to believe that the perpetuation of biases, which disadvantages certain people, would not be in one's rational self interest and thus undesirable. Therefore, the use of big data systems used for policing civilians should be limited and fairer alternatives should be developed.

Third, we consider **virtue ethics**. Virtue ethics captures a number of moral theories that focus on the goodness of a moral agent according to an assessment of her virtues, which are dispositions to act, exercised through an agent's practical reasoning (Annas, 2005, p. 516). The virtues constitute a person's moral character and aim at what the ancient Greeks coined *eudaimonia*, or human flourishing. The cultivation of virtues is another central part of virtue ethics theories. Such theories stipulate that humans engage in different habitual practices (such as having a conversation), to which certain standards of excellence can be assigned (which determine whether one has a good or a bad conversation), which in turn allow for the cultivation of certain virtues (e.g. honesty). For instance, through having initial conversations with her parents, a child gets to know how to converse in the right way – by not telling lies for instance. An assessment of R&I according to virtue ethics would typically assess the practices that are engaged in and could have the following form. Some social media shape the practice of conversation in such a way that people are incentivised to construct their lives in a way contrary to their own everyday experiences. As Vallor indicates, the carefully edited "streams of personal and career triumphs" don't reflect the inconsistencies of everyday reality, thereby artificially "inflating" our sense of personal worth (Vallor, 2016, p. 174). This can damage a person's character, for instance by obstructing the practice of "having a honest conversation with ones friends". Excessive social media use by early practitioners, such as children, should therefore be strongly reconsidered.

Fourth, we consider **care ethics**. Theories captured under the heading care ethics are relatively recent and are also designated by labels such as the ethics of love and relational ethics (Held, 2005, p. 537). Even though it is not entirely clear whether care ethics can be considered a comprehensive theoretical approach, a number of features can be distinguished (Held, 2005). First, it has a focus on the agent attending to the needs of a particular other agent, for whom responsibility is taken. Second, emotion is taken into account as a crucial

factor in the epistemology of morality (i.e. emotions contribute to our "knowing" of what is morally good or bad). Third, care ethics privileges the particular and situated over abstract and universal moral claims. Fourth, it aims to rethink the often-traditional notions of public and private, moralising for instance the private sphere of the home. Fifth, a unifying aspect of care ethics is its conception of personhood, which is thoroughly relational and interdependent. As Gilligan found, "the moral person is one who helps others" (Gilligan, 1993, p. 63). An example of an assessment of R&I according to care ethics would be the following. The implementation of a certain robotic system to lift patients in health care settings prevents nurses from touching their patients, because they control the machine instead of directly attending to the patient. (Van Wynsberghe, 2013). This would negate the relational bond between the caregiver and the patient, and would therefore constitute a questionable practice from the perspective of care ethics.

Why virtue ethics?

As a concluding step, we need to decide which ethical theory to select as a point of departure and provide a justification for this choice. We argue that virtue ethics is the most promising approach because, in line with our recommendations, it allows for (1) dealing with the uncertainly of technological change in a way that does not (solely) rely on prediction and (2) offers a promising way for thinking about the role of everyday practices to ethically assess the role of technologies in R&I settings. However, we will also explore some remaining limitations of virtue ethics.

First, when considering the ex ante stage in R&I we face the problem of uncertainty of technological change. For instance, during the early years of certain nanotechnology research (e.g. strains of molecular nanotechnology research), no concrete applications of the technology can yet be thought of that would have a predictable impact on the world. In such a situation, multiple courses of action would be possible. An ethicist arguing within a consequentialist framework could still try to foresee and predict the manifold of technological innovations and corresponding potential future scenarios. However, as Vallor argues, the "sociotechnical opacity" that results from the uncertainty of technological change will be a constraint on our any ethical framework that cannot be overcome (Vallor, 2016, p. 10). Another issue is that *systematic* ethical impacts of particular technological innovations are similarly difficult to predict. Technological innovations can create unexpected side effects such as climate change, financial crises, mass-scale privacy infringements and global pollution. In line with MacIntyre, we can relate this to the challenge posed originally by

Machiavelli of *fortuna*: the fundamental unpredictability in human affairs (MacIntyre, 2007, p. 93). When reasoning from a consequentialist perspective, these outcomes will not appear in any argument as long as they remain unpredictable (think of an argument for or against the building of factories during the industrial revolution, when climate change was still unthought-of). Especially in cases of emerging technologies that are radically "open" (i.e. are not yet considered in conjunction with any concrete applications), our observations seem to be a good reason to start looking beyond the means of prediction in dealing with emerging technologies. In order to find an alternative, Vallor asks what kind of ethical framework offers us the best *strategy* to be able to deal with sociotechnical opacity, and argues that this would be virtue ethics. She argues for virtue ethics because it stipulates that moral expertise (or the ability to act virtuously) does not result from following general moral principles based on anticipated consequences but instead is reflected in the capacity to generate such principles (Vallor, 2016, p. 24). As such, virtue ethics tries to stipulate how humans can cultivate their characters in such a way that they will be more able to deal with morally challenging situations, such as acting under the condition of sociotechnical opacity. In the ex ante stage of R&I, ethicists might focus on what virtues should be cultivated (e.g. prudence, self-control) in order for researchers and policy makers to be able to deal with sociotechnical opacity in the best possible way; for instance by instituting certain ways of schooling, training programmes or raising awareness of ethical impacts of existing technological applications. In short, while virtue ethics does not offer a *solution* to the problems posed by the unpredictability of technological change in consequentialist approaches, it offers a promising alternative approach.

Second, we argue that virtue ethics is a more promising starting point than deontology because of the different answers these theories provide to the question of how to formulate proper interventions, based on an ethical challenge. Deontological approaches typically intervene by formulating certain requirements, restrictions or permissions: by regulating particular activities. In certain instances, this can be very fruitful, for example in the case of regulating data use through data protection protocols in ICT research. That is, deontological approaches predominantly offer ways to *regulate* ethics in R&I. However, we focus on *practising* rather than on regulating ethics. This means that in R&I contexts, we are interested in ways of integrating practises related to ethics with or alongside the concrete work researchers engage in on a day-to-day basis. In a broader context, which included potential use and application in a societal setting, we are interested in the practises different users of technologies engage in. As a starting point, we therefore want to adopt an approach that

offers an understanding of the role of everyday practices in the shaping of our ethics. Such an understanding would for instance help us examine the research practices of an engineer working under pressure of increasing his company's market share as happened with the Volkswagen emission scandal (Ewing, 2017). For this purpose, virtue ethics is suitable because its main focus point is the role of everyday practices in the cultivation of virtues such as honesty or self-control. Another reason for considering virtue ethics instead of deontological approaches is the complexity of contemporary technical practices. As Vallor concedes, technical practices such as "tweeting" are too complex to be understood using a formal system of general requirements, restrictions and permissions (Vallor, 2016, p. 27). Sending a tweet has many intricate impacts in many different contexts (e.g. it might trigger a bidding mechanism for data brokers to display a proper advertisement in response, it might invoke political anger by a distant reader following a certain hash tag, it might invoke happiness for a good friend). Therefore, we should not categorically allow or disallow the act of tweeting for no general rule would be sufficient for covering all particular contexts in which a tweet can have an impact. In many cases tweeting might be allowed while in some cases tweeting might be disallowed, as was the case with Microsoft's "racist" tweetbot Tay (Gibbs, 2016). We have to note that we do not conclude that deontological considerations will altogether be absent from our novel method. As will become clear in chapter 6, deontological notions such as distributive justice will be part of our method for practising ethics, but only within a framework that has virtue ethics as its starting point.

Third, we choose virtue ethics over care ethics because as Halwani (2017) shows, care ethics has a number of crucial problems that are resolved when care is subsumed in virtue ethics, as Vallor (2016, p. 138) does in her work. Focusing on the primacy of care in personal relations, care ethics faces criticisms that (1) it cannot deal with situations in which someone cares for an evil person, because it does not provide grounds for relinquishing care when running the risk of moral corruption and (2) it does not provide grounds for our obligations towards non-intimate others such as strangers because partiality is considered integral to ethics and no grounds for diverting from it are provided (Halwani, 2017). Virtue ethics can to some extent deal with these problems because it allows for a consideration of personal care as well as impersonal justice (considering a person's practical wisdom within a certain context, in which the virtues of care and justice can play a role) (Halwani, 2017, p. 184). To this we add that care ethics does not offer an adequate way to evaluate the role of technological systems, because it focuses on intimate care-settings and pays little attention to the role of technologies in public life. This creates the problem that a method based on care ethics will

have too small of a scope, by leaving out some of the crucial technological systems that impact our everyday existence. Virtue ethics, as we will show later on, allows for the assessment of sociotechnical systems by focusing on the way in which they mediate our technical practices (e.g. making explicit how financial systems mediate the signing of electronic loan agreements or certain payment habits). Taken together, we have sufficient grounds for taking **virtue ethics** as point of departure for constructing our novel method for practising ethics in R&I.

However, we need to acknowledge some important remaining limitations of virtue ethics. First, with regards to uncertainty of technological change it should be noted that virtue ethics only provides for a response insofar this change does not surpass the limits of human capacity. Philosophers such as Bostrom (2013) argue that technological changes might pose existential threats that humans do not have the inherent capacity for – morally and intellectually - to properly address. This argument poses at least a challenge for virtue ethics. Namely, because the virtues are bounded to the human being as such they are ultimately limited, which means that humans might need to be 'enhanced' to be capable of dealing with ethical challenges that are thus far too complex. Second, even though virtue ethics is better capable of dealing with the problem of evil than care ethics (by incorporating considerations of justice), it does not yet resolve it. That is, acting in accordance with virtue and tending to a certain good can in a pluralist society still lead to conflicts resulting in evil. As Wall argues, acting freely in aiming at the good stands in tension with the historical finitude of human life - the possibility of accidents and unfortunate choices - which can lead to "hubristic domination over the world" or "violence toward the stranger" (Wall, 2005, p. 9). For instance, even though Antigone in Sophocles' play aimed at the good in respecting her family bonds, her actions nonetheless contributed to a tragic course of events (Ricoeur, 1992, p. 245). Virtue ethics is ill equipped do deal with the problem of evil because it only offers a positive account of human action, not a negative one (i.e. accounting for the grounds for limiting human action). Third, a limitation of virtue ethics is that it offers an account of human development but not of an evaluation of human conduct that might for instance be the basis for legislation. As Louden (1984) argues, virtue ethics offers no ground for formally assessing *who* is virtuous, for instance by observing certain behaviour. As such, it does not offer a proper grounds for regulation, for formulating rules to distinguish between those who act in accordance with vice (e.g. criminals) and those who act in accordance with virtue.

In constructing our novel method, we will to some extent deal with these three limitations. We will address the limited character of human capacity by incorporating an

institutional dimension in our framework. Institutions allow for the durability of human action and extend its capacity, for instance by setting up structures for human collaboration that surpass the limits of the individual's intellectual and moral capacities. The other two limitations will be addressed through Ricoeur's efforts to embed Kantian formalism in Aristotelian teleology. As we will discuss in section 6.4, Kant's deontology adds the necessary consideration for limiting human practice to deal with the problem of evil, and also offers a grounds for the formulation of formal rules. Thus, even though virtue ethics will be our starting point, our approach will eventually also incorporate central aspects of deontology.

3.3.2 Choice for a framework for understanding the role of technology

Now we have chosen a theory in ethics, we need to consider ways in which to understand the role of technology in shaping our life world that we attend to when practising ethics in R&I. As we saw in the previous chapter, a consideration of the role of technology is crucial for all methods we analysed. For ex ante methods, we need to consider how emerging technologies might influence our life world in the future. For intra methods, we need to consider how technologies can be designed in such a way that they "embed" human values. For ex post methods, we need to consider how technologies have impacted our life world in order to properly assess their "ethical impacts". We justify the need for a theory of technology by echoing Ihde's concern that any intellectual or ordinary understanding of technology is necessarily based on basic and often conflicting presuppositions concerning technology, for instance instrumentalist or determinist presuppositions (Ihde, 1979, p. 3). This means that there is no "fixed", coherent and ordinary point of departure when considering the role of technology, which consequentially means that without a theory of technology there is no proper way of thinking about notions such as the "embedding" of values in design. Even though a theory of technology does not "fix" the meaning of technology, it provides a coherent way of thinking about it.

For our current purposes, we will discuss a number of prominent approaches that provide ways to theorise the role technologies play in our lifeworld. Even though we will at a later point also discuss the ideas of a number of "classical" thinkers of technology (such as Heidegger), we will now stick to more contemporary approaches that can be associated with the "empirical turn" (Achterhuis, 2001). Notwithstanding the great diversity of approaches, we limit ourselves to the most prominent ones in the field of philosophy of technology: philosophy of society, based on Searle's analytic philosophy, postphenomenology and social

constructivist approaches in STS. After having considered these different approaches and having discussed their shortcomings, we will argue that we need to construct a novel approach that combines some aspects of the above-mentioned approaches but also introduces the crucial notion of *narrative*. We will consequently justify this choice, which will be developed into a full-fledged theory of technology in chapter 5.

Which theory of technology?[23]

Different theoretical approaches have been developed during the previous decades aimed at improving our understanding of the role of technologies in our lifeworld and consequently of the technical practices and outcomes in R&I contexts. In order to cover the spectrum of these different approaches, we start by discussing a particularly "analytic" theory (philosophy of society), followed by a "continental" approach (postphenomenology)[24]. Subsequently, we also discuss social constructivist approaches in STS.

John Searle's **philosophy of society** offers a theory of social reality that states that all human made phenomena, ranging from streets to governments to technologies in general, share a linguistic basis. The origin of certain artificial phenomena, or rather institutional facts, is traced back to linguistic entities called "status function declarations" (Searle, 2010, p. 13). An example of a simple status function declaration is: "I hereby declare that the provided information is true". By agreeing with such a statement in an ICT-mediated setting, the linguistic act of agreeing (the speech act) results in a new reality (Searle, 2006, p. 69): it provides the agreeing party with a new set of digital rights and duties, of constitutive rules, that define the ontology of the respective ICT environment. Status function declarations include both locutionary aspects (linguistic aspects, propositions) and illocutionary aspects (extra-linguistic aspects: intentional states like a beliefs and desires). They are characterised by what Searle calls a "double direction of fit": a notion that refers to the fit between the locutionary, propositional aspect of the declaration and the human directedness to the world implied by the illocutionary aspect (Searle, 2010, p. 12). For declarations, two different illocutionary aspects coincide: (1) the desire to make something the case and (2) the belief to make something the case. In other words, if we declare something to be the case, we are able to create a social reality while desiring it to come about. For example, when a certain person

---

[23] Parts of this section have been adopted from: Reijers, W., & Coeckelbergh, M. (2016). The Blockchain as a Narrative Technology: Investigating the Social Ontology and Normative Configurations of Cryptocurrencies. Philosophy & Technology, 7. http://doi.org/10.1007/s13347-016-0239-x

[24] Although the divide between the "analytic" and the "continental" traditions in philosophy is controversial and contested, it does reflect a distinction in foci (Dolcini, 2007)

declares to become the President of the United States, the propositional form of the declaration "I, Barack Obama, hereby declare that…", fits with the collective desire to bring about a new state of affairs implying a new ontological reality (the new president of the United States). When we apply Searle's theoretical model to capture the ontology of technologies, we could state that technologies indeed can be understood as status function declarations. They are constituted by declarations because their use is dependent on a linguistic, propositional structure that is such that it allows them to bring about their own reality. Moreover, they are constituted by status function declarations because their meaning depends on a coinciding structure of human desires and beliefs. For instance, when using our Internet banking system (a technology) to transfer money, we believe the new state of affairs (a transaction), which coincides, with our desire to bring it about (we wanted the transaction to occur). These desires and beliefs do not belong to the individual but to a collective. We collectively intend status function declarations to become part of our social reality. In other words, the individual act of transacting an amount of digital money[25] depends on the collective intentionality that amounts to the validity of this act.

However, we argue that this framework does not seem to lead to an adequate understanding of the socio-linguistic grounding of technologies. Two main lacunas make Searle's theory inadequate to serve as a solid basis for the examination of the role of technologies in our life world. First of all, Searle leaves the gap between individual intentionality and collective intentionality unexplained, merely stating that collective intentions are biologically primitive phenomena: intentionality in the "we" mode instead of in the "I" mode. By suggesting this reductionist view, he disqualifies the impact of culture that is precisely not reducible to human biology (Heidemann, 1999, p. 259). Since we are particularly interested in understanding how individual intentionality is culturally mediated to arrive at the collective consensus constituted by technologies in use, we are in need of a theoretical framework that does account for the interrelation between individual and collective intentionality. Secondly, Searle's theory does not include an aspect of normativity that is needed to explain why declarations can have a status function at all (Heidemann, 1999, p. 260). In the case of technologies such as Internet banking systems, we would want to explain why we assign a status function to them. In more common terms, we would want to explain why people assign value (not just economic value, but also emotional and political values) to digital numbers on screens when they engage in Internet banking. This is not a

---

[25] For a more elaborate treatment of the ontology of digital money, using Searle's philosophy of society, see the *Critique of Digital Money* (Reijers, 2014)

trivial point, for the meaning of technologies in use depends on their relation to human normative values.

In order to overcome the problematic aspects of Searle's philosophy of society, we turn towards alternative theories in philosophy of technology and in STS. Initially, we consider **postphenomenological** theories of technology that focus on the role of technological mediation. As Ihde concedes, instead of understanding technologies as formations of formal rules, we should aim to understand the way in which their materiality shapes our experience of the world (Ihde, 2009). We might for instance say that a technology such as glasses constitute an "embodied" experience of the world or, as Verbeek puts it, that technologies such as scientific instruments make our objects of experience "present in a specific way" (Verbeek, 2005, p. 141). Consequently, we can analyse a technology by conceptualising the kind of relationship it constitutes between a human and its life world. Postphenomenology offers four types of basic relations: embodiment (e.g. glasses being experienced as part of the body), hermeneutic (e.g. reading a thermometer to know the temperature), background (e.g. being environmentally aware of the air-conditioning) and alterity (e.g. treating an ATM as a "quasi-other" technological agent) relations. One of the major advantages of postphenomenology is that it offers a normative framework for understanding how technologies shape human morality (Verbeek, 2006), because it makes explicit how technologies constitute new relations between humans and their lifeworld that can have moral significance. For instance, an ultrasound scan constitutes a new relation between parents and the phenomenon of the foetus, which is made "present" in a specific way and as such carries with it ethical impacts such as a redefined "responsibility" for the health of foetus in the womb.

However, the focus on the material aspects of technologies in postphenomenological theories covers-up the important role of linguistic and symbolic mediations (Coeckelbergh, 2015) that is at stake when aiming to understand the mediating role of many types of technologies (such as Internet banking systems). For instance, postphenomenology fails to provide an account of how certain cultural artefacts that influence our understanding of technology (such as Orwell's *1984*) play a role in the way in which technologies mediate our perception (e.g. while looking at a monitor that is connected to CCTV cameras). Moreover, the focus of these theories on the individual mediation captured by the human-technology-world relationship fails to provide for an explanation of how "being-with-each-other", social relationships (Van Den Eede, 2010) are shaped by technologies. Therefore, the conceptual leap from technological mediation at the individual to the collective level remains as

problematic as in Searle's social ontology. For instance, postphenomenology cannot adequately explain how the individual act of retrieving cash money from an ATM (an alterity relation) is related to the collective intentionality on which the monetary system depends.

Alternatively, we consider approaches in **STS** that unlike postphenomenological theories are more focused on the role of language, by capturing technological mediation using notions of "de-scribing" "scripts" of technological objects (Akrich, 1992) or the "interpretative flexibility" of artefacts (Pinch & Bijker, 1984). Works in STS focus on the relevance of the mediation of the collective: mapping networks of social groups or of human and non-human actors. For instance, Jasanoff argued that "socio-technical imaginaries", which are "collectively held, institutionally stabilised, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology" (Jasanoff, 2015, p. 6), shape technology design. These socio-technical imaginaries can incorporate accounts of technological artefacts in popular fiction, but also institutional narratives of nationhood and citizenship. On a slightly different note, expanding on works such as Winner's analysis of the politics of architectural structures (Winner 1980), Edwards argues that socio-technical infrastructures are designed according to "mutual orientation" of normative goals of both small social groups and large institutions towards a design of a socio-technical system (Edwards, 2004, p. 22). Dawson and Buchanen articulate an informative perspective on this kind of approaches in STS, arguing that technology change happens through the outcomes of competing *narratives* (Dawson & Buchanan, 2005), a process of interaction between interpretations of technologies that would be characterised by Pinch and Bijker as moving towards "closure" (Pinch and Bijker, 1987, p. 44). In a similar vein, Joerges (1999) argues, against Winner, that narratives play an important role in politics of technologies. Turning to technologies in R&I contexts, we could state that people construct narratives about artefacts, which are related to the expectations and negotiations various individuals and organisations hold. As such, development of technologies are indicative of a politics, understood as interactions between social discourses and social imaginaries.

As the foregoing illustrates, STS approaches focus predominantly on interpretations and narratives *about* technology. As Pinch and Bijker explain, a distinct social group might have a certain interpretation of a technology that subsequently influences a discourse between different actors (relevant social groups). This focus on narratives about technologies as elements of a social discourse puts human agency somewhat in the forefront. Even though it is argued that a design is "co-produced" (Jasanoff, 2015, p. 16) or that artefacts, like words,

are also tools of politics (Joerges 1999); it is co-produced by different groups of people. It therefore seems that these approaches insufficiently answer to the valuable lessons gained from postphenomenological research, namely that humans and technologies co-shape reality: that a certain agency has to be ascribed to technologies as well (Verbeek, 2005, p. 112). Moreover, it seems that the narratives themselves are viewed as pre-given and that the technology design is a consequence of interaction between these narratives. For instance, different social groups are said to have different interpretations of a technology and the design is changed accordingly. Conversely, in line with Searle, we should say that technological structures, as consisting of systems of status function declarations, do not just incorporate interpretations or delegate scripts, but *create* or *constitute* new social realities. Therefore, we do not merely want to ask how different interpretations of technologies shape their designs, but also how the technology itself in turn shapes our understanding of the world we live in. Taken together, we summarise below (table 4) why none of the prominent approaches we discussed seems to offer an adequate account of the role of technology in our lifeworld that we might use for our novel method for practising ethics in R&I.

Table 4: Listing of the approaches for understanding technological mediation and their advantages and disadvantages.

Approach	Advantages	Disadvantages
Philosophy of Society	• Provides an account of the relation between technology and language • Links technologies to other aspects of human social reality (institutional facts)	• Explanatory gap between individual and collective intentionality • Neglect of normativity in the shaping of social reality
Postphenomenology	• Provides a framework to analyse particular human-technology-world relations • Provides an account of the ways in which technologies mediate human morality	• Neglects the role of language and linguistic artefacts in conceptualising our understanding of technologies • Does not offer adequate means to understand how collective intentionality is mediated by technologies
STS	• Looks at the way in which language, in terms of discourse and narrative, influences technology development	• Focuses on discourse and narrative *about* technology, leading to a neglect of the mediating role of technology

Approach	Advantages	Disadvantages
	• Provides an account that looks at the role of collective actors (e.g. social groups)	• Narratives are assumed to be pre-given, rather than mediated by technologies in use

Why an alternative approach based on narrative theory?

From the foregoing discussion, it follows that we need an alternative approach to understand how language, collective intentionality and technology are intertwined in the normative shaping of our lifeworld. Going back to the idea that a "competition" between narratives can result in technology change, as Dawson and Buchanen (2005) suggest, we propose to re-frame the problem they present by asking: how do humans and technologies *co-shape* the narrative structures that have the potential to transform our understanding of our technologically mediated social reality? Such an approach might answer to the concern raised by Feenberg (2009, p. 228) about postphenomenological theories of technology and theories in STS, namely that they provide for a convincing relational ontology but insufficiently offer a corresponding hermeneutic theory of meaning (Feenberg, 2009, p. 228). At the same time, we have to resist ideas of technological determinism such as the one proposed by Winner (1980), because our framework should explicitly acknowledge the agency of both humans and technologies in the mediation of social reality. These concerns cause us to turn to narrative theory, and notably to the work of Paul Ricoeur (1983) to which we will fully turn in chapter 5.

Regarding the ontological significance of narrative, there are various philosophical views on how this concept can contribute to our understanding of the social world and on the way in which they shape social reality. Some scholars consider narrative as an instrumental cognitive ability or linguistic tool, whereas others consider it as an ontological category connected to the way humans are in the world (Meretoja, 2014, p. 89) or understand human life itself as having a narrative character (Macintyre, 2007, p. 114) Another theoretical division with regard to the role of narrative exists between an empiricist tradition that denounces narrative as a fundamental philosophical concept (e.g. see Strawson, 2004) and a hermeneutic tradition that instead rejects the idea of experience unmediated by narratives. The latter tradition holds that all representations of the human social world are mediated by human-linguistic interpretation (Taylor, 1971, p. 4), that subjectivity is always mediated by language, "by 'signs, symbols and texts'" (Meretoja, 2014, p. 96). Ricoeur belongs to the philosophical tradition that conceptualises narrative as being deeply connected to human

lives, as a fundamental mediator of human social existence. Thus, rather than viewing narrative merely as an instrumental or discursive tool that shapes the design process of technologies, we argue that it should be understood as a fundamental ontological aspect of human social reality.

Beyond philosophy, multiple scholars have shown how a narrative ontology can improve our understanding of concrete aspects of our social world. For instance, Bruner explains how we can increase our understanding of human psychology, as embedded in a cultural context, by looking at how a "text affects the reader" (Bruner, 1986, p.4). As such, he claims that narratives can "make events" and even "make history" (p. 42). In a similar vein, Czarniawsma employs a methodology for organisation studies that allows for an understanding of the "reflexive nature" of the human condition as the basis for collective action (Czarniawsma, 1998, p. 77). She understands an organisation as a story, understood as a social construct that is shaped by human interaction and interpretation through narratives. Gotham and Staples show how the significance of narrative goes beyond our understanding of history and can help analyse "human agency in processual, action-oriented ways" (Gotham and Staples, 1996, p. 492). As such, it is argued that a narrative understanding of our social world would improve sociological inquiries. These and many other scholars have shown how a narrative ontology can be fruitfully employed in studies of different aspects of our social world. At a later point, we will build on this idea, by exploring how narrative theory, focusing on the work of Ricoeur, can assist us in studying the role of technologies in our lifeworld. Taken together, we conclude that none of the existing approaches we discussed (philosophy of society, postphenomenology and STS) offer a satisfactory understanding of the role of technology in our lifeworld, and that we need to construct an **alternative approach** that takes the notion of **narrative** as its starting point.

3.3.3   Choice for an approach for stakeholder participation

So far, we have chosen virtue ethics together with a theory of technology departing from the notion of narrative as our starting points for constructing a novel method to address the three over-arching aims that were presented earlier in this chapter. However, we still need to consider ways in which we can address the more practically oriented common challenge, of appropriate stakeholder participation in the process of practising ethics in R&I. Stakeholder participation in the R&I process can be understood in three different ways: as aiming (1) to engage individuals that are directly related to the design or development of technologies (e.g. users, designers, researchers, ethicists) (see e.g. Borning et al. 2004), (2) to have different

individuals or groups (e.g. CSOs, companies, government agencies) that represent certain interests participate in the R&I process (see e.g. Blok et al., 2015), and to engage the public at large[26] in the governance of R&I processes (see e.g. Lezau & Soneryd, 2007). The methods we reviewed predominantly focused on the first understanding of stakeholder participation[27]. In what follows, we will first identify some central requirements for the organisation of stakeholder participation that should guide the development of our method. Second, we will present a framework for developing a particular participation tool that will help us to demonstrate the use of our method in R&I practices.

Requirements for stakeholder participation

In order to get a sense of the aspects of stakeholder participation that we should take into account, we will consider some central requirements put forward in the literature included in our review and auxiliary sources in the field of RRI. Unfortunately, most of the methods we previously discussed (section 3.2) do not offer suitable, concrete guidance for *how* to organise participation. That is, many methods mention that participation should happen, but only a few provide an answer to how this should be done. For instance, the ATE method mentions: "it may be recommended, if possible, to solicit participation from different stakeholders" (Brey, 2012b, p. 313) but does not discuss how this could be done. Similarly, the EIA method stipulates that stakeholders should be identified and should be consulted in the process of practising ethics (Wright, 2014, p. 165), but again no ways are proposed for how to do this. Another example is the approach of eTA, which prescribes that processes of practising ethics should be opened up for "public participation and influence" and that "relevant stakeholders" should be involved (Palm and Hansson, 2006, p. 550), but also remains silent as to how to accomplish this. Due to this lack of discussions of stakeholder participation, we will therefore have to consult literature beyond that included in our literature review.

First, we will consider stakeholder participation in terms of involving different individuals and groups with particular interests and to some extent the public at large in the R&I process. This topic is extensively discussed in the RRI literature, which emphasises the aspect of governance (e.g. ways of impacting the decision-making process). An example of

---

[26] We should note, however, that members of "the public" should be distinguished from "stakeholders" (Lezaun & Soneryd, 2007, p. 292), because the latter have delineated opinions and interests whereas the former do not.
[27] This is probably due to the focus of these methods on technological innovation, which puts the emphasis on the research of novel concepts, the design of novel technologies and the impact of these technologies in terms of use. Broader issues such as those pertaining to the governance of the introduction of innovations in society (e.g. considering the climate impacts of new types of fuels) are therefore often not addressed.

stakeholder participation along these lines concerns CSOs dealing with health or with environmental issues participating in the commercial development of technological innovations in the food industry (Blok et al., 2015). Another example, focusing on involving the public, is the organisation of a public debate on genetic modification in the UK with the aim of producing citizens' feedback for government policy (Lezaum & Soneryd, 2007). These forms of stakeholder participation focus on the *governance* of the R&I process in such a way that innovations are embedded in society according to a process of deliberation (i.e. subject to public debate and consultation of stakeholders) and responsiveness (i.e. subject to governance mechanisms such as democratic decision-making procedures) (Owen et al., 2008, p. 38). Moreover, participation is organised in such a way that stakeholders are jointly able to address ethical "grand challenges", such as climate change (Schomberg, 2013).

For practising ethics in R&I, therefore, stakeholder participation will need to be incorporated in such a way that it (1) opens up debate concerning the impact of technological innovations, (2) includes mechanisms for democratic governance, and (3) allows for societal actors to work together to address grand challenges. However, at the same time the limitations of these three requirements need to be acknowledged and incorporated. With regards to the first, (a) it should be acknowledged that debate always involves a tension between the public (the laypeople) and stakeholders with a clear interest (Lezaun & Soneryd). Concerning the second, (b) it should be noted that in governance mechanisms there is a tension between moving towards consensus and contestation, between the possibility to find common grounds and to lack common grounds (Blok, 2018). With regards to the third, (c) we should acknowledge that participants in a process of participation have different power positions and often promote different interests, goals, and possibilities for change (Blok et al., 2015). As will be clarified in section 6.4.1, these requirements call for an institutional dimension of a novel method for practising ethics in R&I, which includes considerations of conflict resolution and democratic governance.

Second, we turn to considerations of stakeholder participation understood as directly engaging types of individuals in the design process. A promising approach for understanding how stakeholder participation could be organised in this regard is VSD (Friedman et al., 2006). This approach engages with *participatory design* (Borning et al., 2004), which offers ways to have stakeholders participate in R&I processes. However, VSD in its original form puts questions such as "how are stakeholders affected" under the rubric of conceptual investigations, which means that an ethicist is supposed to "identify benefits and harms" (Friedman et al., 2006, p. 88) of stakeholder groups, instead of having the stakeholders

themselves participate in the process to allow them to indicate what the potential harms might be. To find a response to this problem, we turn to the "tools and techniques" that are offered by approaches in *participatory design* (Brandt et al., 2013). These tools and techniques include "telling activities" (e.g. sharing narratives about the impacts of a technology), "co-creation" (contributing to prototyping, using generative tools for co-creation), "enacting scenarios" (staging use in real-use contexts) and using "design games" (having a rule-governed setting to explore possible uses of a technology).

Approaches in participatory design offer these different tools and techniques according to a number of central principles. Kensing and Greenbaum (2013) offer six general principles for participatory design. First, it should aim at equalising power relations, which means that those participating in a design process should be on an equal footing regarding the input they can provide. Second, it should aim at democratic practices, meaning that educated and engaged participants act both on their own interests and on a notion of the common good. Third, it should lead to situation-based actions, meaning that tools for participatory design should be usable in concrete everyday context, which in our case predominantly implies that they should be usable in concrete R&I practices. Fourth, it should lead to mutual learning, meaning that experts learn from non-experts and vice-versa. Fifth, it should not merely be a theoretical approach but has to be translated into practical tools to be rendered useful. Sixth, it should allow for alternative visions about technology, meaning that different perspectives on a particular technology should be incorporated in the design process. We will take these six principles of participatory design as requirements for the way in which we will organise stakeholder participation in our novel method.

To do this, we should, however, also consider the compatibility between participatory design and the other choices made for the framework leading to our novel method. We argue that the tools and techniques in participatory design are very much aligned with narrative theory, because many of them relate to the use of narratives ("telling" narratives and "enacting" them) to facilitate the participation process. Moreover, Winner has argued for the use of participatory design in ethics of technology and has connected this with the cultivation of "virtues of citizen participation" (Winner, 1995, p. 81). He argues that through participation in R&I processes, stakeholders engage in a community where "making things" (design) "and taking action" (practising ethics) are one and the same. He concedes that this engagement might cultivate the virtues of participating stakeholders, by developing a form of practical wisdom in their dealings with technologies.

Method for a participatory tool

Stakeholder participation is a dimension of practising ethics in R&I that does not merely relate to reflection (e.g. reflecting on the ethical impacts of technologies) but also to practical organisation (e.g. getting people in a room to discuss or work together). For this reason, we believe that it is insufficient to merely provide the outlines of possible ways to organise participation (e.g. prescribing ways of democratic decision-making), but that we additionally should also put "participation" *in* concrete practice. Next to formulating the requirements for stakeholder participation, we therefore also started to develop a practical tool that would facilitate practising ethics in a concrete R&I context, namely for R&I projects in the ADAPT Centre[28], both in R&I teams and in classroom settings. Because the process of developing the Ethics Canvas started in parallel with the construction of the literature review and the novel method, it is aligned with the core ideas in this thesis but not fully embedded in the literature it uses. For instance, it does not yet align with our methodology for practising ethics in R&I derived from the work of Ricoeur. Instead, in order to create the Ethics Canvas, we have pragmatically selected approaches in business model development (cf. Lucassen et al., 2012) for organising stakeholder participation in accordance with principles of participatory design (Gregory, 2003). This selection was pragmatic insofar our team of developers[29] was already familiar with applied ethics on the one hand and business development on the other. Nevertheless, we justify this general orientation towards business modelling and participatory design approaches by pointing at (1) their use in a non-exhaustive variety of R&I contexts, (2) their focus on iterative deployment throughout the R&I process, (3) their use in practical contexts (they organise discussion and participation as integral parts of R&I practices) and (4) their focus on the participation of non-experts (they focus not just on experts - which would be ethicists in our method - but on many types of stakeholders).

Taking the six requirements for the organisation of stakeholder participation into account, we can consider a particular participatory tool that would be useful in translating our novel method to concrete R&I practices. The methods that we discussed in the foregoing sections only offer only a few resources for enabling R&I practitioners to participate. The most promising approaches in this regard are the ETHICS approach (Mumford, 1995), which explicitly discusses how participation should be organised in R&I settings (systems design in particular) and the Ethical Matrix (Mepham, 2000), which provides a visual tool that

---

[28] "The ADAPT Centre"…"focuses on developing next generation digital technologies that transform how people communicate by helping to analyse, personalise and deliver digital data more effectively for businesses and individuals". Taken from https://www.adaptcentre.ie/about on 28-11-2018.
[29] Wessel Reijers, David Lewis, Arturo Calvo, Killian Levacher, Harshvardhan Pandit.

researchers can use to discuss how a technology impacts different types of stakeholders according to a number of ethical principles (taken from the princiflism approach). The ETHICS approach is promising because it is closely aligned with the workplace of the R&I practitioner and the Ethical Matrix is promising because it offers a easy-to-use visual tool that can be used by researchers to reflect on their work on an iterative basis. However, while learning from the strengths of these approaches we will not adopt them directly into our novel method because the ETHICS approach has the shortcoming of being based on weak conceptual basis coming from ethical theory (Stahl, 2007, p. 487) and the Ethical Matrix suffers from the problems of being unable to weigh the ethical problems it uncovers and the potential instigation of conflicts between stakeholders.

For our purposes, we wanted to find an approach that is (1) closely aligned with concrete research practices in the workplace of R&I practitioners in alignment with the principles of participatory design, (2) easy-to-use in order to be capable of iterative deployment in these practices and (3) as much as possible related to our theory of technology that is rooted in the notion of narrative. First, this meant that our approach should preferably not be alien to R&I practices but already in some way used in many different R&I contexts (i.e. in different disciplinary fields). Second, this meant that our approach should not depend on substantial prior knowledge (e.g. knowledge of theories in normative ethics), should not demand heavy documentation and should be capable of being used dynamically in the R&I process (e.g. on a monthly basis by a research team, to iteratively reflect on the team's work to disclose potential ethical impacts). Third, this meant that our approach should have some affinity with narrative theory and should allow participation based on shared narratives rather than on following rules (e.g. checking compliance with regulations).

Even though we lack an exhaustive overview of the different participatory tools that fulfil the above-mentioned requirements, we found the promising tool called "the Business Model Canvas" (BMC), originally developed by Osterwalder and Pigneur (2010), which facilitates and structures a participatory process for the construction of a business model. The BMC consists of a visual canvas with different boxes that capture core elements of a business model, such as customer segments and value proposition. It has been widely adopted within business settings, but crucially also within R&I settings (e.g. ICT start-ups but also ICT departments in universities). This means that the approach is already closely aligned with practices in the workplace. It also accords with the second requirement because, as Lucassen et al. argue (2012, p. 14), "it effectively models explicit information of both tangible and intangible aspects of the business and communicates this information in a highly accessible

manner to parties unfamiliar with the modelling technique". Finally, it shows affinities with narrative theory, which is confirmed by Kuparinen's characterisation of the BMC as a "narrative business model" (Kuparinen, 2012, i). Participants (e.g. researchers working on an R&I project) are triggered to engage in a discussion of different narrative scenarios related to the boxes on the canvas. Certainly, the BMC has little in common with approaches for practising ethics in R&I because it focuses on a very different problem (creating a business case, rather than disclosing potential ethical impacts). Nevertheless, we can refocus the design of the BMC to make it accord with our aim of using it for organising participation in the disclosure stage. We will discuss how this can be done in chapter 7.

3.4   Chapter Summary

In this chapter, we discussed shortcomings of the existing methods for practising ethics in R&I, formulated recommendations based on these shortcomings, and based on the recommendations constructed our framework for a novel method. Eventually, we decided to take (1) virtue ethics of technology in conjunction with a (2) theory of technology based on the notion of narrative as starting points for our theoretical framework that in turn will be (3) translated into a practical, participatory approach, which led us to consider the BMC as a suitable participatory tool.

In chapters 4, 5, 6 and 7, we will construct a novel approach for practising ethics in R&I based on the framework we laid out in chapter 3. At the end of chapter 7, we will reflect on the extent to which we have managed to incorporate the three over-arching aims as well as the common challenge of practising ethics in R&I (see section 3.2). In this reflection, we will consider whether the choices we made in constructing the framework adequately incorporate the recommendations we formulated.

3.5   References

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

Akrich, M. (1992). The De-scription of Technical Objects. In W. Bijker & J. Law (Eds.), *Shaping technology/ Building society*. Camebridge, Massachusetts: MIT Press.

Albrechtslund, A. (2007). Ethics and technology design. *Ethics and Information Technology*, 9(1), 63–72. http://doi.org/10.1007/s10676-006-9129-8

Annas, J. (2005). Virtue Ethics. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of Biomedical Ethics*. Oxford: Oxford University Press.

Blok, V., Hoffmans, L. and Wubben, E. F. M. (2015). Stakeholder engagement for responsible innovation in the

private sector: critical issues and management practices, *Journal on Chain and Network Science*, 15(2), 147–164. doi: 10.3920/JCNS2015.x003.

Blok, V. (2018) From participation to interruption: Toward an ethics of stakeholder engagement , Participation and Partnership in CSR and Responsible Innovation. In: Von Schomberg, R., & Hankins, J. *Handbook Responsible Innovation: A Global Resource*, Cheltenham: Edward Elger, pp. 1–22.

Bombard, Y., Abelson, J., Simeonov, D., & Gauvin, F.-P. (2011). Eliciting ethical and social values in health technology assessment: A participatory approach. *Social Science & Medicine*, 73(1), 135–144. http://doi.org/10.1016/j.socscimed.2011.04.017

Borning, A., Friedman, B., & Kahn, P. H. (2004). Designing for Human Values in an Urban Simulation System: Value Sensitive Design and Participatory Design. In *PDC-04 Proceedings of the Participatory Design Conference*, Vol 2, Toronto, Canada (pp. 68–71). Palo Alto.

Borning, A., & Muller, M. (2012). Next steps for value sensitive design. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems* - CHI '12, 1125–1134. http://doi.org/10.1145/2207676.2208560

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management*, *18*(3–4), 285–298. http://doi.org/10.1080/09537320600777002

Brandt, E., Binder, T., & Sanders, E. B.-N. (2013). Tools and Techniques. In J. Simonsen & T. Robertson (Eds.), *Routledge International Handbook of Participatory Design*. London: Routledge.

Brey, P. (2000). Disclosive Computer Ethics: The Exposure and Evaluation of Embedded Normativity in Computer Technology. *Computers and Society*, *30*(4), 10–16.

Brey, P. (2012a). Anticipatory Ethics for Emerging Technologies. *NanoEthics*, *6*(1), 1–13. http://doi.org/10.1007/s11569-012-0141-7

Brey, P. (2012b). Anticipating ethical issues in emerging IT. Ethics and Information Technology, 14, 305–317. http://doi.org/10.1007/s10676-012-9293-y

Brink, D. (2005). Some Forms and Limits of Consequentialism. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.

Bruner, J. (1986). *Actual Minds, Possible Worlds.* Cambridge, Massachusetts: Harvard University Press.

Coeckelbergh, M. (2015). Language and technology: Maps, bridges and pathways. *AI & Society*, Forthcoming, 1–37.

Copp, D. (2005). *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press. http://doi.org/10.1093/0195147790.001.0001

Cotton, M. (2009). Evaluating the "ethical matrix" as a radioactive waste management deliberative decision-support tool. *Environmental Values*, 18(2), 153–176. http://doi.org/10.3197/096327109X438044

Clarke, D. L. (2009). Moral principlism alone is insufficient, and traditional moral theories remain important for practical ethics. *SAJBL, 2*(2), 54–58.

Clouser, K. D., & Gert, B. (1990). A Critique of Principlism. *The Journal of Medecine and Philosophy, 5*, 219–236.

Czarniawsma, B. (1998). *A Narrative Approach to Organization Studies.* Thousand Oaks: Sage Publications.

Dawson, P., & Buchanan, D. (2005). The way it really happened: Competing narratives in the political process of technological change. *Human Relations, 58*(7), 845–865. http://doi.org/10.1177/0018726705057807

Dolcini, N. (2007). The Analytic/Continental Divide: Entities and Being. *Soochow Journal of Philosophical Studies, 16*, 1-14.

Edwards, P. N. (2004). Infrastructure and modernity: force, time, and social organization in the history of sociotechnical systems. *Modernity and Technology*, 1, 185–225. http://doi.org/0262633108

Ewing, J. (2017). Engineering a deception: what led to Volkswagen's diesel scandal. Retrieved October 3, 2017, from https://www.nytimes.com/interactive/2017/business/volkswagen-diesel-emissions-timeline.html

Feenberg, A. (2009). Peter-paul verbeek: Review of what things do. *Human Studies, 32*(2), 225–228. http://doi.org/10.1007/s10746-009-9115-3

Felt, U., Fochler, M., Muller, a., & Strassnig, M. (2008). Unruly ethics: on the difficulties of a bottom-up approach to ethics in the field of genomics. *Public Understanding of Science*, *18*(3), 354–371. http://doi.org/10.1177/0963662507079902

Ferrari, A. (2010). Developments in the debate on nanoethics: Traditional approaches and the need for new kinds of analysis. *NanoEthics*, 4(1), 27–52. http://doi.org/10.1007/s11569-009-0081-z

Flipse, S. M., van der Sanden, M. C. A., & Osseweijer, P. (2013). The Why and How of Enabling the Integration of Social and Ethical Aspects in Research and Development. *Science and Engineering Ethics*, *19*(3), 703–725. http://doi.org/10.1007/s11948-012-9423-2

Friedman, B., Borning, A., Davis, J. L., Gill, B. T., Kahn, P. H., Kriplean, T., & Lin, P. (2008). Laying the Foundations for Public Participation and Value Advocacy: Interaction Design for a Large Scale Urban Simulation. *Design, 1*(2), 305–314. http://doi.org/10.3233/IP-2009-0172

Friedman, B., Kahn Jr., P. H., & Borning, A. (2006). Value Sensitive Design and Information Systems. In K. E. Himma & H. T. Tavani (Eds.), *Human-Computer Interaction and Management Information Systems: Foundations* (pp. 1–27). John Wiley & Sons, Inc. http://doi.org/10.1145/242485.242493

Gaus, G. F. (2001). What is deontology? Part one: Orthodox views. *Journal of Value Inquiry, 35*(1), 27–42. http://doi.org/10.1023/A:1010359613871

Genus, A. (2006). Rethinking constructive technology assessment as democratic, reflective, discourse. *Technological Forecasting and Social Change*, 73(1), 13–26. http://doi.org/10.1016/j.techfore.2005.06.009

Gibbs, S. (2016). Microsoft's racist chatbot returns with drug-smoking Twitter meltdown. Retrieved October 3, 2017, from https://www.theguardian.com/technology/2016/mar/30/microsoft-racist-sexist-chatbot-twitter-drugs

Gilligan, C. (1993). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, Massachusetts: Harvard University Press.

Gotham, K.F., & Staples, W.G. (1996). Narrative Analysis and the New Historical Sociology. *The Sociological Quarterly*, 37(3), 481-501.

Gregory, J. (2003). Scandinavian Approaches to Participatory Design. *International Journal of Engaging Education, 19*(1), 62–74. http://doi.org/http://www.ijee.dit.ie/contents/c190103.html

Groves, C. (2013). Horizons of Care: From Future Imaginaries to Responsible Research and Innovation. In: *Shaping Emerging Technologies: Governance, Innovation, Discourse*, edited by Konrad, K., Coenen, C., Dijkstra, A., Milburn, C., and Van Lente, H., 185–202. Berlin: IOS Press.

Habermas, J. (2000). *The inclusion of the other: Studies in political theory. Political Theory*. Cambridge, Massachusetts: MIT Press.

Heidemann, C. (1999). On Some Difficulties Concerning John Searle's Notion of an "Institutional Fact." *Analyse & Kritik, 20*, 143–158.

Held, V. (2005). The Ethics of Care. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.

Ihde, D. (1979). *Technics and Praxis*. Dordrecht: D. Reidel Publishing Company.

Ihde, D. (1990). *Technology and the Lifeworld*. Bloomington and Indianapolis: Indiana University Press.

Ihde, D. (2009). *Postphenomenology and Technoscience*. New York: Sunny Press.

Jasanoff, S. (2015). Future imperfect: Science, Technology and the Imaginations of Modernity. In S. Jasanoff & S. Kim (Eds.), *Sociotechnical imaginaries and the fabrication of power* (Vol. 1, pp. 1–49). http://doi.org/10.1017/CBO9781107415324.004

Kant, I. (2002). *Groundwork for the Metaphysics of Morals*. (J. B. Schneewind, M. Baron, S. Kagan, & A. W. Wood, Eds.). Educational Philosophy and Theory (Vol. 47). New Haven: Yale University Press.

Kensing, F., & Greenbaum, J. (2013). Heritage: Having a Say. In J. Simonsen & T. Robertson (Eds.), *Routledge International Handbook of Participatory Design*. New York: Routledge.

Kiran, A. H., Oudshoorn, N., & Verbeek, P.-P. (2015). Beyond checklists: toward an ethical-constructive technology assessment. *Journal of Responsible Innovation, 2*(1), 5–19. http://doi.org/10.1080/23299460.2014.992769

Kuparinen, P. (2012). *Business Model Renewal and its Networking Aspects in a Telecom Service Company*. Master Thesis. Tampere: Tampere University of Technology.

Latour, B. (1994). On Technical Mediation - Philosophy, Sociology, Genealogy. *Common Knowledge*, 3(2), 29–64. http://doi.org/10.1111/j.1365-294X.2010.04747.x

Le Dantec, C. A., Poole, E. S. E. S., & Wyche, S. P. S. P. (2009). Values as lived experience: Evolving value sensitive design in support of value discovery. *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*, 1141–1150. http://doi.org/10.1145/1518701.1518875

Lee, J. H. (2010). The problem of "thick in status , thin in content" in Beauchamp and Childress' principlism. *Journal of Medical Ethics, 36*, 525–528. http://doi.org/10.1136/jme.2009.031054

Lezaun, J. and Soneryd, L. (2007). Consulting citizens: Technologies of elicitation and the mobility of publics. *Public Understanding of Science*, 16(3), 279–297. doi: 10.1177/0963662507079371.

Lucassen, G., Brinkkemper, S., Jansen, S., & Handoyo, E. (2012). Comparison of Visual Business Modeling Techniques for Software Companies. In M. A. Cusumano, B. Iyer, & N. Venkatraman (Eds.), Software Business: *Third International Conference, ICSOB 2012*, Cambridge, MA, USA, June 18-20, 2012. Proceedings (pp. 79–93). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-30746-1_7

Lucivero, F., Swierstra, T., & Boenink, M. (2011). Assessing Expectations: Towards a Toolbox for an Ethics of Emerging Technologies. *NanoEthics*, *5*, 129–141. http://doi.org/10.1007/s11569-011-0119-x

MacIntyre, A. (2007). *After Virtue: A study in moral theory* (Third Edit). Notre Dame, Indiana: University of Notre Dame Press. http://doi.org/10.1017/CBO9781107415324.004

Manders-Huits, N. (2011). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*, *17*(2), 271–287. http://doi.org/10.1007/s11948-010-9198-2

Markus, M. L., & Mentzer, K. (2014). Foresight for a responsible future with ICT. *Inf Syst Front*, (January),

353–368. http://doi.org/10.1007/s10796-013-9479-9

Mepham, B. (2000). A Framework for the Ethical Analysis of Novel Foods: The Ethical Matrix. *Journal of Agricultural and Environmental Ethics, 12*(2), 165–176. http://doi.org/10.1023/A:1009542714497

Meretoja, H. (2014). Narrative and Human Existence: Ontology, Epistemology, and Ethics. *New Literary History*, *45*(1), 89–109. http://doi.org/10.1353/nlh.2014.0001

Mumford, E. (1995). *Effective Systems Design and Requirements Analysis - The ETHICS Approach*. London: Macmillan Press Ltd. http://doi.org/10.1017/CBO9781107415324.004

Nordmann, A. (2007). If and then: A critique of speculative nanoethics. *NanoEthics*, 1(1), 31–46. http://doi.org/10.1007/s11569-007-0007-6

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. No publisher. Retrieved from BusinessModelgeneration.com.

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E. and Guston, D. (2013). A Framework for Responsible Innovation. In Owen, R. and Bessan, J. (eds) *Responsible Innovation*. London: John Wiley & Sons, Ltd. doi: 10.1063/1.1764494.

Page, K. (2012). The four principles: can they be measured and do they predict ethical decision making? *BMC Medical Ethics, 13*(1), 1-10. http://doi.org/10.1186/1472-6939-13-10

Palm, E., & Hansson, S. O. (2006). The case for ethical technology assessment (eTA). *Technological Forecasting and Social Change, 73*(5), 543–558. http://doi.org/10.1016/j.techfore.2005.06.002

Pandza, K., & Ellwood, P. (2013). Strategic and ethical foundations for responsible innovation. *Research Policy*, *42*(5), 1112–1125. http://doi.org/10.1016/j.respol.2013.02.007

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artifacts: or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science, 14*(3), 221–232. http://doi.org/10.1177/030631284014003004

Polansky, R., & Cimakasky, J. (2015). Aristotle and Principlism in Bioethics. *Diametros, 45*, 59–70.

Reijers, W. (2014). *Critique of Digital Money* (Master's Thesis), (pp. 1–114). University of Twente. http://doi.org/10.13140/RG.2.1.4747.7288

Schöne-seifert, B. (2006). Danger and Merits of Principlism. In C. Rehmann-Sutter, M. Düwell, & D. Mieth (Eds.), *Bioethics in Cultural Contexts* (pp. 109–119). Heidelberg: Springer.

Schroeder, D., & Palmer, C. a. (2003). Technology assessment and the ethical matrix. *Poiesis & Praxis*, *1*, 295–307. http://doi.org/10.1007/s10202-003-0027-4

Searle, J. R. (2006). Social Ontology: Some Basic Principles. *Anthropological Theory, 6*(1), 51–71.

Searle, J. R. (2010). Making the Social World: The Structure of Human Civilization. Oxford: Oxford University Press. http://doi.org/10.1093/acprof:osobl/9780195396171.001.0001

Shilton, K. (2014). This is an Intervention: Foregrounding and Operationalizing Ethics During Technology Design. In K. D. Pimple (Ed.), *Emerging Pervasive Information and Communication Technologies (PICT)* (pp. 176–192). Heidelberg: Springer. http://doi.org/10.1007/978-94-007-6833-8

Spahn, A. (2015). Mediation in Design for Values. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 251–266). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-6970-0_9

Stahl, B. C. (2007). ETHICS, Morality and Critique: An Essay on Mumford´s Socio-Technical Approach. *Journal of the Association for Information Systems, 8*(9), 479–490. http://doi.org/Article

Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. *NanoEthics*, *1*(1), 3–20. http://doi.org/10.1007/s11569-007-0005-8

Taylor, C. (1971). Interpretation and the Sciences of Man. *The Review of Metaphysics*, *25*(1), 3–51.

Vallor, S. (2016). *Technology and the Virtues*. Oxford: Oxford University Press.

Van Den Eede, Y. (2010). In Between Us: On the Transparency and Opacity of Technological Mediation. *Foundations of Science, 16*(2–3), 139–159. http://doi.org/10.1007/s10699-010-9190-y

Van de Poel, I. (2009). Values in Engineering Design. *Philosophy of Technology and Engineering Sciences* (Vol. 9). Elsevier. http://doi.org/10.1016/B978-0-444-51667-1.50040-9

Van de Poel I. (2013). Translating Values into Design Requirements. In: Michelfelder D., McCarthy N., Goldberg D. (eds) *Philosophy and Engineering: Reflections on Practice, Principles and Process*. Philosophy of Engineering and Technology, vol 15. Springer, Dordrecht

Van Gorp, A. (2005). *Ethical issues in engineering design: Safety and sustainability*. Simon Stevin Series in the Philosophy of Technology. PhD Thesis, University of Delft.

Van Wynsberghe, A. (2013). Designing robots for care: care centered value-sensitive design. *Science and Engineering Ethics, 19*(2), 407–33. http://doi.org/10.1007/s11948-011-9343-6

Von Schomberg, R. (2013). A vision of responsible innovation. In Owen, R., Heintz, M., and Bessant, J. (Eds.), *Responsible Innovation*. London: John Wiley. doi: 10.1002/9781118551424.ch3.

Verbeek, P.-P. (2005). *What things do; philosophical reflections on technology, agency, and design*. Pennsylvania: Pennsylvania University Press.

Verbeek, P.-P. (2006). Materializing Morality. *Science, Technology & Human Values, 31*(3), 361–380. http://doi.org/10.1177/0162243905285847

Verbeek, P.-P. (2008) Morality in Design: Design Ethics and the Morality of Technological Artifacts. In: Peter Kroes, P., Vermaas, P.E., Light, A., Moore, S. A. (Eds.), *Philosophy and Design*. Springer, Dordrecht

Wall, J. (2005). *Moral Creativity - Paul Ricoeur and the Poetics of Possibility*. Oxford: Oxford University Press.

Wenstøp, F., & Koppang, H. (2009). On operations research and value conflicts. *Omega*, 37(6), 1109–1120. http://doi.org/10.1016/j.omega.2008.10.001

Winner, L. (1995). Citizen Virtues in a Technological Order. In A. Feenberg & A. Hannay (Eds.), *Technology and the Politics of Knowledge* (pp. 65–84). Bloomington: Indiana University Press.

Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136. http://doi.org/10.2307/20024652

Wright, D. (2014). Ethical impact assessment. *Ethics, Science, Technology, and Engineering, 163*(c), 163–167. http://doi.org/10.1016/j.clsr.2011.11.007.Wright

Zwart, S. D., van de Poel, I., van Mil, H., & Brumsen, M. (2006). A network approach for distinguishing ethical issues in research and development. *Science & Engineering Ethics*, *12*(4), 663–684.

4 Outlining Virtue Ethics of Technology

4.1 Introduction

In this chapter, we explore the use and limitations of virtue ethics of technology as the ethical theory that forms the basis of our novel method. The "revival" of virtue ethics - understood as a renewed interest in Aristotelian ethics in particular - has produced a great amount of scholarly works in the past decades. However, virtue ethics has only recently been introduced in the writings on ethics of technology. Charles Ess (2009) is one of the first scholars who sketched the outlines of a virtue ethics of technology and others have used virtue ethics in their ethical reflections on technology (cf. Bolsin et al., 2005; Coeckelbergh, 2012). After these first explorations, Vallor's recent book *Technology and the Virtues* can be regarded as the first systematic and comprehensive account of a virtue ethics of technology approach (Vallor, 2016). Virtue ethics has been a welcome addition to ethics of technology, because it offers an alternative, fruitful response to tensions in utilitarian and duty-based approaches that are very dominant in the field. Additionally, as Vallor argues, virtue ethics does not offer straightforward *solutions* to complex ethical issues as many approaches try to do, but instead provides a *strategy* for "cultivating the type of moral character that can aid us in coping, and even flourishing" (Vallor, 2016, p. 10) with and under the challenging conditions of contemporary emerging technologies, such as surveillance technologies and human enhancement technologies.

Recently, virtue ethics has also become a topic of interest in the RRI literature, though in this context it is still scarcely discussed. For instance, Meljgaard et al. (2018) discuss how the Aristotelian concept of phronesis could guide efforts to educate researchers. They argue that by focusing on the cultivation of practical wisdom, educators could incorporate reflective aspects in their teaching of RRI and include external actors (the *phronimoi*) in the teaching process. This ought to incline R&I practitioners to connect their efforts with praxis, through caring and committing to contributing to the improvement of different practices. In a similar vein but looking at a different issue, namely of wicked problems that can arise in the context of innovation (e.g. climate change), Blok et al. argue that MacIntyre's virtue ethics offers a valuable perspective for considering the types of "virtuous competencies" (Blok et al., 2006, p. 23). They argue that by engaging at activities that aim at internal sustainable goods

together with other stakeholders, practitioners would gain a better capacity to respond to the challenge of wicked problems.

Notwithstanding the important progress that has been made, we argue that virtue ethics as an ethics of technology is still incomplete for our purposes. First, it attributes moral agency almost exclusively to humans by emphasising how *we* should cultivate *our* values in order to cope with the ethical challenges of emerging technologies. This largely disregards the valuable lessons learnt in recent works in philosophy of technology (cf. Ihde, 2009; Verbeek, 2005) and STS (cf. Latour & Porter, 1993; Pinch & Bijker, 1984) that show how humans and technologies co-shape reality: how technologies themselves contribute to the cultivation of virtues and how therefore "practices", "narratives" and "the good life" can only be understood within a context of technological mediation. We argue that in order to fully understand how humans and technologies "co-shape" or "co-cultivate" the virtues, we need an account of technological mediation that provides for the necessary conceptual framework to understand how this process takes place. Second, in its current form virtue ethics of technology does not offer a distinct theory of "technomoral" virtue. For this reason, we reconsider MacIntyre's work that offers a theory of practice as a *unified* account of virtue. In line with MacIntyre (2007), we ask: how do technologies shape technical practices, the narratives in which these practices are embedded and accordingly an overall conception of the good life?

In order to provide the basis for our novel approach and to construct an argument that will lead us to complement our ethical theory with an appropriate theory of technology, we engage in three steps. First, we briefly present the contemporary field of virtue ethics and the central questions it puts forward. Second, we discuss the recent introduction of virtue ethics in the field of ethics of technology and notably the novel insights that Vallor has provided. Third, we provide an elaborate critique of Vallor's work, present a case for the missing notion of technological mediation, reconsider MacIntyre's theory of practice and discuss its limitations.

4.2 A Very Short Introduction to Virtue Ethics

Virtue ethics builds on the idea that human agents possess certain consistent dispositions or states of character, called the virtues, which inclines them to engage in right action in a particular context. It does not focus on the evaluation of particular actions, but on evaluating the agent's character. The seminal paper that has been responsible for the revival of Aristotelian virtue ethics in the Western world is G.E.M. Anscombe's *Modern Moral*

Philosophy, in which the author states that when we have the outlines of a philosophy of psychology, in which notions such as "action, intention, wanting and pleasure" can be accounted for, we might *eventually* start advancing the concept of "virtue" as the basis for a form of ethics (Anscombe, 1958, p. 12). Even though Anscombe did not herself present a theory of virtue ethics and was reluctant to adopt the Aristotelian system due to its perceived lack of clarity (notably regarding concepts such as eudaimonia or "human flourishing"), a novel approach in moral philosophy emerged that started to develop the Aristotelian notion of virtue, with Philippa Foot (1978), John McDowell (1979), Rosalind Hursthouse (1999), and Alasdair MacIntyre (2007) as its main proponents. What these scholars have in common is not only their focus on virtue, but - in line with Anscombe - also a strong disagreement with the established Kantian and utilitarian traditions in moral philosophy. Yet, Martha Nussbaum (1999) contests whether virtue ethics genuinely offers a *distinct,* or incommensurable approach that can be contrasted with Kantian ethics and utilitarianism, because thinkers in these traditions also assign importance to the notion of virtue in their theories. We tentatively agree with her assessment, but also argue that virtue ethics at least presents a *refocusing* or correction of modern moral philosophy by a return to the notion of virtue as developed in ancient Greek ethical theory. Accordingly, a notable difference between virtue ethics and other approaches in moral philosophy is the *emphasis* on the agent as virtuous or vicious rather than on the rightness or wrongness of an action.

In the literature on virtue ethics, the following central questions are posited: (1) *what* is virtue; (2) *which* (cardinal or principal) virtues can we distinguish; and (3) *how* are virtues "cultivated"? A human being has certain dispositions that we commonly say belong to her "character". Some of these can be regarded as "general beneficial characteristics" (Foot, 1900, p. 2), which can be conceptualised as "virtues". It is the possessing of virtues, or rather of a certain kind of "virtuous being" that determines the goodness of a moral agent. The virtues themselves are socially cultivated, rather than naturally acquired, but do depend on a teleological notion of the good life for human beings (the human *ergon,* or "function") (NE, 1.7, 1098a15-20, trans. Irwin). This teleological notion is captured by the concept of *eudaimonia,* which cannot be easily translated into the more familiar concept of happiness (which might be understood as a pleasurable state), but rather designates the end of a human life as a whole that needs no qualification. While other ends in a human life, such as health, can have a qualification (e.g. we are healthy to live in accord with eudaimonia), eudaimonia does not have any further qualification (e.g. it does not make sense to state that we live in accord with eudaimonia to be healthy). Accordingly, a virtue needs to be a disposition that is

acted upon in accordance with eudaimonia and that *endures* rather than one that might be deployed in a single action or that may lay dormant in an individual. An agent can therefore be evaluated according to an evaluation of enduring dispositions to act in certain way in particular circumstances and not according to an evaluation of a single action (e.g. one apparently kind act does not make its agent kind). Contrary to Kantian ethics and utilitarianism, which argue from the idea of ethical principles towards an idea of the good life, virtue ethics holds that acting on the virtues in line with eudaimonia yield rules of action (Hursthouse, 1999, p. 39). For instance, virtue ethics maintains that by engaging in just practices we become just people and therefore acquire a notion of rules that just practices adhere to (e.g. a sense of procedural justice).

Based on a certain conception of what virtue is, distinct versions of virtue ethics offer different heuristics or lists of virtues. MacIntyre discusses types of these heuristics, such as the lists of Christian virtues, Homeric virtues and Benjamin Franklin's utilitarian virtues (MacIntyre, 2007, p. 185). Unlike the heuristics we discussed in the previous chapter, a heuristic in a certain version of virtue ethics stands for a distinct, systematic theory of virtue (MacIntyre, 2007, p. 183). Aristotle articulated the most influential list of virtues in Western philosophy. He distinguished between virtues of character and virtues of thought, which relate to one-another. Virtues of character belong to the part of the soul that is not rational itself but nonetheless can "share in reason" (NE, 1.13, 1102a14-18, trans. Irwin). In other words, these are virtues that allow a person to have certain emotional dispositions or rather *feelings* such as fear, confidence, pleasure and pain, in the right way that is mediated by the right reasons. A way to understand such a virtue is by considering the mean within two types of emotional extremes: of excess and of deficiency. For instance, an excess of fear leads to cowardry while a deficiency of fear leads to recklessness. The virtue of courage represents the mean disposition between these extreme emotions, and is mediated by rational deliberation – notably by the intellectual virtue of *phronesis* or prudence. Amongst the virtues of character we can count courage, temperance, generosity, magnanimity, mildness, truthfulness, wit, friendship and justice. Virtues of thought are dispositions to act that are fully rational (NE, 1.13, 1103a1-5, trans. Irwin), belonging to the human intellect. In other words, these virtues do not mediate our feelings, but aspects of reason. The virtues of thought are craft knowledge, scientific knowledge, prudence, wisdom and understanding. Prudence, or *phronesis*, which is a disposition to grasp the truth concerned with (a particular) action about what is good or bad for the human function, takes a central position because it connects

all the virtues by linking practice with reflection. As such, prudence is a required virtue for all the other virtues to be cultivated and acted upon.

Additionally, theories in virtue ethics make explicit *how* virtues are cultivated, or how a human being becomes virtuous. MacIntyre is perhaps the scholar who has constructed the most widely adopted theoretical basis for answering this question. He even takes the position, one that we will adopt hereafter, that the "how" question is the one that unifies different, seemingly incommensurable theories of virtue (MacIntyre, 2007, p. 186). His framework, in conjunction with Vallor's, will be central to the discussions in the remainder of this chapter. Virtue, for MacIntyre, is "an acquired human quality the possession and exercise of which tends to enable us to achieve those goods which are internal to practices and the lack of which effectively prevents us from achieving any such goods" (MacIntyre, 2007, p. 191). He argues that to understand the logical development of the concept of virtue we need to pay attention to the different *stages* in which virtues are cultivated. First, there is the stage of *practice,* in which a human being needs to engage in order to cultivate her virtuous being. MacIntyre explains that a practice should be understood as a "socially established human activity through which goods internal to that activity are realised" (MacIntyre, 2007, p. 187). Goods internal to a practice are those related to the desire or wish to excel or improve in the respective practice. An example would be practicing a musical instrument in virtue of improving the mastery of playing it. As such, internal goods are opposed to external goods, which are reasons for engaging in an activity that are external to it. An example would be working to acquire money, because a monetary reward is external to the actual activity of work. The notion of practice used by MacIntyre relates to one of the three senses of *praxis* that Irwin discusses in his authoritative translation of Aristotle's *Nicomachean Ethics* (Aristotle, 1999, p. 315). First, praxis can be used broadly for all voluntary, intentional actions. This type of praxis includes those actions performed to harness the necessities in life (e.g. nutrients). In its second, narrower sense, praxis can be rational action based on a decision. This relates to the realm of crafts, or technê. Finally, praxis can have a third, most narrow meaning of rational action, which is an end in its own right. It is such instances of praxis, designated as "practices" by MacIntyre, that enable an agent to cultivate her or his virtues.

Second, according to MacIntyre practices through which virtues are cultivated should be understood as being embedded in a narrative order of a human life. This relates to the notion that virtues are not related to individual actions, but to the purpose of a human life as a whole (Aristotle's eudaimonia). MacIntyre argues that this purpose is made *intelligible* by means of

a narrative order, which provides a life with a history in which practices and the cultivation of virtue can be understood. He states that narratives provide us with proper answers to the question: "what is she or he doing?" (MacIntyre, 2007, p. 206). We use narratives to situate a practice in a certain social setting, for instance the practice of "marrying" in the social setting that includes two people's life histories and narratives relating them to their friends and family, the witnesses of and participants in the specific practice. Similarly, intentions for actions and engagement in certain practices can be explained through the construction of narratives. For instance, the writing of a philosophy book can be rendered intelligible by constructing a narrative history of a philosopher's professional career, including an account of the development of his ambitions and ideals. As such, narratives provide some kind of an epistemological correlate to the ontology of situated practices. An important consequence of MacIntyre's turn to narrative is the novel characterisation of an agent as not merely an actor, but also as an author (MacIntyre, 2007, p. 213). We will return to this point, specifically in relation to the use of technologies, in the next chapter.

Third, practices that cultivate virtuous being and that are embedded in a narrative order of a single human life should be understood in the broader context of what MacIntyre designates as a "moral tradition" or community. A moral tradition can be understood as the system of particular moral practices and narratives that aim at an overarching idea of the "good life" and shape the virtues it promotes, such as the Christian moral community, the Confucian moral community or a moral community shaped by Greco-Roman ethics. In line with this idea, MacIntyre emphasises that different moral traditions offer different lists of virtues that in term embody different theories of virtue. The notion of "tradition" is needed in order to surpass the single life of an individual with its own narrative order and extend it to the community within which this life can be evaluated as having been lived in accordance with eudaimonia. In light of this argument, MacIntyre claims that every moral philosophy such as virtue ethics has a particular sociology, or theory of the social world, as its counterpart (MacIntyre, 2007, p. 225). To summarise MacIntyre's theory of virtue, which will remain with us throughout the next chapters:

- Socially established human activities, or **practices**, that realise internal goods, allow human agents to cultivate their **virtues**.
- Practices are intelligible to these agents due to a **narrative** order that encompasses a human life as a whole, rendering agents both actors and authors.

- A human life and the related social established notion of happiness in a human life are embedded within a living **moral tradition**.

4.3 Technology and the Virtues

Having provided a sketch of virtue ethics, we move to the contemporary use of this ethical theory in our reflections on technology. Shannon Vallor offers a first comprehensive ethical theory of technology that adopts the virtue ethics approach. Of the earlier mentioned scholars (Anscombe, Foot, McDowell, Hursthouse and MacIntyre) in virtue ethics, Vallor's framework seems to be most closely related to that of MacIntyre, since she argues that the main aim of her book is to offer "an ethical *strategy* for cultivating the type of moral character that can aid us in coping, and even flourishing" (Vallor, 2016, p. 10 - original emphasis) under the challenging conditions of emerging technologies. MacIntyre's framework is also the only one next to Aristotelian, Buddhist and Confucian frameworks that she explicitly acknowledges and adopts. To arrive at this point, the leading question of Vallor's book is to figure out "what we will *do* with these technologies once we have them, and what they will do with *us*" (Vallor, 2016, p. 5 - orginal emphasis).

In constructing her theory of virtue ethics of technology, Vallor argues that she contributes to the existing virtue ethics tradition in three significant ways. First, she strengthens the criticism raised against utilitarian and Kantian ethics by conceptualising "acute technosocial opacity" (Vallor, 2016, p. 6) that renders us unable to make decisions based on universal rules or foreseen consequences of technologically mediated actions. Second, she constructs her framework by drawing from Aristotelian, Confucian and Buddhist ethics, by which she manages to propose a virtue ethics of technology that can potentially be embraced by different cultures - perhaps even globally - and is therefore less affected by the presumed Western bias that a purely Aristotelian ethics. Because of the global character of emerging technologies, most notably global information and communication technologies, this is a valuable aspect of Vallor's theory. Third, she offers a detailed account of the practice of cultivation of virtue and a list of "technomoral"[30] virtues, both of which are responsive to the challenges posed by contemporary emerging technologies.

---

[30] We will not adopt Vallor's notion of "technomoral virtues" but refer to "virtues" instead, since we argue that virtues are in each case cultivated by practices that are mediated by technologies. This renders the adjective "*techno*moral" superfluous. Additionally, the emphasis on the term "moral" introduces the problematic of this term's connotation with "obligation" and "cultural norms", which disregards the distinction between ethics and morality, as discussed for instance by Ricoeur (1992, p. 170)

Vallor's theory consists of three main parts: one that discusses *how* virtues are cultivated through technical practices, one that discusses *which* "technomoral" virtues should be cultivated in order to deal with acute technosocial opacity, and one that discusses specific case studies of paradigmatic contemporary technology that make use of her theoretical framework. In the first part, Vallor presents seven "core elements" or perhaps rather "conditions" of the practices that cultivate the "technomoral" virtues[31]. First, these practices should be engaged with *habitually*: meaning that one repeats them, guided by moral exemplars and eventually gets accustomed to them in a positive sense (Vallor, 2016, p. 74). Second, these practices should be engaged with from within a relational understanding: meaning that they should be understood in the context of our relations with other members of the moral community, making them responsive to the particular aspects of these relations (Vallor, 2016, p. 83). Third, these practices should be engaged with in a context of continuous reflective self-examination: meaning that the agent while engaged in practice reflects on her weaknesses and faults, both pertaining to the generality of the human being as well as to her particular character (Vallor, 2016, p. 90). Fourth, these practices should be guided by a sense of "intentional self-direction": meaning that the agent positions her practices within a moral will to attain moral goodness (or rather, to keep striving for eudaimonia) (Vallor, 2016, p. 97). Fifth, these practices lead to the cultivation of virtues by means of a habit of "moral attention": meaning that the agent is attentive to problematic moral situations both in a cognitive sense (knowing why a situation is morally problematic) and through engaging in an appropriate emotional response (Vallor, 2016, p. 104). Sixth, these practices require a sense of prudential judgement: meaning that the subject has to be able to choose well amongst a range of alternative courses of action given a particular situation (Vallor, 2016, p. 109). Seventh, these practices require an appropriate extension of moral concern: meaning that virtuous practices should be directed at "the right beings, at the right time, to the right degree, and in the right manner" (Vallor, 2016, p. 117). Vallor makes a distinction between the first four elements and the latter three, arguing that the first four *enable* practical wisdom (*phronesis* or prudence for Aristotle) and the latter three *complete* or conclude it.

In the second part of her work, Vallor presents a list of virtues that she argues answers the *"what"* of "technomoral" virtue, and thereby complements the *how* of the cultivation of

---

[31] Our presentation of the seven core elements of the practices for the cultivation of technomoral virtue is a very brief summary of the lengthy discussion presented by Vallor. Therefore, some important aspects of each of these core elements might be missing in our account. However, the general picture arising from this discussion should at least provide for a reliable overview of the aspects of each core element that are crucial for our later analysis.

virtue as shaped by the seven core conditions of practices. However, she answers a specific type of "what" question, namely "what, or which virtues can we distinguish?" In answering this question, she draws from the three virtue ethics traditions she includes in her approach, Aristotelian, Confucian and Buddhist, and synthesises them within the context of a reflection on the impacts of emerging technologies. The following list of virtues results from her deliberation (Curzer, 2017; Vallor, 2016):

- **Honesty**, related to need to address the changing nature of interpersonal communication through digital media.

- **Self-control**, related to the need to deal with the increasing range of temptations offered through digital media.

- **Humility,** related to the need to balance our tendency to under- or over-estimate the benefits and harms of emerging technologies.

- **Justice,** related to the need to address the increasing unjust distribution of resources and power through technological communication channels.

- **Courage,** related to the need to overcome the challenges humans face of existential proportions (e.g. climate change) that correspond with technological progress.

- **Empathy,** related to the need to deal with the influence of digital culture on our interpersonal relations (e.g. violent discussions on social media).

- **Care,** related to the need to properly balance impacts of technologies on human intimacy (e.g. through the introduction of care robots).

- **Civility,** related to need to respond to individualisation that comes along with emerging technologies and the neglect of public life.

- **Flexibility,** related to the interdependence between moral cultures caused by technologies.

- **Perspective,** related to the need for a better understanding of the impacts of emerging technologies.

- **Magnanimity,** related to the need for leadership in dealing with the impacts of technologies.

- **Wisdom,** related to the need to be knowledgeable of the technological world in dealing with its impacts.

Vallor shows convincingly how these virtues are relevant in the 21st century, in a context of increasing and ubiquitous impacts of technologies. However, as Curzer indicates, she less

convincingly defends these virtues as somehow typically "technomoral", or uniquely applicable to our current technological condition (Curzer, 2017). Thus, even though these virtues are *relevant,* they might not be *typical* for our current age. Moreover, her list of virtues does not seem to correspond with a *distinct* ethical theory, but rather largely accords with Aristotle's theory of virtue. As Mcrae argues, Aristotle's theory of virtue would have been sufficient to provide Vallor with the list of "technomoral" virtues, and it seems insufficiently clear how Buddhist virtue ethics has a crucial influence on this list (Mcrae, 2017). Nevertheless, Vallor provides an account of the virtues that at least is shown to be relevant for technological use-contexts in the 21st century and that could be embraced by different virtue ethics traditions. For this reason, and because the construction of a heuristic of virtues falls outside of the scope of our current endeavour, we adopt Vallor's list of virtues, with some notable adaptations:

- **Wisdom** will be rephrased as **prudence**, in accordance with the Aristotelian concept of phronesis, which Vallor actually refers to when she discusses wisdom. Phronesis is in the leading translations of Aristotle's work translated as prudence, and wisdom translates into another virtue of thought: of *sophia*. Hence, we align ourselves with the leading translations, in order not to confuse prudence as *phronesis* with the Aristotelian virtue of wisdom, *sophia*.

- **Empathy** will be rephrased as **friendship**, in accordance with the criticism that empathy itself should *not* be considered a virtue because (i) considered as a voluntary disposition it is a skill and (ii) considered as an involuntary disposition it is a capacity (Battaly, 2011). We therefore retain Aristotle's related virtue of friendship.

- We will sustain the Aristotelian distinction between virtues of character and virtues of thought, which implies that the missing virtues of thought are added (next to prudence: **craft knowledge, scientific knowledge, wisdom and understanding**).

In the third part of her work, Vallor presents a number of illuminating case studies of paradigmatic contemporary technologies (social media, surveillance technology, robotics, human enhancement technology) and analyses these using her virtue ethics of technology approach. We will not elaborately delve into each of these case studies, but instead give a brief account of *how* these are conducted. First, Vallor introduces each case by means of a short historical narrative about the technology in question and relates it to relevant scholarly critiques. Second, she discusses each technology by means of some relevant technomoral

virtues. Third, she offers solutions based on the discussion, usually in terms of what virtues should be cultivated in what way. To illustrate: for her analysis of social media (Vallor, 2016, pp. 159-187) she first discusses the rise of social media in contemporary society and some of the problems it brought according to a number of scholars, for instance anxiety and loneliness in children and privacy concerns. Second, she analyses social media by discussing the virtues of self-control (social media being very addictive), empathy (social media causing less face-to-face contact and attention), humility, honesty and perspective (social media causing distortions of information) and civility (social media individualising the common good). Thirdly, she presents three ways of dealing with the problems she describes: (1) paying attention to the cultivation of character and not merely to progressing technology, (2) creating better spaces for technomoral education and (3) recognising and promoting individuals and groups who show leadership in promoting technomoral virtues.

Considering Vallor's theory and case studies, we can make a preliminary assessment of the use of her approach for constructing a method for practising ethics in R&I. According to the recommendations we formulated in the previous chapter, we can firstly state that virtue ethics of technology provides a fruitful alternative to other approaches that focus on emerging technologies by focusing on our abilities to deal with potential impacts rather than on the prediction of these impacts. Secondly, Vallor's virtue ethics of technology offers a resolution to value-conflicts because she does not take values in relation to actions but takes the virtuous character of the agent as the starting point for moral deliberation, which is the reason why virtue ethicists argue that the issue of value conflicts does not affect virtue ethics (cf. Hursthouse, 2006, p. 108). However, Vallor's theory does not offer ways for considering how her approach could be used in everyday R&I practices, for instance for a researcher who could ask: how might the design of my technology impact the virtues of its users? Notably, this becomes clear when we consider the types of solutions Vallor puts forward in her case studies. These are generally related to *managing* the design and use of technology. Policy makers can for instance use them to start cultivating the virtues of technology users by providing better forms of education and industry representatives can use them to find "exemplary" individuals or groups that can help promote the right kinds of virtues. However, for *practising* ethics in the context of R&I no complete guidance is provided by Vallor's theory. This is related to her lack of engagement with the notion of technological mediation, which we will elaborate on in the next section.

4.4 The Need for an Account of Technological Mediation

It would be convenient to simply adopt Vallor's virtue ethics of technology and use it as the basis for constructing a novel method for practising ethics in R&I. However, we will raise two main criticisms that will lead us to construct a complementary approach of technological mediation in the next chapter. These criticisms are that (1) Vallor focuses on a one-sided conception of agency, thereby paying insufficient heed to the notion of technological mediation and (2) she does not provide a distinct theory of the cultivation of "technomoral" virtue, leading us to reconsider MacIntyre's theory of practice. To support our argument, we will discuss a widely debated case study originating from work in moral psychology around the "Milgram experiment" that is frequently interpreted as an attack on virtue ethics.

4.4.1 The absence of technological mediation

Our first criticism is that Vallor presents a one-sided notion of agency. Even though she positions her framework as being in line with the "empirical turn" (Vallor, 2016, p. 31) of philosophy and ethics of technology (Achterhuis, 2001) she does not pay sufficient heed to one of the central concepts of this body of work: namely the concept of technological mediation. On the one hand, Vallor argues that one of the aims of her framework is to offer an account of what *we* can *do* in order to cultivate our technomoral virtues by virtue of which we will be able to cope with technosocial opacity (Vallor, 2016, p. 6). This aim, we concede, is attained to a large extent by means of the theory she constructs, notably by stipulating what types of virtues we should cultivate in order to be able to deal with technological change. On the other hand, however, she claims to provide a convincing account of what emerging technologies *do* to *us*, which is a claim that we argue is not reflected in the contents of her theory. This aim resonates with the idea proposed by scholars of technological mediation, foremost perhaps reflected in Verbeek's book "What Things *Do*" (Verbeek, 2005 - emphasis added). As such, human virtues, understood as dispositions[32], as well as the practices by which humans cultivate their virtues are mediated by technologies. Understanding the process by which technological mediation occurs is therefore crucial for understanding how

---

[32] "Dispositions" and "traits" seem to be handled as synonyms in writings on virtue ethics. However, etymologically, they have distinct meanings. "Trait" originates from the Latin verb "trahere", which means to draw (along) and thereby indicates the position of a drawer with regard to that which is drawn (or: someone projecting a trait on something, saying (on inspection) "this knife has the trait of sharpness"). "Disposition", however, originates from the Latin verb "ponere", which simply means to place, to put (to have a disposition is to tend to place oneself, to put oneself in a certain way) (Partridge, 2006). Disposition, therefore, belongs more clearly to the agency of a subject (being disposed to *x*), while a trait belongs more clearly to the knowledge of an observer (knowing something to have a certain trait). As a conclusion, we will refrain from using the term "traits" when discussing the virtues and use the term "dispositions" instead.

not only humans can cultivate their virtues, but how humans and technologies can "co-cultivate" the virtues.

We illustrate the problem of the missing notion of technological mediation by focusing on a point in Vallor's work at which she defends her virtue-based approach against criticisms of virtue ethics coming from moral psychology. She discusses the argument that is related to the famous Milgram experiments in which "research subjects were asked to 'punish' a screaming 'victim' with realistic (but simulated) shocks at the polite request of an experimenter" (Merritt, Doris, & Harman, 2010). Even though the "victim" was an actor who had to pretend that he was hurt by the simulated electric shocks, the experimental setup was sufficiently convincing for the participants to be conceived as being real. The overall outcome of this experiment was that 33 out of 40 research subjects continued the experiment after initial protests of the victim and that 26 of them continued through agonised screaming and a final unresponsive silence. Moral psychologists argue that experiments like this one show "that the difference between good conduct and bad appears to reside in the *situation* more than in the person" (Merritt et al., 2010, p. 357 - emphasis added). This is conceived as an attack on the notion of human character that is central to virtue ethics, because our character allegedly should enable us to act morally right in a consistent manner. One can start to criticise this interpretation of the findings by arguing that the concept of "situation" is black-boxed in this approach and that the particular context of the experiment is left un-examined. However, Vallor takes another approach to defend virtue ethics, arguing that her approach holds because virtue is "*by definition* exemplary rather than typical" (Vallor, 2016, p. 22). The fact that a minority of research subjects refrained from following the requests of the experimenter is seen as decisive for acknowledging the correctness of virtue ethics. In other words, the agency of the human subjects who resisted the request for engaging in the experiment is taken as a reflection of a categorical[33] form of their strength of character that in turn *proves* the validity of a virtue ethics approach.

This defence seems to be unsatisfactory for a number of reasons. First, the claim that virtue is cultivated by means of following the practices of exemplary figures does not entail that virtue is "by definition" exemplary rather than typical. Instead, virtue ethics traditions allow for different degrees in virtuous character (hence, having a degree of virtuous character *is* typical), in which exemplary members of a certain moral community (the *phronimos,* in

---

[33] Their virtues character is to be interpreted categorically because, for Vallor's argument of "proof" to hold, the research subjects resisting to engage in the experiment should remain to do so – even in different, more extreme experiments or situations.

Aristotle) display the highest degrees. Particular communities can consist of only vicious or only virtuous people, or can be mixed. Second, different experiments in moral psychology show different levels of "virtuous behaviour" (Merritt et al., 2010, p.p. 356-357)[34], understood as consistency in engagement in particular actions, which implies that someone whose virtuous character enables her to cope with one challenging moral setting might be unable to cope with another, depending on the context. This does not mean that such a person's character is not "virtuous", but instead that whether the degree to which her virtuous character is sufficiently cultivated to deal with a particular situation depends in part on the situation, or is *mediated* by the situation, and not merely by the agent's character. If we accept this claim, moral psychologists seem to be at least to some extent warranted in drawing their conclusions. Third, the supposed warrant that a minority of people will be able to deal with situations like the one presented in this experiment, which – importantly – is *not* technosocially opaque (people are aware of the way in which their actions conducted through technologies have certain consequences) is one that no moral philosophy can be satisfied with and especially not one that introduces the additional factor of technosocial opacity. In a world of technosocial opacity, the virtuous conduct of a small minority is not sufficient for safeguarding the flourishing of humanity; and can even less be expected given the greater complexity and opacity as compared to the Milgram experiment. For instance, we can consider whether a non-digital native with the highest degree of virtuous character would be sufficiently capable of dealing with the novel setting of ubiquitous digital technologies. The answer would probably be negative.

Nonetheless, this does not disqualify the project of constructing a virtue ethics of technology. Rather, it shows that "technomoral" virtues are co-shaped by both people's characters *and* by "settings". A philosopher of technology would criticise the conclusions drawn by moral psychologists on the basis that the factor of technological mediation is fully "black-boxed" by sketching the concept of "setting" as something that is somehow separated from a "person". Instead, she would argue that on the one hand "setting" – in contrast to what Vallor argues – should be taken seriously, but that on the other hand the concept of setting should be turned into a concept of technological mediation. Especially in the case of the Milgram experiment, the factor of technological mediation seems important for

---

[34] Notably, four out of five of the experiments presented by Merritt et al. have a technology as an important mediator of the "situation", be-it money, a prison (as a disciplinary technology), or mechanical noise (Merritt et al., 2010: 357). One example is ambiguous, for the variable is "being in a hurry", but one does not need many persuasive arguments to at least consider the possibility that "being in a hurry" in our contemporary society is highly technologically mediated (e.g. see Bucher et al., 2013).

understanding the "situation". The technology of an electronic system being triggered by a voltage lever constitutes a process of what Coeckelbergh designates as "distancing" (Coeckelbergh, 2015): the research subject does not simply "hurt" a victim, but hurts the victim in a mediated way, through a device that presents no feedback with regard to the severity of the harm and in an architecture that separates the research subject from the victim and thereby creates a moral distance between them.

Recent scholarly work that revisited the Milgram experiment seems to be largely in agreement with the above. For instance, Haslam et al. (2014) argue that the setup of the experiment, which included aspects like its "location, the appearance and behaviour of the experimenter, and the *technical apparatus*" (Haslam et al., 2014, p. 275 – emphasis added) mediated the interpretation of research participants of their actions as taking place in the context of the paradigm of scientific research. In other words, the situational aspects provided research participants with an interpretative framework (related to the authority of science) through which they could understand the technical practice they were engaging in. In a similar vein, Burger (2014) discusses situational aspects of the experiment that mediated the actions of the research participants. He explicitly focuses on the role of the technical device used in the experiment (a lever indicating voltage levels), arguing that both its detachment from the victim's responses and its capacity to incrementally administer electric shocks contributed to the choices made by the research participants (Burger, 2014, p. 492). Even though these scholars do not unpack the notion of technological mediation, they provide good reasons for integrating the notion of "setting", which includes the use of technology, in a virtue ethics account of the Milgram experiment.

The illustration of the Milgram experiment points at a general tendency that runs throughout Vallors work, of emphasising human agency in dealing with technology and insufficiently discussing the impact of technologies on this agency. This can also be observed in the solutions she puts forward for problems posed by technology, be-it "improved technomoral education"(Vallor, 2016, p. 204), "cultivating technomoral humility" (p. 207) or cultivating "renewed technomoral courage" (p. 218). Next to these suggestions, no systematic guidance[35] is provided regarding the way in which we can understand how technology design impacts the virtues and how we could design technologies to enhance the cultivation our virtues. In other words, Vallor offers systematic theoretical guidance for cultivating the

---

[35] Vallor does occasionally discuss specific design features, for instance concerning data storage for child safety applications (Vallor, 2016, p. 206). However, these discussions are largely anecdotal and are not supported by a theoretical, systematic account of how the technology design (e.g. child safety application) mediates the virtues and how different alternatives could either enhance or obstruct cultivation of these virtues.

virtues to deal with technological change, but does not offer similar guidance for designing technologies in ways that cultivate and do not obstruct the virtues. In accordance with the recommendations presented in the previous chapter we do need such guidance, notably for being able to consider how technologies might "embed" or "incorporate" certain virtues. Hence, we argue that Vallor's virtue ethics of technology needs an account of technological mediation to be suitable as a basis for our novel method.

4.4.2 Revisiting MacIntyre and remaining shortcomings

Our second criticism is that Vallor does not provide a distinct theory of "technomoral" virtue, in the way characterised by MacIntyre (MacIntyre, 2007, p. 183). Vallor concedes that she answers the "what" question of virtue by providing a list of "technomoral" virtues. However, the more fundamental "what" question that is posed in virtue ethics, namely "what is virtue" and – by extension – "what is technomoral virtue" remains largely unanswered by Vallor, which is reflected in the varied and loose way in which she defines virtues, which are being referred to as "traits", "dispositions", but also as "discerning skills" and "recognitions" (Vallor, 2016, pp. 37,126). Even though she indicates that her account of virtue ethics is in accord with MacIntyre's, she does not seem to pay sufficient heed to his argument that his theory of practice is what *unifies* the different theories of virtue he discusses (MacIntyre, 2007, p. 191). In fact, Vallor takes a different approach, by pragmatically synthesising conditions for cultivating virtues and lists of virtues in different virtue ethics traditions, without however providing a distinct, unified theory of virtue. We argue that even though the synthesising of different virtue traditions has been fruitful, not paying heed to the central role of MacIntyre's theory of practice has led to a missed opportunity of providing a distinct theory of virtue that draws from contemporary reflections on technology and technical practice.

MacIntyre's most notable contribution to the field of virtue ethics is his theory of practice that largely incorporates and goes beyond Aristotle's conditions for virtuous practice as presented in Book 3 of *Nicomachean Ethics*[36]. That is, Aristotle provided an account of *conditions* of virtuous practice: that it should be voluntary, based on a rational decision, be open to deliberation, and be in accordance with the wish of the agent; leading to the conclusion that the virtues and vices are within the agent's power. For Aristotle, these conditions ought to be met for a practice to lead to the cultivation of the virtues. What

---

[36] Even though MacIntyre surpasses Aristotle in this respect, he at the same time acknowledges that his account "captures much, but very far from all, of what the Aristotelian tradition taught about virtue" (MacIntyre, 2007, p. 203).

MacIntyre crucially adds to Aristotle's conditions of praxis, is an account of how a practice relates to a human life and to society as a whole (or, as we discussed above, he adds a complementary theory of social reality). Going back to the discussion of Milgram's experiment and the crucial notion of its technologically mediated "setting", we can see how MacIntyre's theory of practice might offer a better starting point than Vallor's synthesised conditions for virtuous practice. That is, MacIntyre enables us to understand the practice of participating in a scientific experiment and the role of symbolic and technological mediations in this practice. For instance, his theory allows us to show that causing harm to a victim as happened in the Milgram experiment relates to a socially established activity through which goods *external* to that activity (honour or monetary reward as result of participating) are realised[37]. He thereby points at the importance of considering the setting of an *institution* in which the experiment takes place that is the social bearer of the practice (MacIntyre, 2007, p. 195).

Considering the Milgram experiment in light of MacIntyre's theory, we can state that the institutional setting of participants and the "victim" being in different rooms mediated the practice in question. Participants were not able to directly communicate with the victim and for instance ask the victim questions or touch the victim's body. They only had mediated and limited access to the alleged victim's suffering, comparable to some degree for instance with a drone pilot attacking targets in a desert in Afghanistan. Moreover, the "action on a distance" as enabled by the pushing of a button mediated the practice by preventing the participant from receiving direct feedback, as would for instance be the case when punching the victim and being confronted with a verbal or bodily response. These are significant aspects to consider when discussing the activity as a "practice" that either obstructs or cultivates the virtues of the participants in the experiment. Considered at a higher level, of "narrative structure", the question of "what are you doing" will not merely be answered by "pushing a button to apply electric charges to the victim". Instead, it will be answered in narrative form, which includes accounts of participating in scientific experiments, relating to an authoritative scientist and of the alleged system behind the button and the wires connected to the alleged victim. Finally, we can consider the practice as well as the narratives through which it is rendered intelligible as embedded in a moral community. In this community, science has a certain authoritative status; technology is often considered as a neutral

---

[37] According to MacIntyre's theory, we cannot label this activity a "practice" due to the fact that goods external to the activity realise it. However, we will criticise MacIntyre's idealised notion of practices at a later point, which allows us to uphold the label here.

instrument, and so forth. We can access all these aspects of the Milgram experiments through MacIntyre's theory of practice.

However, MacIntyre's theory has certain drawbacks that will eventually lead us to consider Ricoeur's theory of practice in chapter 6 as an alternative and superior theory of virtue. Three of these drawbacks are central. First, MacIntyre posits an opaque and idealised notion of practice[38]. This notion is opaque because it includes criteria of *complexity, coherence, cooperative activity* and *systematic extension* that lack qualification. That is, his theory of practice does not admit for questions such as: "Which activities count as complex and which as simple?", "What does it mean for an activity to be coherent?", "why would only cooperative activities count as practices?", and "what does it mean for internal goods to be systematically extended?" One can for instance question whether "playing the violin", which MacIntyre designates as a practice, is a cooperative activity because one can play the violin exclusively by oneself. MacIntyre's very extensive notion of practice thereby creates a problematic degree of conceptual opacity. At the same time, however, MacIntyre's notion of practice is idealised (cf. Carr, 2003, p. 256). On the one hand, it seems to surpass Aristotle's notion of praxis in its narrow sense - of being the kind of action that allows for the cultivation of the virtues -, by admitting for practices such as playing chess, the violin or the piano, which Aristotle would have arguably considered as belonging to *technê,* to the crafts or arts that serve an instrumental purpose and are not concerned with the virtues[39]. On the other hand, however, MacIntyre counts *only* those activities as practices that realise goods that are internal to the particular type of activity. In this regard, he idealises Aristotle's notion of practice by conceding that for an activity to be regarded as a practice, it necessarily needs to support the cultivation of the agent's virtue. For Aristotle, when we consider the conditions for praxis as discussed above, the activity can also lead to a deficiency or an excess of a feeling or thought; for instance when the agent's wish that constitutes the practice is the *wrong* one. For Aristotle, therefore, a warrior engaging in battle and charging hesitantly and thereby putting his comrades in danger still engages in a practice, though he can also be called a coward (showing a lack of virtue). For MacIntyre, however, this would disqualify the

---

[38] Previously, we only presented a part of MacIntyre's definition of practice. His entire notion of practice encompasses: "any coherent and complex form of socially established cooperative 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, 2007).

[39] In this choice, however, we will follow MacIntyre because we reject Aristotle's strictly instrumental notion of *technê*. As we will argue in chapter 6, instrumentality should be considered as a matter of degree, rather than a matter of different kinds that somehow allow for "pure forms" (e.g. a craft being "purely instrumental).

very activity from being a practice, leading to a much narrower and more idealised notion of practice than the one proposed by Aristotle. The major problem of this narrow conception of practice is that it creates a dichotomy without proper qualification. For instance, a violin player might at first be motivated by her parents to play the violin (e.g. realising the external good of honour) and gradually move to be motivated according to the internal goods (e.g. mastery) of the activity. In MacIntyre's framework, "playing the violin" would thereby move from being a non-practice to a practice. However, it is not at all clear at what point we can designate the activity as a "genuine" practice in its idealised sense. This problem becomes increasingly pressing when activities become more complex, notably in technologically mediated environments (e.g. when considering practices of stock market trading or playing video games).

Second, MacIntyre uses a restricted notion of narrative, being only tied to a person's history in the sense of stories "told in the thick of everyday life" (Ricoeur, 1992, p. 158). As such, MacIntyre argues that we make sense of our practices by relating them to our recounted past experiences and the past experiences of others. However, by taking this idea of personal or common historical narrative as unproblematic, MacIntyre fails to account for the difference between historical and fictional narrative and how these both narrative modes inform our practices. As Ricoeur concedes (Ricoeur, 1992, p. 158), the reinterpretation or refiguration of our practices through fiction produces a thorny problem for MacIntyre's theory. The problem that MacIntyre faces is that, because of its productive qualities, literary fiction can configure novel interpretations of practices and the standards of excellence through which the internal goods to these practices are realised that sometimes are in discord with a moral community. A historical example would be "Uncle Tom's Cabin": a fictional narrative that broke with the ideas of virtuous practices of the moral community of the time (of slave owning and management) that it addressed. As Carr argues, MacIntyre's theory opens up the possibility of a problematic type of moral relativism (Carr, 2003, p. 257), because the criteria for a practice to be desirable or not can only be found in the historical narrative surrounding an agent and her moral community. However, this moral community can have conceptions of practices, their internal goods, and standards of excellence that are quite clearly morally wrong (as illustrated by example of slavery).

Third, MacIntyre fails to properly integrate or relate the three major components of his theory of practice: practice, narrative unity and moral community. For practices, he argues that they are made intelligible through narrative structures but he does not provide an account of *how* this happens. For narrative structures, similarly, he argues that such structures fit

within a moral community that has a certain conception of the good life, but does not provide an account of *how* they fit. As Devine argues, "there is a gap in MacIntyre's account between the concept of a practice and the concept of a tradition" (Devine, 2013, p. 115): it is unclear how heterogeneous sets of practices (e.g. having a conversation, playing an instrument, conducting political debates) combine through narratives to form a tradition. The connection between the different parts of MacIntyre's theory will need to be made explicit in order to move beyond anecdotal evidence. For instance, *that* we commonly explain what we do when engaged in a practice by telling a narrative does not by itself offer sufficient grounds for assuming that therefore practice and narrative are necessarily connected and should be integrated in a single theory of practice.

4.5 Chapter Summary

In this chapter, we took three steps to explore how we can use virtue ethics and the virtue ethics of technology approach as the basis of our novel method, what limitations these approaches have and how we might address these limitations. As such, we considered (1) what we can take from the relevant approaches and (2) what we should leave open for further consideration. First, we considered the contemporary virtue ethics tradition, considered as the revival of Aristotelian ethics in western philosophy. Second, we particularly considered Vallor's virtue ethics of technology approach that due to its thematic focus on ethics of technology seemed most suitable for our purposes. Third, we discussed the limitations of Vallor's approach, moved to MacIntyre's theory of practice to address some of these limitations and consequently raised some shortcomings of MacIntyre's theory as well.

In order to summarise what we take with us from our discussions in this chapter and what issues we leave open for further considerations, we can consider the three central questions of virtue ethics and what preliminary answers we can formulate for these:

1 **What is virtue?** In answering this question, we will stay very close to the Aristotelian tradition, stating that virtue is a state of character that inclines the agent to decide on the right action in a certain context. Furthermore, we agreed with MacIntyre that the answer to this question is dependent on the answer to the question of how the virtues are cultivated, because that is where different theories of virtue converge.

2 **Which virtues can we distinguish?** In answering this question, we largely followed Vallor's heuristic of "technomoral" virtues (though we will from now on dismiss the adjective "technomoral"). We indicated however, that we will replace

"empathy" with "friendship" and "wisdom" with "prudence" and that we will uphold the Aristotelian distinction between virtues of character and virtues of thought.

3   How are the virtues cultivated? In answering this question, we followed MacIntyre's theory of practice, while acknowledging some of its limitations that need to be further considered in the following chapters. To address these, we will need (1) a non-idealised notion of practice, (2) a theoretical consideration of the interplay between historical and fictional narrative, and (3) an integral account of the three "stages" of virtue (practice-narrative-moral community).

In the following chapters, we will largely focus on the answer to the third question and leave the answers to the first two questions mostly in their current shape, though we will briefly reconsider them in chapter 6. Reasons for focusing on the third question are (1) because an elaborate reconsideration of the Aristotelian "ontology" of virtues (question 1) falls outside of the scope of this dissertation, and (2) because the theory of practice has showed relevant to the issue of "setting" and to the related issue of technological mediation that we will focus on in the next sections.

4.6   References

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

Anscombe, G. E. M. (1958). Modern Moral Philosophy. *Philosophy*, *33*(124).

Aristotle. (1999). *Nicomachean Ethics*. (T. Irwin, Ed.). Indianapolis: Hackett Publishing Company.

Battaly, H. (2011). Is Empathy a Virtue? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199539956.001.0001

Blok, V., Gremmen, B. and Wesselink, R. (2016). Dealing with the Wicked Problem of Sustainability: the Role of Individual Virtuous Competence, *Business and Professional Ethics Journal (forthcoming)*.

Bolsin, S., Faunce, T., & Oakley, J. (2005). Practical virtue ethics: healthcare whistleblowing and portable digital technology. *Journal of Medical Ethics*, *31*(10), 612–8. http://doi.org/10.1136/jme.2004.010603

Bucher, E., Fieseler, C., & Suphan, A. (2013). the Stress Potential of Social Media in the Workplace. *Information, Communication & Society*, *16*(10), 1639–1667. http://doi.org/10.1080/1369118X.2012.710245

Burger, J. M. (2014). Situational features in Milgram's experiment that kept his participants shocking. *Journal of Social Issues*, *70*(3), pp. 489–500. doi: 10.1111/josi.12073.

Carr, D. (2003). Rival Conceptions of Practice in Education and Teaching. *Journal of Philosophy of Education*, *37*(2), 253–266. http://doi.org/10.1111/1467-9752.00324

Coeckelbergh, M. (2012). Care Robots, Virtual Virtue and the Best Possible Life. In P. Brey, A. Briggle, & E.

Spence (Eds.), *The Good Life in a Technological Age*. New York: Routledge.

Coeckelbergh, M. (2015). *Money Machines: Electronic Financial Technologies, Distancing, and Responsibility in Global Finance*. Farnham: Ashgate.

Curzer, H. J. (2017). Yesterday's Virtue Ethicists Meet Tomorrow's High Tech: A Critical Response to Technology and the Virtues by Shannon Vallor. *Philosophy and Technology*, forthcoming. http://doi.org/10.1007/s13347-017-0269-z

Devine, P. (2013). The Concept of Tradition : A Problem Out of MacIntyre. *Reason Papers*, *35*(1), 107–123.

Ess, C. (2009). *Digital Media Ethics*. Camebridge: Polity Press.

Foot, P. (1978). *Virtues and Vices and Other Essays in Moral Philosophy*. Berkeley: University of California Press.

Haslam, S. A., Reicher, S. D. and Birney, M. E. (2014). Nothing by mere authority: Evidence that in an experimental analogue of the milgram paradigm participants are motivated not by orders but by appeals to science. *Journal of Social Issues*, *70*(3), 473–488.

Hursthouse, R. (1999). *On Virtue Ethics*. Oxford: Oxford University Press. http://doi.org/10.1093/0199247994.001.0001

Hursthouse, R. (2006). Are Virtues the Proper Starting Point For Morality? In J. Dreier (Ed.), *Contemporary Debates in Moral Theory*. Malden: Blackwell Publishing.

Ihde, D. (2009). *Postphenomenology and Technoscience*. New York: Sunny Press.

Latour, B., & Porter, C. (1993). *We Have Never Been Modern*. Camebridge, Massachusetts: Harvard University Press.

MacIntyre, A. (2007). *After Virtue: A study in moral theory* (Third Edit). Notre Dame, Indiana: University of Notre Dame Press. http://doi.org/10.1017/CBO9781107415324.004

McDowell, J. (1979). Virtue and Reason. *The Monist*, *62*(3), 331–350.

Mcrae, E. (2017). Finding a Place for Buddhismin the Ethics of the Future: Comments on Shannon Vallor's Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. *Philosophy and Technology*, *Advance On*. http://doi.org/10.1007/s13347-017-0287-x

Mejlgaard, N., Christensen, M. V., Strand, R., Buljan, I., Carrió, M., Cayetano I Giralt, M., Griessler, E., Lang, A., Marušić, A., Revuelta, G., Rodríguez, G., Saladié, N. and Wuketich, M. (2018). Teaching Responsible Research and Innovation: A Phronetic Perspective, *Science and Engineering Ethics*. Springer Netherlands, pp. 1–19. doi: 10.1007/s11948-018-0029-1.

Merritt, M. W., Doris, J. M., & Harman, G. (2010). Chapter 11: Character. In J. M. Doris (Ed.), *The Moral Psychology Handbook* (pp. 1–504). http://doi.org/10.1093/acprof:oso/9780199582143.001.0001

Nussbaum, M. C. (1999). Virtue Ethics: a Misleading Category? *The Journal of Ethics*, *3*(3), 163–201.

Partridge, E. (2006). *Origins: A Short Etymological Dictionary of Modern English* (fourth edi). London: Routledge.

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artifacts: or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, *14*(3), 221–232. http://doi.org/10.1177/030631284014003004

Ricoeur, P. (1992). *Oneself as Another*. (K. Blamey, Ed.). Chicago: University of Chicago Press.

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford:

Oxford University Press.

Verbeek, P.-P. (2005). *What things do; philosophical reflections on technology, agency, and design*. Pennsylvania: Pennsylvania University Press.

5 Narrative Technologies[40]

5.1 Introduction

In this chapter, we develop an account of technological mediation that is based on the notion of "narrative". As a starting point for this endeavour, one might look within philosophy of technology (for example revisiting the earlier work of Ihde (1979) on technology and hermeneutics) or one might look elsewhere (philosophy of language and philosophical hermeneutics not usually associated with philosophy of technology). Because of the problems that we encountered following the first route in chapter 3, we choose the latter route in this chapter: we use the later work of Paul Ricoeur to propose a framework to investigate what we call "narrative technologies". Although Ricoeur has recently been brought to the attention by philosophers of technology (cf. Gransche, 2017; Kaplan, 2006; Lewin, 2012; Romele, 2017), a thorough reworking of Ricoeur's narrative theory to construct a theory of technological mediation has thus far not been engaged in.

The impact of narrative on the design and use of technology becomes apparent by reviewing literature in different disciplines. In the field of computer science, game designers argue that "choices about the design and organisation of game spaces have narratological consequences" (Jenkins, 2003, p. 129). Similarly, researchers in artificial intelligence argue that the richness of "narrative intelligence" might solve enigmas resulting from the simplicity of formal understanding of computer intelligence (Mateas & Sengers, 1999). The significance of narrativity for social reality becomes apparent in discussions about narrative in law and history of technology. It is argued that narrative constitutes an important aspect of legal judgement that goes against purely formalist conceptions of law (Patterson, 1990). Moreover, social scientists show that technologies like the Internet are embedded in a "cultural-narrative" (Goodfellow & Hewling, 2005) and that historical events like the Fukushima nuclear disaster are understood according to multiple narratives that reveal an interplay of technology, politics and society (Jones, Loh, & Sato, 2013). Hence, narrativity appears to be an issue of great interest for studies of technology, the humanities and for the social sciences.

---

[40] Substantial parts of this chapter have been adopted from the papers: Coeckelbergh, M., & Reijers, W. (2016). Narrative Technologies: a Philosophical Investigation of the Narrative Capacities of Technologies. Human Studies, 39(3), 325–346.; *and*: Reijers, W., & Coeckelbergh, M. (2016). The Blockchain as a Narrative Technology: Investigating the Social Ontology and Normative Configurations of Cryptocurrencies. Philosophy & Technology, 7. http://doi.org/10.1007/s13347-016-0239-x

The question that guides the investigation in this chapter is whether Ricoeur's narrative theory can inform an approach to technological mediation that meets the criteria we outlined in chapter 3, thus taking into account language and the social. We start by providing a high-level overview of Ricoeur's work and its potential value in addressing the interplay between technology, language and the social, which contemporary theories of technological mediation have thus far inadequately accounted for. Subsequently, we present an exposition of the concepts of emplotment and configuration in Ricoeur's narrative theory, followed by a discussion of the implications of his theory for our understanding of technology. Accordingly, we construct a theory of narrative technology based on Ricoeur's narrative theory that clarifies the notions of technological emplotment and technological configuration.

5.2 The Philosophy of Paul Ricoeur

We start by presenting a high-level overview of Ricoeur's oeuvre. This overview will assist us in grasping the coherence of Ricoeur's thinking and consequently of the relevance of his thought for the theoretical challenges that we face, as presented in chapter 3. No brief outline of Ricoeur's work will enable us to do full justice to its breadth and depth, and we will therefore merely sketch some of its central aspects and its merits for addressing aspects of language and the social.

5.2.1 Ricoeur's philosophical anthropology

Throughout Ricoeur's work runs a concern with a philosophical anthropology that centres in on the critique of the Cogito as introduced by Descartes, which presumed "immediate certitude and self-transparency" (Jervolino, 1990, p. xxi). The goal of Ricoeur's philosophical anthropology is to provide an alternative account of the self and simultaneously of the social world in which oneself is confronted with another (Kaplan, 2003, p. 9). Opposed to the Cartesian notion of the self, Ricoeur posits a notion of the self as *indirect*, which characterises hermeneutics as a philosophy of "detours" (Ricoeur, 1992, p. 17). A detour, in Ricoeur's work, consists of problematising a philosophical position that needs to be resolved (an aporia) by invoking another philosophical position that might seem completely foreign to it, for instance problematising Augustine's conception of time by invoking Aristotle's in *Time and Narrative* (Ricoeur, 1983).

Philosophical reflection, according to Ricoeur, departs from the "fullness of language" that is always already there in human experience and understanding (Kaplan, 2003, p. 20). A human does not find herself having immediate, direct access to a world, using

language as a layer on top of this world to refer to it. Instead, her experience and understanding of her world are mediated through language and can therefore only be understood by means of a detour through language, involving aspects such as symbols, metaphors and narrative.

Even though Ricoeur did not construct a closed philosophical system, there is a clear coherence in his thought and the way in which he approaches his concern with philosophical anthropology. That is, he has traced the way in which language mediates one's understanding of oneself and one's lifeworld through different stages of linguistic operation. Initially, Ricoeur focused on the *symbol* as a theme of philosophical reflection (Ricoeur, 1967). Because symbols operate in the public realm and are socially established and maintained, we can have no "direct" knowledge of them. A symbol is an expression that does not merely express what a speaker means to say but also designates something that lies beyond the speaker's intention, therefore having a *double meaning*. To demonstrate this point, Ricoeur discusses how the traditional symbolism of evil in the Western world, conceived as guilt and sin (a violation of divine command) contains a more primordial understanding of evil as defilement (an expression of chaos in nature) (Lai, 1984, p. 317). Therefore, when we invoke the symbolism of evil through the notion of sin, we indirectly invoke its more primordial meaning of defilement, which nonetheless remains hidden in the common use of the symbolism. Hence, symbols in use carry with them a "surplus of meaning" that relates to their histories and often-mythical origins. They reveal meaning and at the same time conceal meaning, and the task of hermeneutics is to retrieve the hidden meaning concealed in symbols. From the symbol, Ricoeur moved to investigating the role of *metaphor* in the creation of meaning in language (Ricoeur, 2004). Metaphor can be understood as a heuristic fiction that redescribes reality (Kaplan, 2003, p. 48). A famous example of a metaphoric expression is: "all the world's a stage, and all the men and women merely players" (Shakespeare 1623, p. 52). A more contemporary example would be a sentence such as "the human mind is simply a computer". By redescribing our idea of the world through the heuristic fiction of the theatrical stage, or our idea of the human mind through the heuristic fiction of a computer, a new possibility of *being-in-the-world* is opened up. This possibility reveals the capacity of language to *create* meaning, to be productive. As such, a metaphor always carries with it a tension between its literal meaning ("the computer") and its imaginative meaning (the human mind understood *as* a computer).

From his discussion of metaphor, Ricoeur progressed to create an account of how the rediscription of reality could *transform* our being-in-the-world. For this, he turned to an even

higher level of linguistic operation, of *narrative*. Metaphor and narrative remain intimately connected because they both indicate the synthesis of heterogeneous elements. According to Kaplan, "in metaphor it is the ability to perceive resemblance in difference; in narrative it is the ability to construct a meaningful organisation of otherwise unrelated events" (Kaplan, 2003, p. 50). A narrative, such as a folktale, gathers together a heterogeneous collection of events and characters, that are configured in a meaningful whole that we refer to as a "plot". Ricoeur argues that we can understand how textual narratives redescribe reality in which meaningful action takes place. By taking Ricoeur's work as a starting point for thinking about technology and technical practice, we argue in line with Lewin that "we are interested in technology first and foremost because we are interested in human freedom and agency" (Lewin, 2012, p. 59). Ricoeur's ongoing project is searching for a philosophical anthropology, and any corresponding theory of technology will not be "about" technology or technological objects, but about technology's role in the being-in-the-world of human beings. Because we need to proceed through the long detour of symbols, metaphors and narrative to understand what and who a human being is, we need to perform the same task to understand the human as a "tool-making" being.

What makes Ricoeur's theory particularly relavant for our purposes is his general focus on *mediation*, in particular the linguistic mediation of action. In relation to technological mediation, Ricoeur's philosophical anthropology can be considered to be close to the one put forward by Georg Simmel (1900). Simmel considers the human, "paradoxically, as an indirect being" (Simmel, 1900, p. 227). Humans, according to Simmel, are tool-making, technological animals, neither being determined by mechanical instincts, nor having unmediated powers and access to the world (Simmel, 1900, p. 226). Ricoeur, in a similar vein, takes the notion that there is no immediate access to the self, leading to the conclusion of the non-coincidence of man with his "self" (Jervolino, 1990, p. 15). Humans, for Ricoeur, are essentially fallible in the sense that conflict, particularly between interpretations of the self, is the most primordial (original) constitution of the self (Jervolino, 1990, p. 18). Simmel similarly acknowledges this fundamental reality of conflict that arises from the antinomy between the movement of life and the stability of form (Gangas, 2004, p. 18). For both Simmel and Ricoeur, therefore, humans are intermediate beings whose access to the world is in each case mediated. They are philosophers of mediation par excellence.

Additionally, we should acknowledge Ricoeur's philosophy not just as a philosophy *of* mediation, but also a *mediating* philosophy. That is, his work constantly mediates between different philosophical traditions and scholarly traditions outside of the boundaries of

philosophy. This aspect of his work is important to consider, because it points at the potential for bringing disparate ideas together that each have their value in our discussions, such as Searle's philosophy of society and Ihde's postphenomenology. As Davidson argues, "a philosophy of mediation"…"must begin by setting the existing viewpoints (the endoxa), which are all too often disengaged from one another, side by side" (Davidson, 2006, p. 2). In line with this characterisation, Ricoeur consistently follows "the thought of the other to its end in order to surpass that thought" (Kemp, 1996, p. 41). In his writings about language, Ricoeur incorporates ideas from both the Anglo-American tradition (commonly referred to as "analytic") and the European tradition (commonly referred to as "continental"); which motivates him to go as far as to compare the works of philosophers who are at face value each other's opposites, such as Husserl and Wittgenstein (Ricoeur, 2014). Analytic philosophers of language significantly influenced Ricoeur, who draws from Austin's and Searle's speech act theory and Davidson's agentless semantics of action. At the same time, Ricoeur engages with key scholars in the traditions of phenomenology and hermeneutics, such as Gadamer and Levinas. Moreover, Ricoeur draws from the work of historians such as White, semioticians such as Greimas and literary theorists such as Frye. This broad and inclusive character of Ricoeur's work does not only build bridges between different philosophical traditions, but also enables a rich palette of philosophical critique. While building on the works of this great variety of thinkers and drawing connections between their theories, Ricoeur remains critical at all times and incorporates no theoretical insight at face value. As such, we argue that Ricoeur is the philosopher par excellence to remedy polarising tendencies in contemporary philosophy that also affect philosophical accounts of technology, by constructing a critical philosophical mean between extremes.

5.2.2   Language and social reality in Ricoeur

In chapter 3, we established that the theories of technology we surveyed (philosophy of society, postphenomenology and approaches in STS) insufficiently deal with either aspects of linguistic mediation or with of collective intentionality (the social). Consequently, in order to justify the turn towards Ricoeur's work on narrative we have to show that it deals with these two aspects. Unlike philosophers who predominantly focus on the material dimension of technological mediation, Ricoeur has a primary, though not exclusive, interest in human language. In his interpretation of works of language, he especially focuses on the "*grasp of language on experience*" (Ricoeur, 2014, p. 29), on the ways in which human language mediates human experience. To illustrate this grasp of language on experience, we can

consider our knowledge of traffic rules (e.g. when *such-and-such* a situation occurs, I react like *this*) as a determinant of our experience of a traffic situation, or our knowledge of ritualistic protocols during the experience of a ceremony (e.g. the installation of a new president occurs when *such-and-such* declarations are uttered). In both these cases, the way we experience the particular situation or event is mediated by our understanding and use of language. Central to Ricoeur's understanding of language lies the interpretation of text, which he argues can be seen as a model for meaningful *action*. Ricoeur states that the object of the human sciences is meaningful action, and that meaningful action is understood by considering the discourse that shapes it, which in turn is fixed in writing – in the paradigmatic unit of a text (Ricoeur, 1971, p. 322). Like textual interpretation, the interpretation of action is a dialectic of explanation and understanding because like texts, actions are readable, having a meaning that is distanced from the intentions of the actors, and subject to conflicting interpretations (Kaplan, 2003, p. 68). In order to investigate how human action becomes meaningful through discourse, both spoken and written, we need to study how meaningful action is made intelligible in written texts, through language. More specifically, Ricoeur argues that humans interpret their every-day actions as configured by narratives, of which texts are the paradigmatic reifications (Meretoja, 2014, p. 98). A clear example of such a text is a diary, in which personal interpretations of every-day action are recounted and thereby re-interpreted (Hassam, 1990).

Ricoeur's ideas about language are strongly influenced by Heidegger, though unlike Heidegger he is also sympathetic to formalist theories that aim at providing atomistic explanations of language. A major premise that Ricoeur employs throughout his work is that human experience mediated by language is characterised by temporality: by the "within-time-ness" of human experience. Human experience and action, according to Ricoeur, are essentially mediated by language within a temporal setting, and hence are not dependent on direct access to point-like, unmediated experiences of the present. In his paper *Narrative Time*, Ricoeur asserts: "my first working hypothesis is that narrativity and temporality are closely related - as closely as, in Wittgenstein's terms, a language game and a form of life" (Ricoeur, 2015, p. 169). As such, when we consider the "time" of an action (e.g. "it's *time* to go to work"), our temporal understanding of the world is mediated by language in its narrative mode. Just how closely narrativity and temporality are related is the major subject of discussion in Ricoeur's seminal work *Time and Narrative*, which will be the major source for constructing our theory of narrative technologies.

At this point, we have shown what Ricoeur potentially has to offer to address the neglect of language in contemporary philosophy of technology. But what role does the social play in Ricoeur's work? Throughout his work, Ricoeur stresses the importance of interpreting human existence by considering its embeddedness in human social reality. In his work, "the mediation of action by the imagination" as recounted in the narrative mode, "links individual with social action by bridging the realms of discourse and action (Kaplan, 2003, p. 81). In line with the "within-time-ness" of human beings, Ricoeur argues: "the time of narrative is *public* time" (Ricoeur, 1980, p. 175 - emphasis added). Moreover, he asserts that: "public time"…"is not anonymous time of ordinary representation but the time of *inter-action*. In this sense, narrative time is, from the outset, time of being-with-others" (Ricoeur, 1980, p. 188). Reconsidering Van Den Eede's critique of postphenomenological theories of technological mediation, we can observe that he uses a term that is almost similar to Ricoeur's "being-with-others": the technological mediation of "being-with-each-other" (Van Den Eede, 2010, p. 140). Thus, Ricoeur asserts that the social is explicitly present in narrative time, which remains a time of interactions between people without being made entirely anonymous, separated from human experience and action (as for example can conversely be said about the time of natural science that is derived from natural laws rather than from human experience).

Consequently, by accounting for ways in which our interactions with narratives are embedded within our temporal existence, Ricoeur explains why narratives can configure social reality: because they configure narrative plots that refigure social events (Borisenkova, 2010, p. 93). Emplotment, which is the process that defines any narrative structure, has an outspoken normative character because the characters in a narrative are not just neutral "doers" as Searle would portray them but are "endowed with ethical qualities" (Ricoeur, 1983, p. 59). Unlike generalised "doers" like the rational economic man who figures in economic theories, acting according to coherent, non-normative motives, characters in narratives can be good or evil, rational and irrational; the protagonists or antagonists of the narrative structures. These features of Ricoeur's theory enable him to go beyond Searle's formal approach, providing a holistic, normative account of linguistic mediation of our social world.

Taken together, we can convincingly draw two conclusions with regard to the use of Ricoeur's work: (1) that Ricoeur's philosophy can be characterised as both a philosophy of mediation and a mediating philosophy and (2) that it consistently pays heed to the linguistic and social aspects of mediations of the human lifeworld. These conclusions support our use

of Ricoeur's work in response to the problems of theories of technological mediation we encountered in chapter 3, and pave the way towards our upcoming task, of connecting Ricoeur's philosophy with a notion of *technological* mediation.

5.3 Exploring Ricoeur's Narrative Theory

In one of his major works that consists of three volumes, *Time and Narrative*, Ricoeur (1983; 1985; 1988) constructs a comprehensive narrative theory. This theory revolves around a basic model that designates the way in which a narrative, considered according to the paradigm of the text, mediates human reality. The starting point for the formulation of Ricoeur's theory is the hypothesis that "*time becomes human to the extent that it is articulated through a narrative mode, and narrative attains its full meaning when it becomes a condition of temporal existence*" (Ricoeur, 1983: 52 - emphasis in original). Ricoeur considers human experience, that is irreducibly temporal and social, to be shaped through a *narrative mode*.

5.3.1 The model of emplotment

What then, is this "narrative mode"? Ricoeur claims that the *mediation* between time and narrative that is implied in this mode revolves around a process that he designates as *emplotment*. He derives his theory of emplotment from Aristotle's *Poetics* (1902)*,* grounding it in three stages of Aristotle's core concept of *mimesis*, which he characterises as "the active process of imitating or representing something" (Ricoeur, 1983:, p. 33). Aristotle says about mimesis that "the instinct of imitation is implanted in man from childhood" and that this is an "instinct for 'harmony' and rhythm" (Aristotle, 1902, p. 15). Accordingly, the significance of understanding this process of imitation goes beyond the realm of literature and finds its philosophical significance in its purported grounding in human existence in general. Aristotle's definition of mimesis is derived from the overriding principle of *muthos*, which designates the *plot* of a narrative. "The plot is the imitation of the action: - for by the plot I mean here the arrangement of incidents" (Aristotle, 1902, p. 25). From this, Ricoeur derives that *emplotment* designates the organisation of events[41] by which people represent action in a plot. Paradigmatic examples are works in the genres of *tragedy* and *comedy* in which characters imitate probable accounts of human action structured according to a play script or scenario, which is an organisation of events.

In a narrative such as a tragedy, the plot *configures* different elements like characters, motivations and events in a meaningful whole. Emplotment, in other words, creates a

---

[41] Ricoeur explicitly refers to "events" rather than "incidents" because he later in his work concludes that an event as implied in a text might cover a very extensive time-span (Ricoeur, 1983, p. 217).

harmonious concordance out of discordant, heterogeneous elements. Therefore, the movement of a plot accords with a teleological principle, being an "inexorable movement that drives the story toward an anticipated conclusion" (Dowling, 2011, p. 6). It makes the reader or spectator say: *this story makes sense*. How then, does emplotment shape the human experience and understanding of temporal existence? Essentially, this amounts to the movement of *prefigured* time that becomes a *refigured time* through the mediation of *configured time* (Ricoeur, 1983, p. 54). To illustrate this: when reading Plato's allegory of the cave our prefigured time involving ideas of human knowledge (before reading) changes by means of interaction with the text in the configured time (during reading) and is subsequently synthesised with our experience of the world in the refigured time (after reading). The reading of Plato's allegory mediates our experience of the world because it for instance leads us to consider our own experiences as analogous to the shadows cast in the cave in which we are supposed to be imprisoned as presented in the allegory. The three phases that constitute the movement of emplotment are defined as follows:

- **Mimesis$_1$**: the prefigurative phase. This phase consists of the understanding of the world of action: its semantics, its symbolic order and its temporality.
- **Mimesis$_2$**: the configurative phase. In this phase, characters and events are organised in a meaningful whole, a plot.
- **Mimesis$_3$**: the refigurative phase of reading. This concludes the narrative circle, of applying narrative to the prefigured world of action.

Mimesis$_1$ refers to our prior understanding of the human world of action. According to Ricoeur, this prior understanding can be analysed by considering its three basic conditions: human competence to identify action in terms of its *structure,* human competence to identify the *symbolic* mediations of actions and the human understanding of the *temporal* elements of action. As such, mimesis$_1$ indicates the initial moment at the start of the reading of a text, a moment that is embedded in a social context – in a human "repertoire" from which we engage with new social phenomena. However, mimesis$_1$ extends beyond the particular text because it concerns the realm of human action in a social setting that is already mediated by narrative. This means that the way we experience our temporal, social existence is in each case shaped by narrative structures (Ricoeur 1983, p. 54). For example, we understand the act of going through the passport control at an airport because we understand the structure of the act (if I'm asked my passport, I hand it over), we understand its symbolic mediations ("EU"

desk for EU citizens) and we understand its temporal setting (first I do *x*, then the official does *y*). At the same time, we also understand ourselves, and our life worlds through narratives about our national identities (e.g. "I'm a citizen of the Netherlands"), about the economy (e.g. "I lost my job due to the financial crisis"), and about technologies (e.g. "robots are going to render many jobs superfluous"). Thus, whenever we engage with human language we act from a cultural basis, which means that our understanding is shaped by structures, symbols and temporal elements that are a part of our collective narrative "repertoire" (Ricoeur, 1983, p. 64). This view is largely in line with approaches in STS, such as Jasanoff's account of socio-technical imaginary, which is conceptualised as a collectively held repertoire of narratives ("visions of the future") embedded in "shared forms of life and social order" (Jasanoff, 2015, p. 6). However, it is only the starting point of the model of emplotment, for we are interested foremost in the ways that these "shared forms of life" come to be and are transformed.

Mimesis$_2$, the phase of narrative configuration, is central in Ricoeur's work and makes explicit how a narrative can configure prefigured time (Ricoeur, 1983, p. 64). This phase denotes the process of emplotment: the mediation of prefigured time by means of a *plot*. The plot mediates between individual events and the whole of a story, it brings together heterogeneous factors (characters, goals, interactions) belonging to the realm of action (mimesis$_1$) into a syntagmatic order and it mediates the temporal dimensions of prefigured time. For instance, in the well-known narrative Oedipus Rex several heterogeneous characters (Oedipus' mother, his lover, his father, the king), events (prophecies, trials and murders) and goals (trying to evade the prophecy, aiming to know its truth) are brought together in a syntagmatic[42] narrative whole, which constitutes the story's surprising, but acceptable, resolution (Ricoeur, 1983, p. 65). As such, the plot is the organisation of elements of a narrative (characters, events) that makes it possible for someone interacting with a text to follow it to a certain conclusion. Ricoeur elaborates upon the mediation of prefigured time by introducing the idea of two temporal dimensions in the process of emplotment: chronological and non-chronological dimensions of narrative time. The chronological dimension is concerned with an episodic arrangement of events, which characterises the narrative in terms of sequence of events (firstly *this* happened, secondly *this* happened) (Ricoeur, 1983, p. 66). The non-chronological dimension of narrative time is concerned with the configuration of events in a meaningful, temporal whole. This implies that the organisation of events is made

---

[42] "Syntagmatic" refers to the relation between different linguistic elements that occur sequentially in discourse.

intelligible, or rather "followable" (referring to the human ability to "follow" a story) in such a way that the order of events leads to a conclusion that characterises the narrative as a whole; that renders its resolution, according to Ricoeur, acceptable for the reader. It is through the mediation of the non-chronological temporal dimension of a narrative that we can proclaim: *this story makes sense.*

The world of the text and our human world intersect at the moment of refiguration, **mimesis₃** (Ricoeur 1983, p. 71). Refiguration is therefore the third conceptual moment in Ricoeur's model: the moment at which the narrative circle has been closed, or rather completed, and the lifeworld of the reader is transformed. At this point we should note that, Ricoeur acknowledges that the relation between narrative structures and our understanding of the social world is a circular one. That is, the organisation of the narrative structure helps us to understand the social world, but at the same time the understanding of the social world is the basis for any novel narrative structure. To understand this relation we have to deal with a "hermeneutic circle" that consists of the stages of prefigured, configured and refigured time. In opposition to a mere tautology (i.e. "experience *is* narrative structure"), this hermeneutic circle is, in the words of Ricoeur, a "healthy" one, "in which arguments advanced about each side of the problem aid one another" (Ricoeur, 1983, p. 76). More specifically, the dimensions of narrative structures that are revealed through textual hermeneutics (e.g. temporal dimensions) are understood in terms of their relation to the world of action, and our engagement with the world of action in turn feeds into the practice of textual hermeneutics. In what follows, we will explore how Ricoeur's model of emplotment can be used to conceptualise the way in which technologies configure human experience and understanding of the social world.

5.3.2   Narrativity and technological mediation

How could we employ Ricoeur's narrative theory to understand technological phenomena, related to the practises and outcomes of R&I? To answer this question, we have to explore in what sense technologies can have narrative qualities themselves, rather than being merely shaped by narratives that are constructed *about* them. As an initial concern, we face the challenge of reconciling Ricoeur's narrative theory with an understanding of technology. At face value, textual narratives and material technologies seem to be far apart. Consider for instance the apparent gap between literary texts that Ricoeur deals with in the second volume of *Time and Narrative* like Proust's *Remembrance of Things Past* (Ricoeur, 1985) and technologies like cars, computers and cameras. Written stories are made up of words and

sentences while cars are made up of materials like plastics and steel. However, we have reasons to question this apparent dichotomy. Ricoeur points at the *generality* of his narrative theory by mentioning a taxonomy of classes of narrative that includes "myths, folklore, fables, novels, epics, tragedies, dramas, films, comic strips, to say nothing of history, painting and conversation" (Ricoeur, 1985). Moreover, he claims that the narrative understanding engraved in history also applies to disciplines like "cosmology, geology, and biology" (Ricoeur, 1983, p. 135). Thus, his theory of narrative configuration is meant to go beyond the works of literary fiction and historical narrative and to include visual objects like paintings and knowledge in the scientific field of biology. Taking narrative as an ontological concept, as a concept that mediates human experience and understanding in each case, we can see how it does not remain an exclusive concern of the understanding of texts. Arguably, the process of emplotment does not only happen through the interaction with actual written texts, but also through interaction with cultural forms such as theatre plays, movies, paintings, and – as we argue – technologies. Considering the basic notion of emplotment, no absolute separation can be enforced between what one might colloquially want to call "soft" cultural phenomena like texts on the one hand and "hard" technological phenomena on the other hand like airplanes. The next chapter will present a more in-depth argument explaining why a strict separation between text and technology cannot be sustained. For now, it suffices to argue that both our understanding of a text and our understanding of a technology are actualised through inter-action. Insofar as interaction is made intelligible through narrative, as we might recall from our discussion of MacIntyre, emplotment governs the meaning of both text-in-reading and technology-in-use. What we additionally argue, alongside Ricoeur, is that narrative understood along the lines of the paradigm of the text can inform our understanding of technological mediation, because the process of emplotment is most readily accessible in written texts.

Once we dismiss the strict dichotomy between our understanding of technologies and texts, we open up the possibility of employing the latter to investigate the first. To do so, we need to ask: to what extent can it be said that a technology brings about a "plot" in similar ways as a text does? In other words, the central question for thinking about technologies becomes: can technologies *configure* our narrative understanding in similar ways as written texts do? Kaplan (2006) provides some initial answers to this question. He argues that narrative theory can be used in order to interpret the way in which humans "read" technology (Kaplan, 2006, p. 49). Moreover, he points out that there are certain ways in which humans can construct plots to understand technology, for example by telling about the motivations for

designing a technology (Kaplan, 2002, p. 4). However, taking into considerations the lessons learned from theories of technological mediation, we want to make explicit how humans and technologies co-shape each other by using the notion of narrative technologies. In other words, we argue that humans do not only read technologies, but technologies on the other hand "read" humans, insofar as what is experienced by a user must first be constructed in the technology (Ricoeur, 1983, p. 50). Ricoeur defends this argument in a crucial passage:

> "It is the act of reading that accompanies the narrative's configuration and actualises its capacity to be followed. To follow a story is to actualize it by reading it. And if emplotment can be described as an act of judgment and of the productive imagination, it is so insofar as this act is the *joint work of the text and reader*, just as Aristotle said that sensation is the common work of sensing and what is sensed" (Ricoeur, 1983, p. 76 - emphasis added).

In other words, a reader is only able to imagine and judge the course of events in a narrative insofar as the text simultaneously offers the resources for following and actualising it. Thus, if we then take Ricoeur's narrative theory as our guiding light, we need to see using or "reading" a technology similarly as a reciprocal process. Therefore, the term "configuration" as it is used in mimesis$_2$ seems most suitable here, implying that a configuration of a technology by a human denotes a corresponding configuration of the human by the technology.

To substantiate the move from text to technology (though not substituting the one for the other), let us then briefly reflect upon the characteristics of emplotment that can be found in the design and use of technologies. In order to defend the claim that technologies configure our narrative understanding we need to show that – just as texts – they are involved in the organisation of events. Moreover, we need to defend the claim that the organisation of events includes both chronological and non-chronological dimensions of narrative time. As a starting point, we consider the way technology design is characterised in design literature to ascertain a clearly narrative structure that is used to characterise the functioning of technologies. As an example, we might consider the passage about the "confirmation principle" in the use of dialog boxes in software, as captured by Lidwell, Holden and Butler (2003, p. 44 - emphasis added):

> "Confirmation using a dialog involves establishing a **verbal interaction** with the **person** using the system. It is most commonly **represented** as a **dialog box** on a software display

(e.g., "Are you sure you **want** to delete all files?"). In this method, dialog boxes directly ask the **user** if the **action** was **intended** and if they would like to proceed."

In this passage, which serves as a guideline for technology design, heterogeneous factors including characters (people and, more specifically, users), events (action, verbal interaction, representation) and motives (want and intention) are brought together in a coherent whole, in what we might characterise as a plot. This observation is in line with Kaplan's claim that technology design involves narrative. However, we do not only want to show that designers express their ideas in a narrative mode to understand the technologies they create; we want to go further by showing that technologies themselves *configure* a plot. Through reading, a textual narrative actively configures the prefigured understanding that a reader has of her social world (Ricoeur 1983, p. 53). For instance, a reader's understanding of the impact of surveillance technologies might be transformed through a reading of Orwell's *1984*. Similarly, we maintain that a technology can actively configure a person's narrative understanding of her social world. The interpretation of a technology-in-use entails a coinciding active process of "reading" and "writing", understood as two aspects of the practical activities of writing and reading a text. That is, while we use a technology, we follow its process of emplotment (mimesis$_2$) and through this our experience and understanding of our social world is transformed (mimesis$_3$).

In the use of a technology, the technology and its user *together* configure an organisation of events. For instance, we can say that a car, as a technology, configures events such as "starting the engine" and "adjusting the mirrors" in a meaningful whole. Many aspects of this process are related to a prefigured understanding of the world. For instance, before a person has ever driven a car she will already have an understanding of the way the traffic functions and of the car as a cultural artefact (including understanding of for instance the environmental impact of the use of cars and the impact of driving too fast). Through interaction with the car, however, this prefigured understanding is *configured*; the understanding of both the traffic and the car as an artefact is altered and integrated into a new understanding of the social world[43]. As such, a technology and the technological system in which it is used are actively involved in shaping the way we understand our activities, experiences and relations with other people. The notion that both technologies and humans

---

[43] An existing empirical study about the narratives of older women driving cars clearly shows the dependency of the understanding of driving a car on both the "practical" narratives (narratives about the practice of driving) and "experiential" narratives (narratives arising from the experience of driving) (Siren and Hakamies-Blomqvist, 2005)

configure an organisation of events is important since it conceptualises, from within a narrative approach, a key insight in contemporary philosophy of technology since Heidegger: technologies are not mere instrumental means, but mediate meaningful human action. Thus, by using Ricoeur's work on narrative we can convincingly support the claim that technologies have the capacity to configure plots, understood as organisations of events.

Unfortunately, most scholars in philosophy of technology dealing with the ways in which technologies mediate the human life world (cf. Feenberg, 1999; Ihde, 2009)[44] do not consider Ricoeur's work in their theories. This is somewhat surprising, particularly because Ihde's first book on hermeneutics explicitly dealt with Ricoeur's earlier work (Ihde, 1971). This, we believe, can partly be explained by noting that Ihde wrote his first book before Ricoeur's "turn from eidetic, descriptive phenomenology to hermeneutic phenomenology in which signs and symptoms mediate understanding" (Kaplan, 2003, p. 19). By failing to take up the fundamental hermeneutical questions raised by Ricoeur, Ihde only manages to expand hermeneutics to the material sphere by postponing engagement with these questions (Lewin, 2012, p. 58). We argue that his postponing of relevant questioning has resulted in the problems of postphenomenology that we identified earlier: accounting for the dimensions of language and the social in technological mediation.

After an extensive period of relative silence on Ricoeur's work, however, David Kaplan (2006), building on his account of Ricoeur's critical theory (Kaplan, 2003), drew a connection between Ricoeur's work and the philosophy of technology. Kaplan suggests that Ricoeur's hermeneutical method as well as his analysis of the hermeneutic circle between human experience and narration can be fruitful in discussions about technology (Kaplan, 2006, p. 43) because these elements can enrich the analysis of technological mediation by including notions of linguistic and social mediation. Moreover, he argues that the model of the text can be utilised as the model of the mediation of experience by technology (Kaplan, 2006, p. 49), for it can make explicit how humans interpret technologies and how technologies play a role in our narrative understanding. More recently, signs of a renaissance of Ricoeur's work in philosophy of technology can be witnessed. Lewin (2012) uses Ricoeur's work to draw questions of technology in the field of philosophy of technology back towards fundamental questions in philosophy and hermeneutics in general as discussed by Ricoeur. He criticises Kaplan's claim that Ricoeur's conception of technology is "essentialist", and instead positions Ricoeur's latent understanding of technology as an

---

[44] After the "material turn", Ihde largely abandoned Ricoeur's hermeneutic project. In that regard, Ricoeur's work has played a very insignificant role in the early works in the postphenomenological tradition.

answer to Heidegger's positioning of technology as the greatest danger (Heidegger, 1977). That is, while Ricoeur admits - in line with Heidegger - that modern technology conceals its use-context, he at the same time positions narrative, the refiguring of the order of things (Lewin, 2012, p. 64), as the way in which we might obtain a free relation towards technology. Applied to a concrete case, Gransche uses Ricoeur's narrative theory to investigate computer simulations, arguing that "simulations in a broader sense, as well as computer simulations specifically, are poietic operations; they are a creative production and, as processes of mimesis, combine creation with imitation" (Gransche, 2017, p. 48). In a similar vein, Romele uses Ricoeur's narrative theory to discuss "imaginative machines", arguing that digital technologies imitate the human productive imagination. In addition, he concedes "that the human productive imagination takes place in digital technologies" (Romele, 2017, p. 13), showing how the model of narrative configuration can inform our understanding of ICTs.

By means of their discussions, Gransche and Romele have gone beyond Kaplan's initial work by using Ricoeur's narrative theory to understand particular technologies. They also posit an indirect criticism of a part of Lewin's interpretation of Ricoeur, by – in line with our efforts – allowing technologies themselves to have the capacity to be involved in narrative configurations. As such, our account distances itself from Lewin's Heideggerian claim that "behind all technologies is a basic desire to foreground functionality and conceal complex operations" (Lewin, 2012, p. 64). Instead, emphasising Ricoeur's own criticism of Heidegger (cf. Ricoeur, 1988, 1992) and his disagreement with Heidegger's ontology of understanding (Bobb, 2011, p. 339), we state that freedom from the decontextualizing force of technology is not only to be found in the recounting of stories (the narrow conception of narrativity) but also in the engagement with narrative configurations that technologies themselves bring about. Paradoxically, therefore, certain technologies can be the antidote to the ailment brought forward by others, just as utopian narrative can be the antidote to ideological narrative (Ricoeur, 1986). Understanding this dynamic should perhaps be considered the core of the entire philosophical enterprise labelled "philosophy of technology".

The recent revival of philosophical ponderings about technology in light of Ricoeur's work have been of great help for repositioning fundamental hermeneutic questions at the heart of the discipline and for sketching the outlines of a "Ricoeurian" philosophy of technology. However, although the above-mentioned scholars discuss the value of Ricoeur's theory for philosophy of technology and use it to interpret particular technologies, they do not yet offer a comprehensive theory of technology or technological mediation that is inspired by Ricoeur's work on narrative theory. This is what we intend to do in the following section.

5.4   Narrative Technologies

In this section, we propose a theory of technological mediation that is informed by Ricoeur's narrative theory, which we coin the theory of "narrative technologies"[45]. While dismissing any strict dichotomy between technical and narrative understanding, our ontological premises are that (1) human experience and understanding are mediated by narrative structures and (2) that these narrative structures are accessed through involvement with the ready-at-hand, which includes a variety of "things" such as books, movies, houses, computers and trains. Our theoretical framework offers four central concepts that pertain to technological mediation and are informed by essential distinctions made by Ricoeur in his discussions of textual hermeneutics. To consider technologies in relation to the paradigm of the text, we offer two central concepts that capture distinctions between technologies: *textuality* and *literacy*[46]. "Textuality" pertains to the extent to which technologies approximate the paradigm of the text, whereas "literacy" pertains to the extent to which humans are attuned to involve themselves with the configurations technologies bring about. To consider the modes of technological configuration that are at stake when interpreting a particular technology, we offer two additional central concepts[47]: *temporality* and *distancing*. "Temporality" pertains to the dimension(s) of time and "distancing" pertains to the abstraction(s) from the world of action that technological configurations bring about.

5.4.1   Technology, textuality, literacy

Initially, we consider two central concepts that pertain to technologies as particular phenomena that humans encounter in their world – similar to particular instances of narrative such as novels, movies, theatre plays, and so on. Not all technologies actively configure our narrative understanding, even though they are in each case involved in prefigured time. As such, they differ in that they have more or less in common with the paradigm of the text, in terms of their textuality. Furthermore, just as readers have to be literate to a certain extent to read a text, users of technologies have to equally possess a "literacy" to engage in the process of technological emplotment. We use the notion of "literacy" in a broader sense than in its

---

[45] In doing this, we follow Bobb's characterisation of Ricoeur's philosophy as a "reversed ontology" that departs from the ontological presupposition of starting "outside" of our self-understanding, as technological beings, and moves towards this understanding through textual hermeneutics (Bobb, 2011, p. 244). Hence, to construct our theory of technological mediation we first need to make explicit its ontological presuppositions and from that basis move towards the epistemological dimensions borrowed from narrative theory.

[46] In an earlier paper, "literacy" was instead conceptualised as "interpretative distances" (Reijers & Coeckelbergh, 2016).

[47] In an earlier paper, "temporality" was conceptualised as "activity" and "distancing" was conceptualised as "abstraction" (Coeckelbergh & Reijers, 2016).

common usage, in which it is characterised as the capacity to read and write in a particular natural language. Here, it is understood (1) praxis-oriented, extending beyond the mastery of a particular natural language and including the mastery to engage in technological configuration in terms of "coding", "building", "hacking", "designing", "tweeting", and so on, and (2) contextually, meaning that to be literate in for instance using a computer does not only depend on one's own capacities but also on the design of the respective computer, the languages it supports, the user-interface it has, and so on.

First, concerning **textuality**, we need to acknowledge that technologies differ in the extent to which they bring about technological emplotment because some technologies - in particular ICTs - are very similar to the paradigm of the text while others - such as a hammer – are very different from it. The reason for this difference, which will become more explicit in the following sections, can be found in the capacity of these technologies to through repeated use transform the experience and understanding of their user's social world. Just as there is a strong difference – though not a discontinuity – between early forms of symbolic communication such as cave paintings and elaborate texts such as Shakespearean plays, and thus a difference concerning the *possibility* of doing textual hermeneutics, there is a difference – though not a discontinuity – between ancient, less textual technologies and modern, more textual technologies. In common words, we could explain this difference by saying: the more a technology-in-use *tells us* something about ourselves and the world we live in, the more it can be regarded as being textual. The notion of textuality assists us in determining the horizon of interpretation, and thereby also postulates a limit for any hermeneutic inquiry of technology. That is, we argue that we can only interpret a technology in use as far as it approaches the paradigm of the text. Consequently, the further a technology is removed from the paradigm of the text, the more we rely on recounted narratives, on testimonies, of its use and significance. This limitation corresponds to the limitation of textual hermeneutics, which as a mode of investigation becomes less feasible the more the phenomenon that is interpreted is removed from the paradigm of the text. The limit that follows from the concepts of textuality is informative insofar as it guides the type of investigation that is appropriate for considering technological mediation: the more a technology is removed from the paradigm of the text, the less we can rely on interpretations of its structure (i.e. design) and the more we rely on recounted accounts of its use[48].

---

[48] This claim should be understood in relation with an earlier one, which stipulated that we understand action (i.e. technologies in use) through the interpretation of recounted action in narratives.

The extent to which a technology approximates the paradigm of the text is determined by its capacity to actively "emplot" human narrative understanding. Some technologies have little in common with the paradigm of the text, and play for the most part a role in our prefigured understanding. For instance, a bridge is predominantly part of a prefigured narrative structure in which events and characters are already configured into a plot: it may be a bridge to transport goods and people across the Rhine River. When a bridge gets built, it plays a role in configuring our narrative understanding (for example by disclosing new areas of a country) but it progressively becomes part of our prefigured time. Such an understanding of technology ties into Heidegger's analysis of the bridge (Heidegger, 1977, p. 16): the bridge has become a passive element of human culture in the course of several generations. It configured the narrative understanding of the people who interacted with it once it was build, but in time became an element of their social reality that found *closure*. However, some technologies bring about an *active* process of emplotment of the narrative understanding of their designers and users. Through interaction with these technologies, a process of emplotment is brought about that implies a simultaneous "reading" and "writing" of our narrative understanding. Information and Communication Technologies (ICTs) are exemplary for this type of narrative technologies, being very closely related to the paradigm of the text. This can first of all be derived from their very textual character in a literal sense: many forms of human-computer interaction revolve around mediation by textual information. More importantly, though, ICTs and humans can be said to "co-author" the narratives they engage in. For instance, social media (e.g. Facebook, Twitter) are very text-like technologies, not only regarding their superficial textual aspects, primarily understood as textual elements of their user-interfaces, but also, and more importantly, with regard to the configurative capacities of their code-at-work: their capacities to *organise* characters (e.g. avatars) and events (e.g. social media updates) in a meaningful whole. This does not refer to the actual reading of the code, for instance by a software developer, but rather to the narrative structures configured by the code. For instance, interactions with the code of a social media platform can configure a person's narrative structures that form his understanding of practices related to "liking" and "friendship". Thus, instead of looking at the present-at-hand source code in order to investigate the narrative structures it configures, the notion of textuality makes us look at the ready-at-hand plot, the organisation of characters and events, as it is configured by a technology.

Second, with **literacy** we consider the extent to which humans are attuned to engage in a process of technological configuration with particular technologies. As such, we interpret the

126

extent to which a technology and related technological configurations are accessible or inaccessible for different people. To draw an illustrative comparison that assists in understanding this distinction: consider the narrative configuration of George Orwell's *1984* and the commentaries related to this it (e.g. hearsay about the message of the book). A reader of the book might have his ideas transformed concerning the role of technologies in modern society. Someone else, who didn't read the book, might nonetheless comment on it to illustrate something, for instance by referring to a CCTV camera as an "Orwellian technology". In such a situation, the commentator has a prefigured understanding of what an "Orwellian technology" entails, but fails to have access to the original process of emplotment (the plot in *1984*). Likewise, there is a difference in understanding of the technological configuration of an ICT between a developer who works with its code (for whom the technological configuration is very accessible), its user who interacts with an application interface and a person who only hears or reads *about* the technology without having used it (for whom the technological configuration is inaccessible). In contemporary debates about digital technology, this distinction is often characterised in terms of "digital literacy" (Eshet-Alkali, 2004), which denotes a difference between the modern "digital natives" who know their way around the technological configurations of digital technologies and the "digital illiterates", to whom the respective technological configurations are inaccessible.

Carr captures this difference by conceptualising two types of narratives as the *practical* "first-order" narrative (narratives configured *by* technologies), and the "second-order" narrative (narratives *about* technologies) that have the same subject but a different aim; namely an aesthetic or cognitive one (Carr, 1986, p. 131). This reflects Ricoeur's distinction between "commentary" (which can be a text *about* a narrative) and a narrative configuration itself (Ricoeur 1985, p.68). Even though both these types of narrative structures have the same form, or in Ricoeur's terms the same *schematism,* and are therefore fundamentally interrelated, we need to distinguish between them because they denote an interpretative distance between humans and technologies that can lead to differences in experiencing and understanding the world. That is, the more accessible a technology is and the more one interacts with it, the more proximate one gets to its first-order narrative configuration. Some technologies, such as hammers or bikes, are accessible to most people, who can be considered "literate" with regard to their capacity to access the first-order technological emplotment of hammers and bikes. Conversely, a layperson who is mildly interested in military drones has much less power in co-authoring their narrative structures

and a *different*[49] understanding of them than a core developer of these drones or a military officer operating them. This is not to say that those people interacting with first-order narrative structures *necessarily* have a greater power overall over the technological emplotment of military drones, compared to people who only interact with their second-order narrative structures. For instance, a layperson holding a powerful political position can enforce regulations that strongly influence the prefigured narrative understanding in which the design of military drones takes place. Rather, literacy resonates with what Foucault designates as the authority of expert knowledge (Philipps, 1996) as a characteristic of those people interacting with first-order narrative structures of technologies such as military drones. According to this understanding of power, people who are proximate to the first-order narratives are powerful in the sense that they have acquired the capacity to co-author the process of emplotment. A powerful regulator, when being far removed from the first-order narrative, would not have this power to co-author the process of technological emplotment but only to change the prefigured time in which it takes place.

5.4.2 Modes of technological configuration: temporality and distancing

The following two concepts do *not* pertain to any single technology - considered as a phenomenon comparable to a novel or a theatre play - but to the process of technological configuration. Just as different narrative modes can overlap in a single literary work, such as the modes of historical and fictional narrative, different modes of technological configuration can overlap "in" a singular technology-in-use. The two distinctions concerning technological configuration that we consider are (1) the extent to which a technological configuration brings about a rigid or dynamic *temporality*, and (2) the extent to which technological configuration *distances* itself from the world of action, either engaging with or abstracting from it. Important to note is that the related distinctions between technological configurations are to be considered as capturing differences in degree rather than differences in kind, thereby denouncing for instance the notion that a technology can configure absolute abstracting narrative structures[50]. The distinctions function as epistemological "relay stations"[51],

---

[49] We explicitly refer to the *difference* in understanding of the narrative structure, rather than for instance to a *lesser* understanding of a layperson. In certain cases, remoteness from technological narrative configuration might actually help inform the human about important hidden aspects of this configuration (think for instance about the position of game addicts vis-à-vis non-addicted non-gamers).

[50] Ricoeur deals with this "temptation" to consider "absolute" mediations in volume 3 of *Time and Narrative*, in which he attacks Hegel's notion of total mediation between human culture and the individual (Ricoeur, 1988, p. 202).

[51] Ricoeur uses the term "relay station" to indicate an epistemological structure which enables one to proceed from one level of a hermeneutic analysis to the other (Ricoeur, 1983, p. 182).

constituting the detour that Ricoeur envisages throughout his work, with which we can approximate the narrative structures as configured through technical practices. Both distinctions, of temporality and distancing, capture the difference between what Ricoeur conceptualises as the modes of *historical* and *fictional* narrative, which are defined in accordance with the ways they concretise their respective intentionalities (Ricoeur, 1988, p. 188). That is, whereas historical narrative aims to achieve a level of strict representation of historical events, literary fiction aims at providing both the author of a text as well as its reader with a sense of imaginative freedom, and correspondingly a sense of responsibility (the responsibility of following the plot). We aim to show how these features of historical and fictional narrative correspond with the ways in which technological configurations mediate our sense of time and represent reality. Importantly, Ricoeur stresses that the two major narrative modes of history and fiction should not be seen as absolutely separate, but rather as being interwoven (Ricoeur 1988, p. 99). That is, history and fiction borrow from each other's intentionality (the ways in which they are intended to relate to the world) and eventually adhere to the same model of emplotment that makes them intelligible.

First, we consider how technological configuration mediates our understanding of the world by organising the **temporality** of the plot[52]. In line with Ricoeur's notion of narrative configuration, we argue that technological configuration brings about a "third time" (Dowling, 2011, p. 76), that is distinct from a purely "cosmic time" (understood as movement of worldly bodies) and a "time of the soul" (understood as a result of the internal flow of consciousness). Dowling illustrates the way in which not only a text, but also a material setting can mediate a human's sense of time. He states anecdotally that "the agonizing wait to get medical attention at the emergency room last week may have taken up three or four hours, but the telling of it"-…-"may take only ten minutes" (Dowling, 2011, p. 47). This observation refers to the notion of a "double temporality" that characterises narrative, of the temporality captured by the narrative itself (e.g. characters in a historical narrative about the First World War that engage in events in the course of four years) and the time of the act of narrating (e.g. finishing the reading of the respective narrative in a single day). That is, there is "a *telos* [end] that carries characters forward in a state of imperfect knowledge about the consequences of their actions, with a narrator who, gazing backward on events from a fixed"…"perspective, has arrived at certain conclusions about their meaning or significance"

---

[52] In structural terms, we could state that the distinction concerning temporality is concerned with the use of *tenses* (perfect, imperfect, pluperfect, etc.) and the *grammar of time* ("now", "then", "earlier", "today" "twelve o'clock") in narrative discourse.

(Dowling, 2011, p. 88 - emphasis added). What this teaches us about technological configuration is that the act of *using* a technology (as analogous to the act of reading) does not necessarily coincide with the temporality of the technological configuration that it brings about. For instance, the act of placing a fishing net in a riverbed at dawn might take only a moment, but it involves a narrative structure that stretches throughout the day, including the anticipation of catching some fish before dusk. Similarly, the setting of a calendar entry can be done in a blink of an eye, while it could configure a narrative structure that involves several hours of the day.

How can we comprehend the way in which technological configuration mediates temporality? Ricoeur argues that the organisation of temporality of a narrative depends on two distinct temporal dimensions: a chronological, or episodic, and a non-chronological or configurational dimension (Ricoeur 1983, p. 66). He also refers to these dimensions as the time of the world, public time, and lived time, time of personal experience (Ricoeur, 1988, p. 182). The chronological dimension comes about by means of an episodic sequence of events (in common terms: "first this happened, secondly that happened"). This dimension is eventually directed at abolishing the human sense of temporality (Ricoeur, 1983, p. 160) by reducing temporal experience to "simple succession" of measured intervals. In contrast, the configurational dimension enables a reader (1) to oscillate between the narrative - as a whole - and separate events, (2) to jump between different "times" (e.g. as happens in a flash back), and (3) to create a sense of ending. For technologies that actively configure our social world, this means that they can either enforce a rigid temporal structure on our understanding of the social world or a dynamic one. Consider for example the *car* as a technology. In the practice of driving, some events are determined in a chronological order, like starting the engine *before* driving away. In contrast, adjusting the seat or the mirrors can be done in many different orders; such events are organised according to a non-chronological dimension, involving a sense of personal freedom and responsibility. The non-chronological dimension of configuration also means that a single event only makes sense within the larger whole. For example, adjusting the mirrors refers back to the plot of "driving" as a whole and indeed to the practice-as-narrative: it ties in with the narrative about what good, virtuous driving is, for instance driving includes taking into account and responding to the actions of others[53]. This dimension of narrative time consequently entails a dynamic that closes in on human temporal experience.

---

[53] In the next chapter, we will further elaborate on this intersection between a technical practice and narrative.

We argue that temporality is configured in the process of technological emplotment in two ways: (1) by means of enforcing or relaxing strict successions of events and (2) by means of "connectors", which can be dating mechanisms, version control mechanisms, and tracing mechanisms. The first way is illustrated above by the example of the car and has been applicable to technological configurations throughout human history. The ancient windmill already rigidified temporality in its use by configuring strict successions of events: the brake needed to be removed *before* the wings could be set into motion. However, most of the events configured in the plot of "milling" can be freely organised by the miller, whose use of the mill invokes a similar kind of responsibility as Ricoeur assigns to the reader of a work of fiction: of following the plot. The second way in which temporality is configured relates to what Ricoeur designates as the "connectors" that "make historical time conceivable and manipulable" (Ricoeur, 1988, p. 182). Historians use constructions of dating (linking events to dates and clock times), the succession of generations (the replacement of the dead by the living) and traces preserved in documents and archives (marks left by humans and things, indicating a "here" and "now" in the past). For instance, historians refer to traces contained in historical documents such as the Treaty of Versailles, which was signed on the 28th of June 1918, to construct the historical fact of the signing of the Treaty of Versailles. Technologies, and notably modern technologies, configure similar constructions in order to make historical, public time conceivable. Dating happens through the application of dating mechanisms in technological systems, such as the system time mechanism in computers. The notion of succession of generations relates to technologies mostly in our prefigured time, to narratives *about* technologies. William Turner majestically illustrated this in his painting "The Fighting 'Téméraire'", in which one of the last battle ships that fought during the Battle of Trafalgar is towed by a paddle-wheel steam tug, symbolising the passing of one generation of a technology to another. Yet, in contemporary technologies we see a passing of generations implemented in the form of version control, leading to "old" and "updated" versions of systems. As a figure of speech, this has even been adapted in popular culture, by referring to the predecessors as "1.0" and by referring to the contemporaries as "2.0". Finally, tracing has been implemented in technologies by means of sensing and recording the passing of humans and things, for instance through the sensing and recording of mouse clicks and page views in web browsers.

To illustrate how the two ways of technological configuration of temporality come together, we consider the example of blockchain technologies. A blockchain[54] can be regarded as a public digital ledger (a book of accounts) that contains all the transactions made within its system. "Blocks" are digital, time-stamped records containing the most recent transactions that are cryptographically signed and added to the blockchain in a designated sequence, in a linear, chronological manner (Mcreynolds et al., 2015, p. 3). Whenever a transaction occurs, anywhere in the world, so-called "mining nodes" validate it and add it to the public blockchain, which makes it impossible for the same digital object (which could a monetary transaction, but also a contractual agreement) to be "double spent": to be transacted to different addresses at the same time. This design feature accounts for the so-called "time-stamping problem" by providing "temporal authentication of electronic documents" (Rompay et al., 1999, p. 1), and thereby of digital traces. Miners are the agents that collectively control the computational nodes validating transactions within the network. For Bitcoin, the service these miners provide is guaranteed according to a system of incentives, which currently amounts to the miners being rewarded transaction fees and newly created Bitcoins. We can examine the two ways in which temporality is configured in a blockchain technology such as Bitcoin. First, events are organised in a strict succession with limited dynamism. To make a transaction, a user *first* needs to login to her wallet, *secondly* enter a destination address, and so forth. In other words, the process of technological configuration already organises events in the form of a strict succession. Second, the Bitcoin network (1) records traces of humans and things that left marks "here" and "now" in the past (for instance: signed transactions), (2) dates these traces by time-stamping them and (3) embeds these operations in a logic of version-control that happens by means of "forks" (updates to the basic code that can lead to the branching of different paths of software development). As such, Bitcoin combines the strict succession of events and the three "connectors" that are also used by historians in historical narrative to make historical, public time conceivable and, as Ricoeur argues, to consequently abolish the human sense of time. It would therefore be no overstatement to characterise blockchain technologies as "history machines".

Second, we propose a distinction between modes of **distancing**, between technological configurations that engage with or abstract from the world of action. Crucial to understanding this distinction is the consideration of the significance of historical narrative on the one hand as "standing for" something that really happened and of fictional narrative on

---

[54] The definition provided only captures the meaning of one particular type of blockchain architecture, namely an open blockchain. Other forms are consortium blockchains and private blockchains.

the other hand as instantiating "imaginative variations", concerning events whose actual occurrence is bracketed (Ricoeur, 1988, p. 177). We argue that a similar difference in narrative modes (understood as a difference in degree between two extremes) can be applied to the process of technological configuration, meaning that it can (1) configure narrative structures that engage people with the world of action by means of bringing about imaginative variations, or (2) configure narrative structures that are almost strictly representational ("standing for" events that *really* happened) and thereby abstract from the world of action. The first type of technological configuration offers a great sense of freedom (and, correspondingly, of responsibility) by offering the possibility of a multitude of imaginary variations in which the emplotment of characters and events takes place. Conversely, processes of technological configuration that abstract from the world of action fixate our narrative understanding in the form of a public archive containing both *documents* and *traces*[55] (in the form of traceable transactions of digital objects) that stand for, or represent the past "as it really happened". As Ricoeur importantly notes, the term *really* is signified only through the concept "as" ("*as*" it really happened – representing a reality that has itself become inaccessible) (Ricoeur 1988, p. 155). As Ricoeur argues: "this abstraction is a result of *forgetting* the complex interplay of significations that takes place between our expectations directed toward the future and our interpretations oriented toward the past" (Ricoeur 1988, p. 208 – emphasis added). This forgetting is the main effect of the abstraction brought about by the "standing for" of a technology's narrative configuration.

We need to make explicit what the above-mentioned process of distancing entails[56]. Note that this second distinction effectively captures the *social* dimension of narrative technologies, for these degrees of abstraction mediate intersubjective (or "being-with-others") relations. Ricoeur argues that the conclusion of a narrative does not need to be predictable but rather acceptable. To defend this claim, he analyses the notion of causal explanation in the paradigm of historical narratives (being narratives about "real" events as opposed to fictional narratives). Although he rejects the idea of historical explanation with recourse to laws, he tries to preserve *causal analysis* and *rational explanation* in history

---

[55] Ricoeur argues that it is "the use of documents and archives that makes the trace an actual *operator* of historical time" (Ricoeur 1988, p. 184). In other words, a trace that refers back to something that has been there in the past (such as a fossil, but equally so a validated block on the blockchain that can be "re-traced") combined with the use of documentation or an archive (such as a public ledger that links to digital objects) constitute a sense of historical time.

[56] As Ricoeur also suggests, modern technologies render "time" itself abstract: the machines that measure time, such as digital clocks but the blockchain as well, enable an "abstract representation of time" (Ricoeur 1983, p. 63).

(Ricoeur, 1983, p. 128). If we consider explanation of historical facts, Ricoeur argues, we ask for a necessary condition and not for a sufficient, law-like condition. For example, if we ask: "how was it possible that Austria-Hungary declared war on Serbia in 1914?" we might ask for a necessary condition, which was the assassination of Franz Ferdinand. However, the assassination is not a sufficient condition for the declaration of war. In order to explain historical facts, we need to take into account the teleology that guides the events that make up history, which is grounded in the world of action of individual people. History is concerned with the realm of action, but action placed in a society that has already been configured through narrative activity. For this reason, we can only explain a historical fact by means of retrodiction (putting past events into a series), not by prediction. This claim is important for our understanding of technological mediation, for it enables us to argue against technological determinism. If technological configurations mediate human narrative time, they only provide for necessary reasons and not for sufficient reasons for technologically mediated action. Things could have turned out differently.

On the one hand, technological configurations can engage with the world of action, by organising actual characters and event in a plot. Some types of video games, for instance "open-world role-playing games", offer striking illustrations of this form of configuration, because they allow a player to engage with the world of action of (fictional) characters and events. As such, they configure imaginative variations akin to those configured by works of fiction. On the other hand, Ricoeur explains that narrative structures can be made increasingly abstract by means of construction of so-called quasi-entities that are based on first-order entities, which are actual characters and events (Ricoeur 1983, p. 181). This is done in historical narrative by constructing quasi-entities such as "Germany" and "the Battle of Warsaw", which *stand for* actual characters (e.g. the German minister of foreign affairs) and events (e.g. soldiers performing a charge) and "act" or are "acted in" in ways similar to actual characters and events. For instance, a soldier might be said to participate in "the battle of Warsaw", and "Germany" might be said to have signed the Versailles treaty document. These quasi-entities can be organised in a quasi-plot, which is a plot that is removed from direct interaction of actual characters and events. We argue that similar types of constructions are present in modes of technological configuration. Technologies such as electronic networks and exchanges are quasi-entities that do not directly denote actual people or events. Nonetheless, any attempt aimed at explaining plots that involve these structures will require an activity of referring-back to first order entities: it will require narratives about actual characters that act within a first-order plot. Hence, we argue that abstracting technological

configurations remove themselves from the realm of action by configuring quasi-characters and quasi-events in a quasi-plot (Ricoeur 1983, p. 181).

To illustrate the distinction of distancing in technological configuration, in terms of abstraction from the world of action, we can consider the electronic trading in "derivatives". The construct of a future – a specific type of derivative – was already used in 1730 in Japanese rice markets (Takatsuki, 2008) and has evolved into one of the major financial products that are currently traded in the global digital economy (Pryke & Allen, 2000). A future is a contract with a price agreement between two parties, based on the buying or selling of an asset at a specific moment in the future. For instance, a rice farmer in Japan might agree with a derivative trader that she is guaranteed to sell his future harvest at a certain price. This allows the trader to bear some of the risk of the rice farmer - the harvest might fail, in which case the income of the farmer is still guaranteed - and at the same time make a profit on it by spreading out her own risk amongst multiple farmers. A derivative is a typical example of a linguistic construct that abstracts from the world of action (the farmer who tries to survive by harvesting from his lands) by constructing quasi-entities (e.g. derivative exchanges, credit risks). These are quasi-entities because similar to quasi-entities in historical narrative they configure our understanding of the world without disclosing their mediation of the world of action. That is, when a local derivative exchange goes down, a referral-back to the world of action needs to be made in order to explain the event (for instance referring to a severe drought that destroyed the harvests of all rice farmers who were securing their livelihood through derivative contracts). As Pryke and Allen argue, in our contemporary world in which derivatives as linguistic contracts have merged with digital technologies, they reflect a cultural shift that is an "outcome of a transformation in our experience of everyday temporal and spatial co-ordinates" (Pryke & Allen, 2000, p. 282). Derivatives have become technologically mediated contracts that automatically respond to changes in the quasi-plots they configure (such as price fluctuations, risk indicators), thereby increasingly obscuring the way in which they configure the world of action. However, as the financial crisis in 2008 showed, a referral-back to the world of action, of families losing their income and therefore being unable to pay their mortgage, was necessary to explain how the abstract complexities of derivative trades could have contributed to a global financial catastrophe (Hellwig, 2008).

5.5 Chapter Summary

In this chapter, we used Ricoeur's narrative theory to construct a theory of technological mediation. We established four hermeneutic concepts that make explicit how technological emplotment mediates our social world (see table 5 below). First, we considered *textuality,* which denotes a distinction between technologies that are passive elements of prefigured time and technologies that actively configure characters and events in a plot. Second, we considered *literacy,* which denotes a distinction between technologies that are accessible, meaning that many people have access to their first-order emplotment, and technologies that are inaccessible. Third, we considered *temporality,* which denotes a distinction between technological configurations of dynamic, lived time, and of rigid, public time. Fourth, we considered *distancing,* which denotes a distinction between technological configurations that engage with and abstract from the world of action. Before we proceed to the next chapter, it is important to stress that no direct *normative* judgement can be derived from the hermeneutic distinctions we established. Each type of technology and technological configuration has to be understood within the frame of a certain technical practice. For instance, for certain monetary practices it is beneficial to have processes of technological configuration that abstract from the world of action, while these same processes might be detrimental to practises in care-settings. In the next chapter, therefore, we will embed the theory of narrative technologies in a theory of technical practice that will allow us to draw a connection with the virtue ethics of technology approach as established in the previous chapter.

Table 5: An overview of (1) the four central concepts of our theory of narrative technologies, (2) the distinctions between technologies and modes of technological configuration denoted by these concepts and (3) particular technologies that illustrate extremes captured by the distinctions, which are also used in the text.

Concept	Distinction	Illustration
Textuality	Between technologies that actively emplot narrative understanding and that become and remain passive.	Social media actively configures narrative understanding, while a bridge becomes a passive element of prefigured narrative understanding.
Literacy	Between technologies that are accessible and that are inaccessible.	A hammer is accessible, while a military drone is inaccessible.
Temporality	Between technological configurations that bring about a rigid, public sense of time, and a dynamic sense of time.	A blockchain technology configures the rigid (public) dimension of time, while a windmill configures the dynamic dimension of time.
Distancing	Between technological configurations that engage the world of action, and abstract from	A video game engages the world of action, while a derivative trading system abstracts

Concept	Distinction	Illustration
	it.	from it.

5.6 References

Aristotle. (1902). *Poetics*. (S. H. Butcher, Ed.) (The Poetic). London: Macmillan and Co.

Bobb, C. (2011). Paul Ricoeur's Hermeneutics Between Epistemology and Ontology. *Logos & Episteme*, *2*(3), 335–345.

Borisenkova, A. (2010). Narrative Refiguration of Social Events: Paul Ricoeur's Contribution to Rethinking the Social. *Études Ricoeuriennes / Ricoeur Studies*, *1*(1), 87–98. http://doi.org/10.5195/errs.2010.37

Carr, D. (1986). Narrative and the real world. *History and Theory, 25*(2), 117-131. http://doi.org/10.2307/369249

Coeckelbergh, M., & Reijers, W. (2016). Narrative Technologies: a philosophical investigation of the narrative capacities of technologies. *Human Studies*, *39*(3), 325–346. http://doi.org/10.1007/s10746-016-9383-7

Davidson, S. (2006). Thinking After Ricoeur. *Journal of French and Francophone Philosophy*, *16*(1/2), 1–8. http://doi.org/10.5195/jffp.2006.179

Dowling, W. C. (2011). *Ricoeur on Time and Narrative: An Introduction to Temps et récit*. Notre Dame, Indiana: University of Notre Dame Press.

Dowling, W. C. (2011). Ricoeur on Time and Narrative - An Introduction to Temps et recit. Notre Dame, Indiana: University of Notre Dame Press.

Eshet-Alkali, Y. (2004). Digital Literacy: A Conceptual Framework for Survival Skills in the Digital Era. *Educational Multimedia and Hypermedia*, *13*(1), 93–106.

Feenberg, A. (1999). *Questioning Technology*. New York: Routledge.

Gangas, S. (2004). Axiological and normative dimensions in Georg Simmel's philosophy and sociology: a dialectical interpretation. *History of the Human Sciences*, *17*(4), 17–44. http://doi.org/10.1177/0952695104048071

Goodfellow, R., & Hewling, A. (2005). Re-conceptualising culture in Virtual Learning Environments: from an "essentialist" to a "negotiated" perspective. *E-Learning and Digital Media*, *2*(4), 355–367. http://doi.org/10.2304/elea.2005.2.4.355

Gransche, B. (2017). The Art of Staging Simulations: Mise-en-scene, Social Impact and Simulation Literacy. In *The Science and Art of Simulation I: Exploring-Understanding-Knowing* (pp. 33–50). Heidelberg: Springer International Publishing. http://doi.org/10.1002/hast.551

Hassam, A. (1990). "As I Write": Narrative Occasions and the Quest for Self-Presence in the Travel Diary. *ARIEL: A Review of International English Literature*, *21*(4), 33–47.

Heidegger, M. (1977). *The Question Concerning Technology*. New York: Garland Publishing.

Heidegger, M. (1996). *Being and Time*. (J. Stambaugh, Ed.). New York: State University of New York Press. http://doi.org/10.1017/CBO9781107415324.004

Hellwig, M. (2008). Systemic Risk in the Financial Sector: An Analysis of the Subprime-Mortgage Financial Crisis. *MPI Collective Goods Preprint*, *2008*(43), 77. http://doi.org/10.1007/s10645-009-9110-0

Ihde, D. (1971). *Hermeneutic Phenomenology: The Philosopher of Paul Ricoeur*. Evanston: Northwestern University Press.

Ihde, D. (1979). *Boston Studies in the Philosophy of Science: Technics and Praxis* (Vol. 33). Dordrecht: D. Reidel Publishing Company.

Ihde, D. (2009). *Postphenomenology and Technoscience*. New York: Sunny Press.

Jasanoff, S. (2015). Future imperfect: Science, Technology and the Imaginations of Modernity. In S. Jasanoff & S. Kim (Eds.), *Sociotechnical imaginaries and the fabrication of power* (Vol. 1, pp. 1–49). http://doi.org/10.1017/CBO9781107415324.004

Jenkins, H. (2003). Game design as narrative architecture. *Response*, *44*(3), 1–15.

Jervolino, D. (1990). *The Cogito and Hermeneutics: The Question of the Subject in Ricoeur*. (G. Poole, Ed.). London: Kluwer Academic Publishers.

Jones, C. F., Loh, S.-L., & Sato, K. (2013). Narrating Fukushima: Scales of a Nuclear Meltdown. *East Asian Science, Technology and Society*, *7*(4), 601–623. http://doi.org/10.1215/18752160-2392860

Kaplan, D. M. (2002). *The Story of Technology*. Retrieved from http://www.pages.drexel.edu/~pa34/The Story of Technology.pdf

Kaplan, D. M. (2003). *Ricoeur's Critical Theory*. *SUNY series in the philosophy of the social sciences*. New York: SUNY Press. http://doi.org/10.1017/CBO9781107415324.004

Kaplan, D. M. (2006). Paul Ricoeur and the Philosophy of Technology. *Journal of French and Francophone Philosophy*, *16*(1/2).

Kemp, P. (1996). Ricoeur between Heidegger and Levinas: Original Affirmation between Ontological Attestation and Ethical Injunction. In R. Kearney (Ed.), *Paul Ricoeur: The Hermeneutics of Action*. London: SAGE Publications.

Lai, W. (1984). Symbolism of Evil in China: The K'ung-chia Myth Analyzed. *History of Religions*, *23*(4), 316–343.

Lewin, D. (2012). Ricoeur and the Capability of Modern Technology. In *From Ricoeur to Action*. London: Continuum International Publishing Group.

Lidwell, W., Holden, K., & Butler, J. (2003). *Universal Principles of Design*. *Design*. Gloucester: Rockport Publishers. http://doi.org/10.1007/s11423-007-9036-7

Mateas, M., & Sengers, P. (1999). Narrative intelligence. *In Proceedings AAAI Fall Symposium on Narrative Intelligence*, 1–10. http://doi.org/10.1016/S0890-4065(99)80003-6

Mcreynolds, E., Lerner, A., Scott, W., Roesner, F., & Kohno, T. (2015). Cryptographic Currencies from a Tech-Policy Perspective : Policy Issues and Technical Directions. In *Financial Cryptography and Data Security: FC 2015 International Workshops, BITCOIN, WAHC, and Wearable. Vol. 8976*. San Juan: Springer.

Patterson, D. M. (1990). Law's Pragmatism: Law as Practice & Narrative. *Virginia Law Review*, *76*(5), 937–996.

Philipps, L. (1996). Discursive deficits: A feminist perspective on the power of technical knowledge in fiscal law and policy. *Canadian Journal of Law & Society*, *11*, 141–176. http://doi.org/10.1017/CBO9780511609800.010

Pryke, M., & Allen, J. (2000). Monetized time-space: derivatives – money's "new imaginary"? *Economy and Society*, *29*(2), 264–284. http://doi.org/10.1080/030851400360497

Reijers, W., & Coeckelbergh, M. (2016). The Blockchain as a Narrative Technology : Investigating the Social

Ontology and Normative Configurations of Cryptocurrencies. *Philosophy & Technology*, *7*. http://doi.org/10.1007/s13347-016-0239-x

Ricoeur, P. (1967). *The Symbolism of Evil*. Boston: Beacon Press.

Ricoeur, P. (1971). The Model of the Text : Meaningful Action Considered as a Text. *Hermeneutics and Critical Theory*, *38*(1), 316–333.

Ricoeur, P. (1980). Narrative Time. *Critical Inquiry*, *7*(1. On Narrative), 169–190.

Ricoeur, P. (1983). *Time and Narrative - volume 1*. (K. McLaughlin & D. Pellauer, Eds.). Chicago: The University of Chicago. http://doi.org/10.2307/1864383

Ricoeur, P. (1986). *Lectures on Ideology and Utopia*. New York: Columbia University Press.

Ricoeur, P. (1988). *Time and Narrative - volume 3*. Chicago: The University of Chicago.

Ricoeur, P. (1992). *Oneself as Another*. (K. Blamey, Ed.). Chicago: University of Chicago Press.

Ricoeur, P. (2004). *The Rule of Metaphor: the Creation of Meaning in Language*. London: Routledge.

Ricoeur, P. (2014). The Later Wittgenstein and the Later Husserl on Language. *Études Ricoeuriennes / Ricoeur Studies*, *5*(1). http://doi.org/10.5195/errs.2014.245

Romele, A. (2018). Imaginative Machines. *Techné: Research in Philosophy and Technology, 22*(1), 98-125. http://doi.org/10.5840/techne201791369

Rompay, B. Van, Preneel, B., & Vandewalle, J. (1999). The digital timestamping problem. In A. Barb, E. C. van der Meulen, and P. Vanroose (eds.), *Proceedings of the 20th Symposium on Information Theory in the Benelux*, Werkgemeenschap voor Informatie- en Communicatietheorie, pp. 71–78, 1999

Simmel, G. (1900). *The Philosophy of Money*. (D. Frisby, Ed.) (3rd ed.). New York: Routledge Classics, 1978.

Siren, A., & Hakamies-Blomqvist, L. (2005). Sense and sensibility. a narrative study of older women's car driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(3), 213–228. http://doi.org/10.1016/j.trf.2005.04.008

Takatsuki, Y. (2008). The formation of an efficient market in Tokugawa Japan. *ISS Discussion Paper Series*, *F-143*.

Van Den Eede, Y. (2010). In Between Us: On the Transparency and Opacity of Technological Mediation. *Foundations of Science*, *16*(2–3), 139–159. http://doi.org/10.1007/s10699-010-9190-y

6 Narrative Virtue Ethics of Technology[57]

6.1 Introduction

In chapters 4 and 5, we presented the outline of virtue ethics of technology and a complementary conceptualisation of technological mediation that is thoroughly informed by Ricoeur's narrative theory. What is still missing at this point is a move from hermeneutic philosophy to a theory of "practical reason" (Kaplan, 2003, p. 45) that can underlie the novel method for practising ethics in R&I. To achieve this, we need to reconnect the understanding of technological emplotment with the world of action, in which we can speak of technologically mediated acting or suffering in accordance with virtue. In alignment with this task, the aim of this chapter is to draw together virtue ethics of technology and the narrative technologies approach, in order to construct an integrated approach that we will call *narrative virtue ethics of technology*. Crucial to this act of drawing together will be the notion of *praxis*, the Aristotelian term for human action, with which both approaches are concerned. For virtue ethics, an understanding of praxis is crucial because it is what cultivates the relevant states of character. For narrative technologies, praxis is central because, as Ricoeur (1980) argues, the plot in narrative is essentially an imitation of praxis.

A direct consequence of this focus on praxis is that our approach will not just be concerned with technologies, understood as discreet objects and artefacts that populate our life world, but also with the more comprehensive notion of technical practice (in which technical things are involved) that cultivates the virtues. For instance, we are not merely interested in the consequences of the use of a social media platform for the psychological well being of its users, but more essentially with the way it mediates the technical practice of "online conversation", which has particular standards of excellence linked to it. Similarly, we are not merely interested in the potential consequences of a piece of software used for financial transactions on the financial freedom of consumers, but more importantly in the way software engineers engaging in the practice of "coding" manage to link their activity to the common good of their society and make prudent decisions regarding this awareness. Our approach thereby aims to answer questions such as: what are the requirements for a technical

---

[57] A part of this chapter has been adopted from the paper: Reijers, W., (2018). Beyond Postphenomenology: Ihde's Heidegger and the Problem of Authenticity. *Human Studies, (Forthcoming)*.

practice to be engaged in virtuously? How are these requirements mediated by standards of excellence and life plans? And how do they relate to a narrative conception of the good life?

This approach follows from the argument that we developed in chapter 4, namely that the technological "setting" in which human action takes place should be taken into account when considering virtue ethics of technology. First, this means that technologies mediate the narrative structures through which human beings interpret their practices. For instance, in our discussion of the Milgram experiment we argued that the use of a technical device to remotely administer electric shocks to a human being mediated the practice of "harming a victim". Second, it means that such technical practices can either promote or obstruct the cultivation of the virtues. In the particular case of the Milgram experiment, the technical practice in which research participants were involved for instance obstructed their courage to protest. The virtues are therefore conceived of as stable but not static, which implies that even though at any given point in time they provide an agent with a stable disposition this disposition can be promoted or obstructed through engagement in certain technical practices, particularly through ones that are frequently engaged in. A striking example in this regard is people's frequent engagement with social media in order to maintain friendships (cf. Vallor, 2016, p. 159). Through the technical mediation of practices such as conversing or discussing, social media platforms allow their users to either cultivate or obstruct their virtue of friendship.

In the wake of our account of narrative technologies, we are left with the question: how does the process of technological emplotment mediate human actions, which are the proper concern of any *ethics*[58]? To formulate an adequate answer to this question, we will have to show how the narrative structures involved in technological emplotment give rise to the relevant type of human activity, namely technical practice that cultivates the virtues. We cannot simply take for granted that a technical practice is something like a human activity involving the use of technical objects, for this will tell us nothing about the roles of narrative and of virtue with regard to this activity. Instead, we need to gain understanding of how both virtue (chapter 4) *and* narrative (chapter 5) fundamentally relate to technical practice. To take on this challenge, we will first construct an argument that shows how technical practice, virtue and narrative are related, which leads us to our further discussions of Ricoeur's work on practice and ethics in *Oneself as Another* (Ricoeur, 1992). Second, we present Ricoeur's theory of practice according to his accounts of practices, life plans and narrative unity of life,

---

[58] For, as Aristotle already argued: the end of ethics – as belonging to political philosophy – is action (praxis) (NE, 1.3, 1095a4-6, trans. Irwin).

and link this theory to our account of technological emplotment. Third, we present Ricoeur's "little ethics" that explicates the "ethical aim", which provides technical practices with a notion of the good life, with and for others, in just institutions.

6.2 Technical Practice that Cultivates the Virtues

In this first section, we present an argument that will serve two purposes: (1) to arrive at a first determination of technical practice that cultivates the virtues according to what Ricoeur calls the "nesting of finalities" (Ricoeur, 1992, p. 178), and (2) to elucidate the role of narrative in a second determination of this notion, which leads to the formulation of the "narrative unity of life" as an unstable mixture of narration and actual experience (Ricoeur, 1992, p. 162). As such, this section functions as a bridge between the previous two chapters and the work that will be done in the remainder of this chapter. Even though we will eventually use Ricoeur's theory of practice in *Oneself as Another* (Ricoeur, 1992), we initially depart from Heidegger's conceptualisation of technical practice in his early work, based on his lectures on concepts of Aristotelian philosophy (Heidegger, 2009) and his Magnus Opus *Being and Time* (Heidegger, 1996). We start with Heidegger because he, unlike Ricoeur, offers an account of technical practice that incorporates a discussion of technologies, or rather of technical things, and because he directly engages in an elaborate discussion of virtue in his work on Aristotle. That is, even though Heidegger distances himself from ethical questions in *Being and Time* (cf. Hodge, 1995)[59], we argue that he does offer an illuminating account of virtuous action and virtuous agents. Additionally, an engagement with Heidegger's notion of technical practice and Ricoeur's critique of Heidegger's philosophy will lead us to understand the indispensible role of narrative.

To interpret Heidegger's notion of technical practice, we will connect his discussion of technics in *Being and Time* with his discussions of technics and of virtue in his 1924 work *Basic Concepts of Aristotelian Philosophy,* which presents his Marburg lectures that focused on his interpretation of Aristotle's works, notably of the *Rhetoric* and *Nicomachean Ethics*. In doing so, we make an important assumption in our reading of Heidegger: namely that Heidegger's work on technology (including his "tool analysis") in *Being and Time* largely derives from his earlier work on Aristotelian philosophy. This assumption has been subject to vast academic debates and our intention in this chapter is not to address these directly. Suffice it to say that we largely side with Brogan, in arguing that we should assign a pivotal

role to Heidegger's interpretation of Aristotle in the development of *Being and Time* (Brogan, 2005, p. 3). The fundamental insight that Heidegger gained from Aristotle concerns the recognition of the manifoldness, or - according to Brogan - more accurately the "twofoldness" of being (Brogan, 2005, p. 20). This twofoldness of being comes to the fore in the many conceptual oppositions Aristotle constructs to characterise phenomena (e.g. *energeia - dunamis, Being – non-Being*), but especially in the two ways in which we *think* being, namely: in terms of beings (e.g. the sun is a star) and in terms of being, or mode of being of Dasein (e.g. being anxious). In a philosophical investigation, this is denoted by the distinction between the so-called ontic and ontological, which always implicate one-another (Brogan, 2005, p. 126).

In line with the above, we argue that the structural discussion of Aristotelian concepts, notably those pertaining to technê and virtue, in Heidegger's 1924 lectures carries over into Division One of *Being and Time*. This implies that we treat Heidegger's discussion of virtue, and especially the virtue of phronesis, as leading to his conception of care (*sorge*)[60] (Heidegger, 1996, p. 169), and his discussion of eudaimonia as leading to his conception of authenticity (*eigentlichkeit*)[61] (Heidegger, 1996, p. 269). This allows us to use Heidegger to formulate a determination of technical practice that cultivates the virtues. However, we should note that from the outset our approach cannot fully be aligned with Heidegger's understanding of technical practice. Especially in his later work *Question Concerning Technology* (Heidegger, 1977), Heidegger makes clear that the way to attain a free (authentic) relation to technology is devoid of connections with public, political life, and thereby rules out any meaningful mediation of technical practices by for instance political interventions in the design process or the formulation of codes of conduct. Heidegger regarded technical activity as inherently inauthentic, and considered the way towards authenticity to be stripped of political significance (i.e. of life in the polis). In order not to align ourselves with this view, will keep a critical distance in discussing Heidegger's notion of technical practice and will eventually go beyond it with the help of Ricoeur's critique of Heidegger's phenomenology (see section 6.2.4). The argument that follows consists of four steps:

---

[60] Brogan supports this interpretation, stating: "Heidegger translates phronesis as Umsicht (circumspection). He also, at least implicitly, offers Sorge (care) as another translation" (Brogan, 2005, p. 16).
[61] Taminiaux supports this interpretation, stating: "Sophia is the pure contemplation by means of which the Greek philosopher who experiences the *bios theoretikos* is immortalized or reaches eudaimonia, a word which Heidegger translates without hesitation as "authenticity" (*Eigentlichkeit*)" (Taminiaux, 1992, p. 7 - original emphasis).

1. Practice engaged in virtuously can be initially characterised as *being awake and succeeding* in acting, which requires a virtuous agent to have the *good* of a practice at her disposal, which lies in its *telos*.

2. The telos of technical practice that cultivates the virtues is two-fold: belonging to a technical thing in its own right *and* pointing beyond the thing, towards a *manifoldness of concern*. This manifoldness of concern implies a nesting of the *in-order-to* of a technical practice with regard to its *for-the-sake-of-which*, of living together in a political community.

3. To respond to the indeterminacy of the good, the for-the-sake-of-which leads to a second determination of technical practice that cultivates the virtues, namely it having to be *authentic*, meaning that its telos lies in the *being-completed* of *Dasein*, which Heidegger grounds in *being-towards-death*.

4. Along with Ricoeur, we criticise Heidegger's notion of authenticity by arguing that (1) being-towards death is an arbitrary grounding of the telos of Dasein, which has equally valid competing accounts, and that (2) phenomenology ultimately cannot solve the *aporias* of time that lead to the conceptualisation of authenticity. This leads Ricoeur to ground authenticity instead in the *narrative mode*.

After having traversed the argument, we will arrive at a determination of technical practice that cultivates the virtues that justifies the move towards Ricoeur's theory of practice and discussion of the ethical aim. That is, we will characterise this central notion as follows: acting (1) awake and succeeding, (2) nesting the in-order-to of the activity with regard to its for-the-sake-of-which, (3) and doing so authentically, which means (4) linking it to the ethical aim that is understood in the narrative mode. Below, we will elaborate on each of these four steps in the argument and explain the meaning of the technical terms that the reader might not be familiar with.

6.2.1   (1.) Virtuous activity as being awake and succeeding

We encounter the notion of virtue (*arête*) in Heidegger's recently translated lectures on basic concepts of Aristotelian philosophy (Heidegger, 2009)[62]. It is important to note that Heidegger's mode of analysis in his early works is similar to Aristotle's, which means that he in each case starts from an analysis of "everydayness", or the testimony of how people

---

[62] Heidegger initially interprets Aristotle's concept of virtue as: a determinate "possibility to be resolved to…" act in a definite moment (Heidegger, 2009, p. 97).

engage in their everyday practices, and moves from this testimony to an existential analysis of ways of being of "Dasein"[63], by which Heidegger designates human existence. At first sight, Heidegger's analysis of a virtuous activity seems to be close to MacIntyre's explication of a practice that we discussed in chapter 4, as the type of activity that aims at goods that are internal to it. Like MacIntyre, Heidegger takes the example of a virtuous musician to illustrate the meaning of virtue:

> "The arête (virtue) of the flute player consists of having the possibility of flute playing at his disposal in a distinctive sense. Such a way of being and living can, however, sleep itself away in a certain sense. One can be competent, and still sleep one's life away. If this way of having the genuine being-possibility at one's disposal is to be an agathon [good], then it must be in the mode of being-awake, and it must itself fulfil the possibility of having it at one's disposal, praxis" (Heidegger, 2009, p.63).

Heidegger proposes two criteria for virtuous activity: (1) being-awake and (2) succeeding (Heidegger, 2009, p. 60)[64]. These criteria derive from our everyday understanding of a person acting virtuously, which in Heidegger's later work is characterised as "anticipatory resoluteness" (Heidegger, 1996, p. 279). Heidegger explains that we can call someone a "competent" (virtuous) fellow (for instance, someone being a competent musician) but that in his virtuosity, this person can sleep through his being-there and can fail to succeed, which according to Heidegger is definitely *not* an instance of virtuous activity.

Even though virtue *can* therefore be involved in the exercise of a craft (technê) (e.g. playing a musical instrument) it has to be distinguished from the mere application of a technique. Moreover, we have to distinguish the training of a technique from the repeating of a serious, virtuous performance (Heidegger, 2009, p. 127). To illustrate this difference, consider the example the technical practice of driving a car. When learning the how to drive a car through training, which includes techniques such as the correct shifting of the gears, one is expected to make errors initially and to learn from those errors towards perfecting the

---

[63] Heidegger captures the difference between testimony of our everyday existence and the existential analysis that follows from it by coining the terms "existentiell" and "existential" (Heidegger, 1996, p. 11). "Existentiell" pertains to our initial understanding of everyday existence, which is given in "ontic" terms, which means that "being" is understood according to "beings" (e.g. "this house consists of walls, windows, and a door"). In contrast, "existential" pertains to an understanding of the structure of the being of "Dasein", which is the being of human beings in "ontological" terms, which means that "being" is understood according to "modes of being" (e.g. "this mood has the character of being-afraid").

[64] Heidegger also specifies that virtue is related to habit (*ethos*) because of its temporal dimension, being characterised by the "stretching across time" (Heidegger, 2009, p. 200).

technique. However, when practising driving on the road in the environing traffic, one is not merely expected to apply the proper technique but to *be awake,* in this sense of having a determinate possibility to be resolved to act in a certain way in a definite moment (for instance, hitting the brake when a pedestrian suddenly crosses the street, while simultaneously being aware of the car approaching at the back), and to succeed in each case, meaning that for instance hitting the pedestrian as an error is in no case praiseworthy – even if one learns from it. We can therefore say that driving virtuously entails a sense of *seriousness,* which Heidegger confirms by stating that "seriousness is expressed by arête [virtue]" (Heidegger, 2009, p. 121). In line with Aristotle, Heidegger claims that we have to connect virtue with the mean (*meson*), which entails that, as a mode of being, virtue is "maintaining the mean". Heidegger approaches the mean initially through *pragmata*, through "things", not insofar as things in general have an "average" mode of being, but insofar things mean something to human-beings (they "matter") in the sense of being "not too much nor too little" (Heidegger, 2009, p. 125).

But *how* can a person engage in an activity virtuously? To answer this question, Heidegger argues that a one's virtue consists in having a good (*agathon*) at one's disposal in undertaking an activity. A good always refers to a limit of praxis (Heidegger, 2009, p. 55). For instance, we could state that the "limit" of the practice of "curing" lies in "health", which in common terms means that once a doctor has made a patient healthy again, that constitutes the limit of his practice of curing. This notion of a limit should not be misunderstood by for instance treating it analogous to the limit of a spatial object (e.g. "the exosphere is the upper limit of the atmosphere"). Instead, this limit should be understood as a "being-completed" of the praxis, as a mode of being, which is captured by the Aristotelian concept of *telos* (Heidegger, 2009, p. 56)[65]. Crucially, this being-completed is to be understood in temporal terms, for "the time for what exists in the moment, is completed in the sense that outside of this time there is no further bit of time to come that also constitutes that thing" (Heidegger, 2009, p. 56).

6.2.2   (2.) Technical practice and the nesting of finalities

Now we have an initial determination of the meaning of practice that cultivates the virtues we need to make the transition to the determination of *technical* practice, in which technical things are involved. This leads us to consider the relation between the Aristotelian virtues of

---

[65] Heidegger therefore insists that telos should not be understood according to the way it is commonly translated, as "aim" or "purpose" (Heidegger, 2009, p. 57).

thought of technê (craft knowledge) and of phronesis (prudence). In his discussion of Aristotelian concepts, Heidegger initially seems to align himself with the interpretation that technê is distinct from phronesis and praxis, stating that the telos (the being-completed) of concern of a shoe by a shoemaker (a craft) is distinct from the telos of concern of "going for a walk", which has its being-completed included in the way of performing the activity (Heidegger, 2009, p. 50). However, he immediately questions this distinction by stating that "there appears a manifoldness of concerns, and their relation to one-another [of one concern vis-à-vis another] is questionable" (ibid). He mentions the raising of horses for military service, as standing in relation to the saddle equipment made by the saddler, and to the tanner who prepares the leather. These technical activities, in turn, stand in relation to war planning and the administration of war for being-with-one-another in the city (*polis*) (Heidegger, 2009, p. 50). In *Being and Time*, this "nesting" of activities belonging to the crafts into the praxis of human beings is made explicit in the following way:

> "The thing at hand which we call a hammer has to do with hammering, the hammering has to do with fastening something, fastening something has to do with protection against bad weather. This protection "is" *for the sake of* providing shelter for Dasein, that is, for the sake of a possibility of its being" (Heidegger, 1996, p. 78 - emphasis added).

This account of a technical practice introduces two relations, namely the "*in-order-to*" (Heidegger, 1996, p. 65), which points at the relation of one thing being done in order to do something else (e.g. hammering in-order-to fasten something), and the "*for-the-sake-of-which*" (Heidegger, 1996, p. 78), which makes explicit what the activity is for, considered in relation to Dasein. The "totality of relevance" (Heidegger, 1996, p. 78) that is implied in this account relates to what Heidegger in his earlier work on Aristotle calls the "*manifoldness of concerns*". He aims to make clear that the "manifoldness of concerns that constitute the being-there of human beings as being-with-one-another" has a *limit* (Heidegger, 2009, p. 51) and that this limit is constituted by a *telos* (a being-completed) for its own (Dasein's) sake, which is the aim of the craft of *politics*. Heidegger thereby explains how technical activities, such as making a saddle, are nested within a manifoldness of concerns that links these activities with the *polis*, with the being-with-one-another of human beings in a political community. To be sure, a technical object such as a shoe has a telos insofar as it is completed in its own right, but at the same time it points beyond itself, towards a mode of concern that

makes the technical practice of shoe-making possible which eventually links up with the telos of a human being, of Dasein (Heidegger, 2009, p. 63).

Thus, our first determination of technical practice that cultivates the virtues involves acting awake and succeeding, by nesting the in-order-to of the activity with regard to its for-the-sake-of-which of living together in the polis, which constitutes the genuine *telos* of the activity. As we will see at a later point in this chapter (section 2), the nesting of the in-order-to with regard to the for-the-sake-of-which corresponds with Ricoeur's notion of the "nesting of finalities" ("finality" understood as telos), which constitutes the reason for formulating his theory of practice (Ricoeur, 1992). However, for reasons that are to be made explicit below, this initial determination is still insufficient and needs a second determination that relies on the examination of Heidegger's notion of *authenticity*.

6.2.3 (3.) Technical practice and authenticity

In *Being and Time,* Heidegger relates his earlier analysis of virtue with the temporal notion of *anticipatory resoluteness* (Heidegger, 1996, p. 281). That is, the primordial experience of temporality that is at stake in virtuous being, according to Heidegger, derives from the "being-a-whole" (i.e. being-completed), which in turn is experienced in anticipatory resoluteness. To understand the notion of anticipatory resoluteness, we have to confront the central Heideggerian notion of authenticity. This notion – a slightly unfortunate translation of the German word "eigentlich", which can perhaps better be understood as "ownedness" – denotes a mode of being which has a telos that belongs to Dasein itself. Authenticity, for Heidegger, is ultimately linked to the temporal "being-towards-death" (Heidegger, 1996, p. 353), in which he sees the outmost possibility of a human life (the possibility of life's impossibility) and therefore the ultimate telos of Dasein. *Telos*, in this understanding, is "being-toward the *end* in such a way that this end constitutes the genuine there, determining, in a genuine [authentic] way, a being in its presence" (Heidegger, 2009, p. 92 - emphasis added). Heidegger thereby completes the circle: virtuous being-in-the-world, as being awake and succeeding, is an authentic mode of being insofar as it is in the mode of anticipatory resoluteness in the face of being-towards-death, the being-completed - or telos - of Dasein.

Hence, the key to an understanding of technical practice that cultivates the virtues according to Heidegger's philosophy lies in the notion of *authenticity*. As Ricoeur argues, "without the guarantee of authenticity" Heidegger cannot ensure the primordiality of his analysis (Ricoeur, 1988, p. 65). Accordingly, any critique of Heidegger's philosophy would need to proceed from this notion of authenticity. Unfortunately, some of the most influential

philosophers of technology, Ihde (1979), Feenberg[66] (1999) and Verbeek (2005) merely mention Heidegger's notion of authenticity as a side issue, failing to see how it is central to the entire Heideggerian project. This also means that a critique cannot merely target Heidegger's understanding of technics[67] but needs to aim at Heidegger's philosophy as a whole.

We face two problems when considering Heidegger's philosophy as the basis for an ethics of technology: (1) the indeterminacy of the good, and (2) the impossibility to ground the good in being-with-one-another, in moral and political life. First, as we discussed, Heidegger insists that there is no good in itself and that a good can only be understood insofar it constitutes a limit of praxis, as a being-completed of an action. As such, the understanding of the good in an *authentic* mode of being can be carried over to the understanding of the good in an *inauthentic* mode of being, for instance the good in "stealing something". We need to recall that we earlier established an equivalence between Heidegger's use of authenticity and of *eudaimonia,* because both determine the good proper to a human being (Dasein). Heidegger notes that the notion of "good", as in being a good human being, can be "carried over" to the notion of being a "good" thief, insofar a thief can be virtuous in his activity by being awake and succeeding in bringing the activity of "stealing something" to completion (Heidegger, 2009, p. 58). It is therefore not possible, initially, to distinguish between a virtuous human being as such and a virtuous thief, or for that matter a virtuous engineer. We are therefore not yet able to make any normative distinction between "good" and "bad" technical practices, or between "good" and "bad" R&I practitioners.

Second, to address this problem, Heidegger only allows us to retrieve authenticity through the individualistic notion of being-towards-death. This notion is individualistic because Dasein "is *completely* thrown back upon its ownmost potentiality-of-being" (Heidegger, 1996, p. 232 - original emphasis) in being confronted with being-towards-death, and is thereby so-to-say completely on its own in coming to terms with this possibility of death. The inauthentic, in contrast, is constituted in "average everydayness" (Heidegger, 1996, p. 235): in the *flight* from death instead of in its resolute anticipation, which reveals itself in the public realm, in which "the They" (*Das Man*) - the anonymous other - is

---

[66] Thomson shows how Feenberg fails to successfully criticise Heidegger's work on technology, focusing on Feenberg's claims that Heidegger's notion of technology was a-historical, fatalist, and one-dimensional (Thomson, 2005). In fact, Feenberg – like Ihde – never seriously problematizes Heidegger's notion of authenticity.

[67] Technê will refer to the virtue of thought discussed by Aristotle, and technics will connote with Heidegger's notion of "*Technik*", which is often erroneously translated as "techno-logy", and which denotes the way of revealing deriving from technê. Technê and technics can therefore be treated as synonyms.

encountered. Heidegger thereby juxtaposes the everydayness of public life, politics, technology, and society in which Dasein's being-in-the-world has the character of "falling prey" to something (Heidegger, 1996, p. 164), with the authentic being-in-the-world of Dasein, which has to assume the mode of anticipatory resoluteness in the face of being-towards-death in order to attain authenticity.

Now, we have several reasons to be critical of Heidegger's understanding of authenticity. First, Heidegger's puzzling connection between Dasein's authentic being and the destiny of a community[68] (*Volk*) (Heidegger, 1996, p. 352) cannot but call for a reflection on the relation between Heidegger's philosophy and his involvement in Germany's Nazi regime. As Fagenblat argues, Heidegger considered the anti-Semitism of National Socialism as more "authentic" and primordial than biologic anti-Semitism, because he saw the question of role of "World Jewry" as a metaphysical rather than a political one (Fagenblat, 2016, p. 148). Without this being a definite objection to Heidegger's work, it should at the very least be a warning against any uncritical adoption of his idea of authenticity. Second, a more pragmatic reason is that practising ethics of technology inescapably takes place in everydayness, in the public realm of research labs, teams of engineers working on R&I projects, and political interventions in R&I processes. Approaching the complexity of this everydayness from the solitary starting point of being-towards-death seems a highly impractical endeavour. However, even though these first two reasons provide a motivation for searching for alternatives, they are not yet sound philosophical reasons for denouncing Heidegger's notion of authenticity. We will therefore engage with philosophical objections to Heidegger's notion of authenticity that will eventually lead us to Ricoeur's marrying of the notion of technical practice and narrative theory.

6.2.4   (4.) Critique of authenticity leading to narrative

To solve the problems in Heidegger's philosophy, two strategic choices can be made. First, the grounding of authenticity in being-towards-death can be refuted, and an alternative can be put forward. Second, in putting forward an alternative the very possibility of arriving at a non-problematic notion of authenticity within the phenomenological tradition can be questioned.

To explain the first strategic choice without submitting to the second, we discuss Hannah Arendt's notion of authenticity. Diametrically opposed to Heidegger, Arendt (1958)

---

[68] Ricoeur provides a more thorough criticism of this "leap" from individual historicality to common history in *Time and Narrative 3*, stating that this transition is unacceptable because it indicates a homology between communal destiny and individual fate (Ricoeur, 1988, p. 75).

positions the *public* realm as "first and foremost a realm of disclosure" (Villa, 1995, p. 139). Instead of orienting the question of authenticity towards death, Arendt turns towards birth and natality. She sees authentic human activity as action that re-enacts the beginning of "being born" (Arendt, 1958, p. 247), beginning something new, which is always totally unexpected and out of the ordinary. This type of action is furthest removed from production, because it takes people out of their daily routines of labour and work and into the public realm of political action. For Arendt, the fabrication of technics is solely guided by an understanding of utility (Villa, 1995, p. 138). The in-order-to in technical activity, for Arendt, therefore provides the contents of the for-the-sake-of-which. To transcend the world of fabrication that is governed by utility, Arendt puts forward the concept of "political action", which "places the world under a new and unexpected aspect" (Villa, 1995, p. 138). A primary example of political action for Arendt is the American Revolution, which constituted a radical new beginning (Arendt, 1958, p. 228). Such political action demands "the affirmation of plurality [being-with-one-another] and contingency [the unexpected]" as well as "commitment to the public way of being-in-the-world" (Villa, 1995, p. 141). Arendt's notion of virtue accords with this demand, since she argues that virtue (and virtuous action) "has always been assigned to the public realm where one could excel" and "could distinguish oneself from all others" (Arendt, 1958, p. 47). For Arendt, therefore, technical practice – together with production – is put out of sight and substituted by authentic [virtuous] action considered as being awake and succeeding in a radically new and unexpected way in the public realm.

As we argued before, we need an account of authenticity to find a response to the indeterminacy of the good that plagues the notion of technical practice. With Heidegger and Arendt we have different - mutually exclusive - ways to ground the notion of authenticity (being-towards-death/being-born), which lead to different understandings of technical practice that cultivates the virtues, but which of these is most adequate? To answer this question, we return to Ricoeur. The main difference between Arendt and Ricoeur is the latter's denial of the possibility of phenomenology to conclusively solve the *aporias* - the fundamental problems or puzzles - of temporality, and thereby those of authenticity, which leads Ricoeur to make both strategic choices discussed above. Ricoeur's formidable critique of Heidegger's philosophy departs from an original proposition: namely not from stating that the unfinished account of phenomenological time in *Being and Time* is a failure, but asserting that this project is an *outermost* attempt to solve the aporias of time (Ricoeur, 1988, p. 92). Ricoeur first of all recognises the crucial point that for Heidegger's project to succeed: the

"conquest of primordial concepts is" … "inseparable from a struggle against inauthenticity", against everydayness (Ricoeur, 1988, p. 65). However, Heidegger's existential analysis has to be based on a testimony of everyday existence, which leads to a lack of a distinction between these two modes of understanding *if* the grounding of authenticity does not succeed. Consequently, Ricoeur questions Heidegger's conception of authenticity as deriving from being-towards-death, and argues that competing tests of authenticity (amongst which we might count the one put forward by Arendt) are equally legitimate (Ricoeur, 1988, p. 67). More forcefully, he even considers an analysis such as Sartre's, which considers death an interruption of the human potentiality-for-being instead of its being-completed, as equally legitimate.

Ricoeur further compromises Heidegger's notion of temporality on which his understanding of authenticity depends, by criticising Heidegger's unavoidable polemic against the ordinary, everyday concept of time that alone enables a conception of the opposite, authentic concept of temporality. He discusses how Heidegger conceives of the ordinary concept of time as a result of the "levelling off" of the authentic understanding of temporality through datability (situating an event in the indifferent "now"), lapse of time (the interval between "since then" and "until") and the publicness of time of preoccupation (the concealedness of our authentic experience of time behind an average, public understanding of time as a series of point-like "nows") (Ricoeur, 1988, p. 84). In other words, Heidegger argues that by publicly "dating" moments in time, using clocks and calendars, and by understanding time as an interval between dates (e.g. "the meeting took place between 1pm and 2pm"), we cover up our original, phenomenological understanding of time. Ricoeur follows Heidegger in accepting that a derivation of human temporality from a series of point-like "nows" (as science might claim) is impossible, but confronts Heidegger by stating that the reverse – the derivation of cosmic time (the time of the world) from a phenomenological account of temporality and Dasein is "just as impracticable" (Ricoeur, 1988, p. 88).

Contra Heidegger, Ricoeur recognises Aristotle's conception of cosmic time, as having something to do with *movement* in the *world*, as being equally valid as Heidegger's phenomenological notion of temporality, derived from being-towards-death. In fact, Ricoeur's entire endeavour in *Time and Narrative* seems to be a response to the aporias of time that were opened up by his critique of Heidegger. He states that "having something to do with movement and something to do with Care [praxis] seem to me to constitute two irreconcilable determinations [of time] in principle" (Ricoeur, 1988, p. 89). At this point, Ricoeur rehabilitates modern scientific contributions to our understanding of time, those

coming from geology, evolutionary theory and astronomy, which contribute a legitimate "diversification in the meanings attached to the term 'time'" (Ricoeur, 1988, p. 90), which are not reducible to an all-encompassing notion of time as a series of "nows". He states that our ordinary conception of time encompasses *both* existential consideration (i.e. "time flies" because we flee from death) *and* cosmic considerations (i.e. "life is short" because of the immensity of cosmic time). These two ways of grasping time overlap in our ordinary, everyday conception of time, which in turn makes us attentive to the incommensurable polarities they entail. Ricoeur therefore concludes: (1) being-towards-death and world-time disclose an irreducible opposition, (2) these two ways of grasping time overlap, (3) Heidegger's analysis of temporality leads to a diversification rather than a homogenisation of our ways of understanding the grasping of time, and (4) *narrative time* should be considered as the bridge between the phenomenological understanding of time and world-time. Finally, therefore, we have a new understanding of the basis for authenticity that allows us to link technical practice with narrative. That is, given that authenticity is a temporal determination of technical practice that cultivates the virtues, referring to the being-completed of such a practice regarding the genuine telos of Dasein, it is mediated by narrative because narrative offers a response to the aporias of temporality. This fourth part of the argument thereby leads us to link our theory of narrative technologies with the understanding of technical practice that cultivates the virtues.

6.3   An Account of Technical Practice

So far, we have connected the notion of technical practice with the notions of virtue and narrative. That is, a technical practice needs a sense of being-completed to be virtuous, which is captured by the notion of authenticity. Authenticity makes the temporality of a technical practice that cultivates the virtues explicit, which distinguishes it from merely applying a technique, and calls for an understanding of narrative as that which responses to the polarities in our understanding of time. Commonly said, we could state that whenever we engage in a technical practice virtuously, we need to be aware of how this practice contributes to the bringing to completion of human life in the social world, which is done through the narrative mode. The saddle maker needs to be aware (a knowing *why*, not merely a knowing *how*) of how his saddle making contributes to the war planning and life in the polis, and be awake and succeeding in his practice by resolutely anticipating the type of war that will be fought, which

might require him to produce lighter or heavier saddles[69]. We thereby return to MacIntyre's use of narrative, namely as that by which we explain our actions. However, Ricoeur positions narrative more firmly within his theory of practice, by inserting narrative theory at two levels of analysis: at the level of practice, where it makes basic actions and action-chains intelligible *and* at the level of life plans, where it makes practices intelligible with recourse to abstract ideals. In what follows, we turn towards Ricoeur's theory of practice in *Oneself as Another* (1992) that we will reinterpret as a theory of technical practice.

6.3.1 Technical practice in Oneself as Another

Without directly referring to Heidegger, Ricoeur confronts the nesting of the in-order-to in regard to the for-the-sake-of in technical practices head on, in the sixth study of *Oneself as Another* (Ricoeur, 1992). He states that in his account of practice, which explicates the "relation between practices and life plans" the "secret of the *nesting* of finalities, one inside the other", can be found (Ricoeur, 1992, p. 178). He argues that this responds to a central difficulty in the *Nicomachean Ethics,* resulting from the paradox that "each praxis has an end in itself and that all action tends toward an ultimate end" (Ricoeur, 1992, p. 178). In other words, we argue that with his conceptualisation of practice, Ricoeur intends to account for the nesting of the in-order-to in regard to the for-the-sake-of-which in a way that does not depend on Heidegger's notion of authenticity. As we will see below, he does this by introducing a number of core concepts: of practices, life plans, and the narrative unity of life. We will reflect on each of these concepts, how they relate to the narrative mode, and how they together form an integral conceptualisation of technical practice. To understand the premises that underlie Ricoeur's theory of practice, we will first return to Arendt's theory of action.

Reflecting on Arendt's theory of action, which encompasses the human activities of labour, work and (ethical) action, Ricoeur accredits her with two important discoveries (Ricoeur, 1983a). First, Arendt distinguishes between the central human activities of labour, work and action by considering their *temporality* (Ricoeur, 1983, p. 61). That is, labour is a timeless activity, in that it is concerned with nutrition (consumption) and the "ceaseless renewal of life" (Ricoeur, 1983, p. 63). Work or fabrication, in contrast, is the source of durability in time – of things that continue to exist beyond the perishing of their makers. Praxis, finally, as an activity initiated by a human actor, is not durable but essentially only

---

[69] Even though we use an example of an ancient craft, this illustration brings us very close already by what we might mean with "responsible research and innovation".

exists as long as the actor sustains it, after which it disperses[70]. Arendt's three basic human activities therefore denote distinct temporal modes of being. Second, Arendt acknowledges the significance of narrative, by stating that praxis only reveals itself to the storyteller (Ricoeur, 1983, p. 68). In other words, it is only in the narrative mode that praxis and human temporality become intelligible, because it places the singularity of a human action as a fragile initiative within the "interplay of circumstances induced by the web of human relationships" (Ricoeur, 1983, p. 67). It seems that Arendt's theory of praxis has had a profound effect on Ricoeur's conceptualisation of practice, by stressing the factors of temporality and narrative. However, as we discussed above, Ricoeur moves beyond Arendt by challenging her notion of authenticity and therefore arriving at a different theory of practice.

Another significant source of inspiration for Ricoeur has been MacIntyre's theory encompassing concepts of practice, life plan, and moral community. Similar to Heidegger, MacIntyre identifies the good with telos[71] (MacIntyre, 2007, p. 52). Knight, in examining the relation between Heidegger and MacIntyre, claims that according to the latter "following moral rules and cultivating such excellences of character as courage, truthfulness, temperance and justice is what progresses us from our 'untutored' state to the human end or telos of rational self-fulfilment" (Knight, 2008, p. 40). Contra Arendt, MacIntyre draws production and action closer together, even to the point of considering the virtues as analogous to *technai*, productive skills, that are acquired in a similar way as an apprentice acquires a craft such as painting (Knight, 2008, p. 41). For MacIntyre, our understanding of practice does not merely depend on some biological account of human nature, but on an account of humans as social beings that engage in *shared* practices. He therefore argues for the need of a sociological account of practices that defines them through socially established standards of excellence. Because we have these shared standards of excellence for both the outcomes of a technical practice and the activity itself, we can discuss two distinct types of goods that are internal to the practice. For instance, we have standards of excellence for painting as an activity ("she is painting very well") and for the outcome ("that is an excellent painting"), which constitute goods internal to the practice of painting. The goods internal to practices, in turn, are nested with regard to goods internal to a certain type of life, for example of "being a

---

[70] For this reason, we do not "make" or produce history as a work, according to Arendt, but we *act* or practice it.
[71] Knight claims the exact opposite, namely that by identifying the good with telos, MacIntyre was refuting Heidegger's philosophy (Knight, 2008, p. 40). However, as we saw before, Heidegger explicitly states that a good is always a limit of praxis, and as such should be considered telos, as a being-completed. However, the difference between MacIntyre and Heidegger lies in the fact that the first accepted the idea of an end, and therefore a good in itself.

painter" (Knight, 2008, p. 41), which constitutes a life plan. Life plans, then, are nested within what MacIntyre calls a moral community.

Even though Ricoeur sees a "felicitous encounter" (Ricoeur, 1992, p. 158) between his and MacIntyre's theories of practice, he is not uncritical towards the latter. On the contrary, and in line with the criticism we developed in chapter 4, he states that MacIntyre's notion of narrative restricts itself to the stories told in the thick of everyday life without engaging the role of fiction, of imaginative variations, and thereby of *innovation* as opposed to tradition. Indeed, it seems that MacIntyre assigns to tradition a similar role as Heidegger assigns to authenticity. A living tradition is conceptualised as a "historically extended, socially embodied argument" (MacIntyre, 2007, p. 222), that provides the for-the-sake-of-which of people's practices and life plans in a moral community. Since the tradition is thereby the source of standards of excellence that determine the goods internal to practices, it is the differentiating factor between those practices that are inauthentic and those that are authentic. Indeed, it seems difficult to envision in MacIntyre how individuals through imaginative variations - through fiction - might challenge historically embedded standards of excellence, and thereby challenge their moral traditions as well. By means of his narrative theory, Ricoeur manages to surpass this difficulty.

To formulate a theory of practice, Ricoeur starts by positing that "narrative theory occupies a central position […] between the theory of action and ethical theory" (Ricoeur, 1992, p. 152)[72]. In other words, narrative allows us to move from a description of an action (e.g. agent *x* declared *y* in order to *z*) to a prescription of action (e.g. agent *x* should not have declared *y* because of *z*). To revise the concept of praxis, he argues that "a hierarchy of units of praxis must be made to appear" (Ricoeur, 1992, p. 153), each containing its own principle of organisation. This hierarchy should not only provide a series of elements or parts that together constitute the notion of practice, but should also make apparent *how* narrative theory connects the elements with one-another, making one element intelligible in terms of another. For instance, it is in the narrative mode that practice can be made intelligible in terms of a life plan: "I'm *farming* [practice] the land because it has been the calling of my family for

---

[72] As an example, consider the following narrative: "Elizabeth, the nurse, pulls the plug, gathers the family members, looks at the mother and says: "may you find peace". The machines die out and a life passes by and slips away." Disjointed, these simple actions of the form "pulling x", "calling y", "saying z" and their related chains of actions do not add up to an intelligible account of the meaning of the entire plot. Only when considered in its narrative unity, the discordant, heterogeneous elements of "pulling a plug, calling the nurse" as a chronological sequence, "a life passing by" in terms of recollected memories and a life "slipping away" in terms of death as the end of time, form a synthesis. It is therefore the connection between a basic action of "pulling the plug", its relation to the practice of patient care and its ethical significance in relation to a life plan (or in this case, the end of a life plan).

156

generations and I want to sustain this *type of life* [life plan] for the next generation". Thus, the narrative mode allows the agent to explain why he engages in a certain practice in terms of his life plan. The three elements that together constitute the hierarchy of praxis that Ricoeur presents are (1) practices, (2) life plans and (3) narrative unity of life. In relation to narrative theory, practices relate to prefigured time (mimesis$_1$), life plans to configured time (mimesis$_2$) and narrative unity of life to refigured time (mimesis$_3$). Ricoeur discusses these elements as stages that allow us to move from a description of action to an evaluation and prescription of action. It might therefore actually be better to characterise each stage as a transition concept: from action to practice, from practice to life plan, and from life plan to the narrative unity of life, and – as we shall see – to the ethical aim.

1. Action <> Practice

Unlike the other thinkers we discussed (Heidegger, Arendt, and MacIntyre), Ricoeur puts great emphasis on epistemological concerns and does not merely ask the ontological question "what is practice?" but also the epistemological question, "how do we understand practice?" It is actually the latter question that is his point of departure, and which makes Bobb designate Ricoeur's philosophy as offering a "reversed ontology", characterised by starting from an investigation of the way we understand ourselves and the world through texts (and in our account, through interaction with technologies) and moving from that investigation ever closer towards an ontology of the self (Bobb, 2011, p. 344). As a first step in his theory of practice, Ricoeur deals with the transition between the recounting of basic actions and action-chains to the narrating of the "global action" that we designate as a practice. Practice, for Ricoeur, is a rich, unifying concept that – in accord with Heidegger's notion of virtue – has to designate an activity that stretches across time. Simply playing a note or a melody is not a practice, but "playing the piano" as in performing a concerto can be considered a practice. It is therefore only with regard to the latter that we can designate a practice as "virtuous", as being-awake and succeeding. In recounting a practice, a text could for instance state: "the farmer's passion for farming became apparent in the resolute but delicate way he placed the plough in the soil to prepare the land". In this example, the verb "place" designates a basic action, the verb "prepare" an action-chain and the verb "farming" a practice. We will shortly reflect on these three concepts.

A *basic action* designates any action that can be performed without depending on another action. Such a basic action can be a gesture (e.g. waving a hand), a posture (e.g. sitting down), or more basically an elementary corporeal action. Such basic actions enable the

notion of the in-order-to, and it becomes clear at this point how Heidegger's primary example of hammering involves such a structure: hammering, as a basic action of the hand, mediated by the hammer, is performed in-order-to fasten something. The in-order-to relation enables the coordination of basic actions in long *action-chains*, which link actions together in causal series in which the agent performing them can intentionally intervene (Ricoeur, 1992, p. 153). This might for instance be expressed in describing a game of chess, by explicating a strategy in terms of a sequence of basic position-changes of pieces on the board. Ricoeur additionally refers to speech act theory to explain how each basic action receives meaning through the notion of a *constitutive rule* (Ricoeur, 1992, p. 154), which for instance changes a gesture of moving a wooden object across a wooden plane into a move in the game of chess. As such, a constitutive rule results from a declaration in the form "resituating this wooden object from this space to that space *counts as* a move of the queen in the game of chess" (cf. Searle, 2010). Ricoeur connects this concept of constitutive rule with MacIntyre's notion of *standard of excellence*, which opens up the possibility for evaluating an action, though not yet in a strictly ethical sense. For instance, certain constitutive rules allow one to evaluate a chain of basic actions made on a chessboard as either valid or invalid moves in the game of chess.

These rules introduce the *interactive* character of practices, because they only exist in virtue of being learned from others or evaluated by others, not merely by an isolated individual (Ricoeur, 1992, p. 156). To be sure, to interact can also mean to abstain from action (e.g. refraining to help someone), which ensures Ricoeur's theory to encompass both *acting* and *suffering* beings[73]. What we can conclude at this point is that practices should neither be understood merely in terms of linear action-chains, nor according to their nesting-relations determined by constitutive rules, but always in terms of both. In line with this finding, Ricoeur makes explicit that his analysis accords with our earlier discussion of Heidegger's notion of technical practice, by stating that through the integration of linear and subordinating (i.e. nesting) viewpoints of basic actions, action-chains and practices, "no action is only *poeisis* [linear] or only *praxis* [nesting]" (Ricoeur, 1992, p. 176).

---

[73] Through being thus conceptualised, practices link up with the ethical aim in two ways: (1) by providing a basis to consider goods internal to a technical practice, and (2) by providing the "content for the empty form of the categorical imperative" (Ricoeur, 1992, p. 177).

2. Practice <> Life plan

At this point, we need to emphasise that the above-mentioned transition from basic action to practice did not yet provide us with a complete account of practice. On the contrary, Ricoeur stresses that to understand practice properly we have to traverse the stages of life plan and the ethical aim. Ricoeur spends relatively little time discussing the notion of life plan, but this is because it introduces the movement of narrative configuration into the fabric of practice, which was already extensively dealt with in *Time and Narrative* (Ricoeur, 1983b). A life plan is not to be understood as an "entity" with definite properties such as "*a profession*", but as a movement that links practices with the wholeness of a human life. Ricoeur's choice of wording seems slightly unfortunate when he designates life plans as "*vast* practical *units* that make up professional life, family life, leisure time, and so forth" (Ricoeur, 1992, p. 157). In fact, a life plan is not to be considered as a unit, as an entity, but as a particular movement of emplotment, which is not necessarily to be considered "vast". This is the case because a life plan seems to be any movement of emplotment that links practices with the narrative unity of life, which can be either "tiny" (for instance, a narrative about visiting the zoo with one's grandfather on a particular afternoon) or "vast" (for instance, a narrative about being a zoo keeper). Both of these constitute a "life plan" for they configure plots with a beginning, middle and end that additionally are connected with one-another (for instance, one became a zoo keeper because of that vital afternoon visit) without the one being subsumed under the other.

Notwithstanding this slight diversion, we retain the crucial addition provided by Ricoeur to the notion of narrative configuration in constructing his theory of practice. This addition consists in characterising the life plan as a "back-and-forth movement between more or less distant ideals" (Ricoeur, 1992, p. 157). In other words, we could state that we understand a technical practice that we engage in, such as flying a miniature airplane, in terms of more or less distant ideals, such as wanting to become a pilot, which are *mediated* by life plans such as partaking in a miniature airplane competition. In this example, the airplane competition constitutes the plot that connects the practice and its standards of excellence with the distant ideal of being a pilot. It is also at this point that narrative configuration through history or fiction enters the stage, for just as recounting a past experience can connect practices with more or less distant ideals, so can works of historical narrative or literary fiction. For instance, the practice of securing one's ICT infrastructure can connect with a distant ideal of a free and open society mediated by the reading of Orwell's *1984*. The difference between Ricoeur and MacIntyre in this case is that MacIntyre only foresees a

159

"bottom-up" movement, from practices, towards life plans and the narrative unity of a human life while Ricoeur characterises it in accordance with a "twofold principle of determination" (Ricoeur, 1992, p. 158), which consists of the two-fold movements of *ascending complexification* and *descending specification*. Ascending complexification denotes the movement from basic actions and practices towards ideals, for instance starting to play the piano at an early age, by pressing some keys, and moving towards the ideal of becoming a professional piano player. Descending specification denotes the movement from ideals towards practices and basic actions, as in the above-mentioned example of reading *1984*, which enables the practitioner to start from ideals and move towards a change in practice and its related basic actions.

3. Life plan <> Narrative unity of life

After having considered life plans, Ricoeur turns to what MacIntyre designates as the "narrative unity of life" (Ricoeur, 1992, p. 160), the overall life biography that supports the notion of a "good life" and provides life with its ethical tenor. As such, a narrative unity of life fulfils the for-the-sake-of-which that we encountered in Heidegger, which returns the telos of the in-order-to of practices back to being-with-one-another in the polis. This narrative unity is therefore necessary for virtue, for being awake and succeeding in a technical practice, which requires the for-the-sake-of-which. However, Ricoeur objects to the idea that one could speak of something like a coherent and complete narrative unity of life. First, we are never the sole authors of a life, but only ever co-authors, just as we are only co-authors in our interactions with technologies. Second, unlike a narrative, a human life cannot be easily considered as "complete", for its beginning in birth and ending in death are inaccessible for a human to "gather together" in a story. Third, a life can consist of several stories, each with their own sense of ending, without necessarily being unified. Fourth, whereas different works of narrative commonly present incommensurable plots, our life stories are entangled with the plots of other people's life stories. Fifth, our recounting of life only applies to past experiences whereas works of narrative can include narrated accounts of future states of affairs. These arguments lead Ricoeur to claim that the narrative unity of life must *not* be understood as a coherent and complete whole (*as* a completed narrative), but rather as an "unstable mixture of fabulation and actual experience" (Ricoeur, 1992, p. 162) that can move *towards* a sense of completion without ever attaining it. To be sure, narrative mediates the sense of unity of a human life but does not offer a complete account of it.

This conclusion necessitates a move from narrative theory to ethics, because we again encounter the problem that we earlier encountered in Heidegger – of being unable to distinguish the good person from the "good" thief. Consider the practice of stealing something, mediated by life plans such as partaking in one's first armed robbery in a criminal gang, which connects it to the distant ideal of being the boss of a crime syndicate. The in-order-to of stealing something is nested with regard to the for-the-sake-of-which of being the political leader of a group. One can therefore be awake and succeeding in practising stealing, and therefore be virtuous in a certain sense. However, the problematic nature of the activity of "stealing" as a virtuous activity provides us with a reason to want to move beyond this notion of virtue, and to be able to distinguish between technical practices that cultivate virtues leading towards *eudaimonia*, and those that do not. Indeed, being a successful thief does not seem to accord with leading a good life. Ricoeur therefore still needs an equivalent for Heidegger's notion of authenticity that he finds in "the ethical aim" (Ricoeur, 1992, p. 169), which we will discuss in the next section (6.4). For now, we can conclude by mentioning how Ricoeur sees that narrative leads to the ethical aim – or perhaps better the "being-completed" of ethical life. First of all, the experiences recounted in the narrative mode always involve estimations and evaluations that enable actions to be "subject to approval or disapproval and agents to praise or blame" (Ricoeur, 1992, p. 164). In fiction in particular, moral judgement is thereby subjected to imaginative variations that allow practitioners to question their practices and related standards of excellence. Ricoeur connects his theory of practice with the main theme of *Oneself as Another*, being the relation between *sameness*, which can be found in lasting dispositions of character, and *selfhood*, which can be found in keeping one's word despite change; which together constitute personal identity. Even though it is outside the scope of our current endeavour to delve into the breath of this thematic, it needs to be pointed out that the narrative mode mediates between sameness and selfhood, by narrativising the ethical aim of the good life and thereby "giving recognisable features to characters loved and respected" (Ricoeur, 1992, p. 166). Finally, the difference between narrative and ethics lies in the above-mentioned problematic of making normative distinctions, which entails that the narrative mode can go to extremes where sameness and selfhood can be negated. In other words, narratives can make us doubt that there is any ground to the question "who am I?" and consequently any ground to people's moral evaluations and estimations. Ricoeur shows that concrete practice intervenes at this point and binds narrative to our everyday existence, stating that the pragmatic answer to the doubt

engendered by narrative is "here is where I stand!" (Ricoeur, 1992, p. 168). And by means of this practical testimony, praxis leads us back towards the ethical aim.

6.3.2   Technical practice and technological emplotment

Even though our theory of technical practice is still unfinished and requires the final step of considering the ethical aim, we can already draw together the account of practice developed in this chapter and the notion of technological emplotment developed in chapter 5. First, we need to emphasise that our theory of technical practice is *not* distinct from Ricoeur's theory of practice for an important reason. That is, by emphasising the continuum between narrative and technological configuration, we consider both the mediation of texts and technologies as essentially mediating praxis, which consequently demands a unified notion of practice. The adjective *technical* therefore merely redirects our focus from textual mediation towards mediation by technics, without losing the fundamental equivocalness of these processes out of sight. Second, we now understand how technological configuration mediates practice and thereby how it mediates virtue. Crucial in this understanding will be the notions of *ascending complexification* and *descending specification*, as those two movements that link practices through technological configuration with more or less distant ideals. Below, we will integrate these concepts into our theory of technological emplotment and show how they relate to its four central concepts of textuality, literacy, temporality and distancing. We will illustrate this by using the example of *driving a car*, incorporating the constitutive rules and ideals comprised by a policy called the "New Driving". This policy was implemented in the Netherlands in 2006, and aimed at mediating the technical practice of driving a car to make it more environmentally friendly (de Goede & Hoedemaeker, 2009). The policy was implemented by making students in driving schools aware of the importance of driving sustainably and integrating corresponding practical principles of "the new driving" in their driving lessons. Guiding principles that were adopted in this policy were for instance: (1) switch to a higher gear as soon as possible, and (2) when driving 80 kilometres an hour, drive in the 5th gear. This example is chosen because it is a clear illustration of the twofold movement that links basic actions to more or less distant ideals.

In conceptualising a particular practice, **textuality** shows us *where to look*. That is, when considering technical objects that are less textual, such as a bridge, we need to focus on the narratives recounted *about* a particular object to understand how it mediates our practices (mimesis$_1$). For instance, one could recount how a particular bridge has been crucial for starting a transportation company that delivers goods coming from one city to another city.

Such a narrative tells us more about how the bridge mediates particular practices than the design of the bridge as such. A car is different, in that its design co-authors the narrative structures that mediate the practice of driving (mimesis₂). In our example, the gearbox in a car enables certain basic actions that can be linked in an action-chain, for instance shifting back gears when closing in on a junction with red traffic lights. These actions are understood according to certain constitutive rules that are considered to be standards of excellence, for instance the rule: switching down gears and pressing the brake in front of a traffic light counts as a valid response in traffic. The practice of driving a car links these basic actions and action-chains with more or less distant ideals. For instance, it can be linked with the life plan of being a taxi driver, in which case the technical practice of driving is nested with regard to the narrative of a person's professional life.

Literacy, then, shows us *at whom to look*, and to whom a particular process of technological configuration is accessible. The technical practice of driving is remarkable in this regard because it is widely shared across people from different walks of life. As such, it is one of the few technical practices that is relatively difficult to master but that is nonetheless mastered by many people in society with varying capacities and dispositions. When investigating the technical configuration of driving a car, we therefore do not have to limit our discussions to the authority of expert knowledge. Instead, we can consider that the notion of *driving virtuously* is accessible to a vast share of the population. The nesting of finalities involved in the technical practice of driving is accessible to all those who are able to drive a car. In other words, almost everyone can become a virtuous driver, whereas for instance only very few people can aspire to be virtuous nuclear scientists. This will be a valuable insight when considering our method for practising ethics, for this method will need to take into account what type of stakeholder participation is relevant when considering a particular technology.

In chapter 5, we already invoked the technical practice of driving a car when considering the **temporality** involved in the process of technological configuration. We argued that the design of the car configures both the chronological and non-chronological dimensions of time in the practice of driving, the latter of which provides the driver with a sense of freedom and responsibility to follow corresponding plots. Concerning "shifting gears", as being nested with regard to the practice of driving, we can see how the manual gear system provides more freedom and responsibility when contrasted with automatic gear systems. The driver is free to determine the order of basic actions and can make mistakes in attempting to apply the technique, but also – and more importantly – regarding "being awake

and succeeding" while being on the road. For instance, shifting gears becomes an integral aspect of responding to a sudden event in a traffic situation, such as the appearance of a speeding car at the rear, which inclines a driver who is awake and succeeding to keep driving in lower gears for a while to catch up speed quickly enough. This is accomplished in the practising of driving according to the movement of *ascending complexification*, because it links basic actions of switching gears with the ideal of being a virtuous driver, who is resolved to act in a definite moment. In other words, the a-chronological dimension of time in the technological configuration of a car with a manual gearbox allows for the action of shifting gears to be linked with the notion of being a virtuous driver.

A similar interpretation follows from the consideration of **distancing** in the technological configuration of a car, whereas the plot remains proximate to the world of action: one engages in driving in close proximity to other people driving their cars and other vehicles in the environing traffic. Self-driving cars would constitute a paradigm shift in this regard, because the models on which their responses would depend abstract from the world of action by linking generalised entities (quasi-characters) with generalised patterns of movement (quasi-events). In the case of a car with a manual gearbox, though, we argue that because the technological configuration is proximate to the world of action, it can bring about imaginative variations. As such, a car provides a "personal" understanding and experience of the practice of driving: because the technology configures basic actions such as shifting gears proximate to the world of action, it allows for a driver to link them with ideals such as those captured by a "driving style". One can link one's basic actions to being a "sportive driver", but also a "road devil". Hence, this type of technological configuration that approximates the world of action both enables virtue and vice.

The policy of "New Driving" discloses the opposite movement, of *descending specification*, for it links remote ideals of environmentally friendly life with action-chains and basic actions. That is, these remote ideals translate into policy that governs the technical practice of driving, which accordingly translates into changes in the constitutive rules that mediate basic actions. Again, technological configuration is important to consider, for it allows for this particular movement of descending specification to unfold. Concerning the **temporality** of technological configuration, the sense of freedom and responsibility brought about by the a-chronological dimension of time allows drivers to both be aware of the environmental impact of their technical practice and inclined to do something about it. In other words, it links the in-order-to of shifting gears with the for-the-sake-of-which of living sustainably in a political community. A similar interpretation can account for **distancing**: the

close proximity to the world of action allows drivers to evaluate their driving style and the driving styles of others according to the basic actions that link to the ideals of a sustainable life.

6.4 Technical Practice and the Cultivation of Virtue

In this final step of constructing our account of narrative virtue ethics of technology, we will consider how technical practice that cultivates the virtues is made possible through what Ricoeur designates as the "ethical aim". The ethical aim points at the *telos* of praxis, the good life for a human being (Aristotle's eudaimonia). This notion completes the determination of virtue, by restating it as being awake and succeeding for the sake of (1) the good life, (2) with and for others, (3) in just institutions. The ethical aim stipulates a three-fold movement, starting from a naïve conception of phronesis, through its actualisation by means of the norm, to a critical conception of practical reason. It is crucial to understand that by positing this movement of the ethical aim, Ricoeur aims to arrive at an account of practical reason that reconciles the deontological tradition, grounded in Kant's notion of *Moralität* (morality), with the teleological tradition, focusing on Aristotle's notion of *phronesis* and Hegel's notion of *Sittlichkeit* (also usually translated as "morality") (Ricoeur, 1992, p. 290).

6.4.1 "Little Ethics" of Technical practice

Ricoeur calls his ethical treatise in *Oneself as Another* his "little ethics", which despite its ironically modest title has an ambitious aim as we can observe above. What the three-fold movement of Ricoeur's ethics offers us is a way in which we can reflect on technical practice by considering how it relates to the ethical aim. Thereby, this ethics provides the final determination of the notion of technical practice that cultivates the virtues, which allows us to distinguish the "virtuous" thief from the virtuous human being. As such, it takes into account the self (the good life for me), the interpersonal (the good life for others) and the societal (the good life mediated by just institutions), which are interdependent (Ricoeur, 1992). It thereby covers the entire scope of technical practices: the ones we perform for our wellbeing, with and for other people, and for contributing to our shared institutions. Each of these three elements of the ethical aim, according to Ricoeur, has to move through the "sieve of the norm" (Ricoeur, 1992, p.170), through their actualisation in a normative structure, and eventually through a reflection on the norm based on a refined conception of the ethical aim. We can illustrate this with our above-mentioned example of the "new driving". First, by reflecting on the technical practice of driving in accordance with the ethical aim, we come to

understand that driving sustainably should be linked to the for-the-sake-of-which of driving. Second, to actualise this aim, a norm is required that enables this notion of the ethical aim to become part of practical life. Third, this norm calls for continual reflection based on the three-fold ethical aim: asking whether it (1) leads to an ethical conviction, (2) entails respect for the norm that does not lead to a lack of respect for persons, and (3) responds to our sense of justice. In what follows, we will trace this movement according to the three stages of the ethical aim: (1) the good life, (2) with and for others, (3) in just institutions. For each stage, we highlight the way in which it further determines the notion of "technical practice that cultivates the virtues".

A. The good life

Ricoeur's conception of the ethical aim seems almost diametrically opposed to Heidegger's notion of being-towards-death, because it does not attempt to ground itself in an existential analysis but instead traverses aspects of life from the self, extending all the way towards the other and public life. For Ricoeur, the "naïve", or perhaps rather "pre-dialogic", initial conception of the good life that makes it possible for people to engage in technical practice virtuously, is founded on the notion of "self-esteem", mediated through "self-interpretation" (Ricoeur, 1992, p. 179). This conception is directly tied with the threefold structure of practice, that moves through constitutive rules[74] and life plans to the narrative unity of life. As such, we can evaluate our actions in the narrative mode and can accordingly evaluate ourselves as the authors of these actions. For instance, we can use the standards of excellence that govern the practice of "playing the piano" to evaluate ourselves as the author of this practice (e.g. the "who" performing the musical piece), which we can extend in light of a life plan (e.g. the "who" that wants to become a professional piano player) and in light of the narrative unity of life (e.g. the "who" that might choose between being a piano player, a teacher, or a soldier). When it comes to the narrative unity of life, we face an open horizon of more or less distant ideals that asks for a never-ending process of interpretation (Ricoeur, 1992, p. 180). This open-endedness points at human beings understanding themselves as both acting *and* suffering beings. That is, in making life-choices, for instance choosing whether to become a professional pianist or a soldier, we only *co-author* the narrative structure that makes such choices intelligible. Unforeseen events might come in between and make us

---

[74] Constitutive rules, Ricoeur argues, as standards of excellence, provide practices with the notion of goods that are integral to them, and can when turned into norms provide contents to the categorical imperative (Ricoeur, 1992, p. 176).

"suffer", for instance if a physical test shows one to have insufficient eyesight to join the army and thereby closes off that choice. What this tells us about technical practice, is that the minimum requirement for such a practice to be engaged in virtuously is that a practitioner can interpret her self-esteem not only through evaluating the respective practice (e.g. creating a good computer program), but also through the related life plan(s) (e.g. wanting to *be* a good programmer) and narrative unity of life (e.g. having chosen well to become a programmer vis-à-vis numerable other choices). This shows us for instance that it is both extremely difficult and undesirable to aim to have someone engage in "ethical practice" concerning a technical practice that he cannot properly evaluate (e.g. a "mindless" task), related to a life plan that he feels alienated from (e.g. a pointless job), and a unity of life that he cannot make sense of (e.g. he has no clue of why this job was chosen). This might seem to be an issue that the human resources department of an R&I institution deals with, but Ricoeur shows how it is also firmly tied to the ethical aim as such.

This pre-dialogic notion of the good life as self-esteem needs to be actualised, which makes Ricoeur turn to Kantian moral philosophy and the norm. With the notion of the "sieve of the norm" (Ricoeur, 1992, p. 215), Ricoeur accomplishes an original cross-fertilisation of Aristotelian and Kantian ethics. That is, he argues both for (1) the necessity of the ethical aim to move through the deontological critique and for (2) the ultimate grounding of the deontological viewpoint in the teleological perspective. Translating the naïve ethical aim into a normative structure is necessary, according to Ricoeur, for two reasons. First, the facticity of an inclination towards evil in the world makes it insufficient to merely have a notion of the good life, but demands the formulation of moral laws that traverse the opposite direction, accruing to "thou shall not be evil" (Ricoeur, 1992, p. 218). Second, in exercising freedom, and in being able to make free choices, we find the fact that we are always free to choose to follow either good or evil maxims. Thus, we need normative structures ensuring that we act in accordance with the maxim that *limits* our actions so that it promotes the wish of what we ought *not* to be, namely: evil. Deontology therefore provides the necessary limitations on praxis in terms of normative structures, which could not be provided by the naïve conception of the good life. In Kantian ethics, this idea is captured by the notion of *autonomy*, which implies obedience to oneself, or self-legislation. Ricoeur connects this idea of self-legislation with constitutive rules that determine the standards of excellence of a practice, namely by stating that the categorical imperative is of the same form: that it internalises a constitutive rule that posits limits on action and at the same time provides a ground for appraisal. By means of self-legislation, we progress from self-esteem to "self-respect", which is "self-

esteem under the reign of law" (Ricoeur, 1992, p. 215). What this tells us about technical practice that cultivates the virtues, is that the practitioner ought not only to be able to evaluate her practices, life plans, and narrative unity of life, but also to limit her choices regarding these practices based on moral laws. This sense of limitation can be accomplished through performing an ethical oath, for instance in professional settings in banking and finance as advocated by Blok (2013). Blok argues that an ethical oath can mediate a technical practice when it takes place in a public context and when it pertains to the sense of self of the practitioner taking the oath. This accords with Ricoeur's theses that (1) selfhood is brought about by "keeping one's word" (Ricoeur, 1992, p. 123), and (2) that the sense of obligation derived from Kantian ethics does not imply any moral solipsism, but rather a moment of universalization that involves "neither you nor me" (Ricoeur, 1992, p. 204), which therefore essentially relates it to public life. In other words, by means of publicly establishing the moral limits of a technical practice, an R&I practitioner can provide content to the categorical imperative as a constitutive rule that mediates her self-respect in acting virtuously.

Finally, Ricoeur shows how self-esteem, through self-respect, turns into *conviction* by reflecting on the application of the norm. By means of this final step, Ricoeur shows that engaging in a technical practice while following a norm is *not enough*, and that one always needs to reflect on this norm by looking at the context in which it is applied and keeping the ethical aim in mind. This claim results from a questioning of the universalism of Kantian ethics from the contextualist viewpoint. Most notably[75], Ricoeur focuses on the encounter between the plea for universality of norms and the weight of the historicity of concrete morality (Ricoeur, 1992, p. 280). When universalist norms are put into practice, they can instigate conflicts based on the historical and cultural context of their application. For instance, Article 17 in the Universal Declaration of Human Rights on the right to property might cause conflict in societies that offer no cultural and historical basis for a compatible

---

[75] Ricoeur provides two additional reasons for questioning Kant's universalism. First, he argues that the Kantian notion of autonomy is really a *political* term that has been turned into a moral one, which depends on the heteronomy of public life rather than on the universalised individual. This is the case because autonomy presupposes the reciprocity of justice and the respect for other persons, in assuming that the self should be treated equally to the other who deserves respect and to the most disadvantaged in society who should be the focus of distributive justice. Consequently, in order to become truly "autonomous", we need the guidance of the *phronimos*, of the master of justice who teaches us how to take responsibility for our own choices (Ricoeur, 1992, p. 276). Second, Ricoeur criticises Kant's criterion of universalization, by stating that the coherence of a system of norms that implies a plurality of duties cannot be derived from the non-contradiction of duties with the categorical imperative. Instead, he states that a system of norms gains its coherence from being *constructed*, by way of refining norms contextually such as in the system of common law. For instance, the context provided by the case of "self-defence" allows for the refinement of the norm "thou shall not kill", by adding a specifying premise. At the same time, such a construction of a normative system ought always to be submitted to a criticism of the potential prejudices and ideological assumptions that might underlie the choices for specifications.

conception of personal property (Donnelly, 1984). The more specific certain norms become, the more likely they are to instigate such conflicts, and Ricoeur sees the essential diversity of goods that plays a role in such conflicts – and consequently the impossibility to have a single conception of distributive justice – as the reason to support the contextualist objection to universalism. For contextualists such as MacIntyre, the justification for norms as standards of excellence is traced back to a historical community, whereas for universalists such as Habermas this justification lies in a-historical principles of communicative action. Ricoeur situates himself at the crossroads between these traditions, by accepting Habermas' project of creating an "ethics of argumentation" (discourse ethics) while also renouncing the idea of ultimate foundation of this ethics in a-historical principles and consequently acknowledging the need to follow inverse path from that of justification: namely that of the actualisation of a normative system through practices, life plans, and the narrative unity of life.

Hence, Ricoeur states that ultimately discussions about autonomy, self-respect and the rule of justice should be mediated by a *reflective equilibrium* between the ethics of argumentation and "considered convictions" (Ricoeur, 1992, p. 289). He argues that argumentation, as an institutionalised language game, always involves other language games, including the crucial one of "narration", which grounds arguments in the world of action that is historically and culturally situated (Ricoeur, 1992, p. 288). The "best argument" concerning a discussion about norms is therefore one that not only fulfils formal principles of communicative action, but that also finds its contextual support in what MacIntyre calls a historical "moral community". To arrive at a reflective equilibrium between the ethics of argumentation and considered conviction, Ricoeur argues that we have to assume the paradox that "one must maintain the universal claim attached to a few values where the universal and the historical intersect" *and* that "one must submit this claim to discussion, *not* on the formal level, but on the level of the convictions incorporated in concrete forms of life" (Ricoeur, 1992, p. 289). The term "value" here is centrally used for the first time by Ricoeur and designates a "compromise concept" rather than a genuine moral or ethical concept. That is, a value can only arise from compromises made between universalist arguments and historical convictions. What this means for technical practice that cultivates the virtues is that the normative structure that governs a certain practice ought to rely *both* on a few values that are universally recognised *and* on the critical discussion of these values within a moral community. For instance, the building of medical equipment that relies on the use of embryos should be governed by universally recognised values, such as those presented in the UNESCO report on the use of stem cells in therapeutic research (Smith & Revel, 2001),

which are justified by considering the universalization principles in discourse ethics. At the same time, these values should be submitted to discussions about the context of their application, for instance about the ambiguous difference between "therapy" and "enhancement". Following Ricoeur, such a difference cannot be established by means of applying a-historical principles but can only arise from the use of narratives and other language games that provide the necessary context for arguments that try to establish the difference. For instance, a narrated account of an innovation relying on embryonic research that makes explicit how this innovation enabled a disabled child to suddenly be able to communicate with other children and partake in everyday life would elucidate how this innovation relates to the "therapy-enhancement" distinction.

B. With and for others

For Ricoeur, the notion of the good life already depends on a dialogical structure that includes otherness, but this otherness is made explicit by the addition of "with and for others". Again, he differs from Heidegger by making *care* dependent on the life with and for others, and by reconceptualising it as *solicitude.* That is, we can have self-esteem only to the extent that we care for the other as oneself (you appear before me as an "I") and for oneself as another (I evaluate myself according to the *phronimos*, the esteem of others). Departing from the Aristotelian virtue of *friendship* as an expression of this mediation*,* Ricoeur states that friendship for the sake of the good needs to be distinguished from friendship for the sake of utility and pleasure – for the latter two are essentially self-regarding. Friendship for the sake of the good is to be based on a mutual relationship between oneself and another, on a basic sense of reciprocity that involves a relation of *giving* and *receiving* (Ricoeur, 1992, p. 188). On the one hand, when the other takes the initiative we receive from the other what is lacking in ourselves by gaining from another's instruction and we give to the other our esteem, which leads to mutual recognition. On the other hand, when we take the initiative we give our sympathy and care to the other who is *suffering*, while at the same time receiving a sense of vulnerability related to the condition of mortality. Hence, this relation implies the mutual constitution of "agent" and "patient" in the process of giving and receiving. In a technical practice that cultivates the virtues, agents and patients are caught up in a reciprocal relationship that joins together "the reversibility of roles" (you regard me as an "I" and vice-versa) and the "non-substitutability" of persons ("each person is irreplaceable in our affection and esteem"), which we spontaneously assume because of the similitude between oneself and another (Ricoeur, 1992, p. 193). In other words, when engaging virtuously in a technical

practice, we spontaneously assume that its for-the-sake-of-which is not merely related to the good life for ourselves but also for others as acting and suffering, who are irreplaceable and essentially like ourselves. For instance, when we engage in the technical practice of "conversing on Facebook", it is not sufficient to link the for-the-sake-of-which of this practice to our personal idea of friendship in the good life, which risks being limited to self-love. Complementarily, we should regard *others* implied in this technical practice, by considering whether it allows us to attend to their suffering and receive their instructions. When electronic messaging on Facebook is merely performed to receive the esteem of others, no reciprocal relation is established, and no solicitude is implied that is necessary for the ethical aim.

As with the pre-dialogic notion of the good life, our understanding of solicitude needs to move through the sieve of the norm in order to be actualised as a "respect for persons". For this, Ricoeur surprisingly turns to the Jewish and Christian conceptions of the "golden rule", which can be stated negatively, in saying that one should *not* do to others what one would hate to do to oneself and positively, in saying that one should *love* another as oneself (Ricoeur, 1992, p. 219). The positive formulation departing from solicitude focuses on agency and includes the possibility of exerting power over someone else, which culminates in the most problematic cases in humiliation: the destruction of another's self-respect. The negative formulation departing from reciprocity focuses on patiency (i.e. being a patient) and balances the positive formulation by placing a limit on practice (thou shall not lie, not kill, not humiliate, and so forth). Ricoeur argues that Kant's second formulation of the categorical imperative is a formalisation of the golden rule, because it superimposes the universal concept of humanity - that by reason of which each is made worthy of respect - on the diversity implied by agency and patiency (i.e. always treat *humanity* never simply as a means, but always at the same time as an end). However, Ricoeur also shows that Kant cannot bring this formulation to its logical conclusion because it still depends on the idea of respect of persons, which can only be demonstrated by grounding the "fact of morality" on the fact that persons exist and that the end (telos) of a human being is distinct from the end of things that have utility and can bear a price. Ricoeur argues that this fact of morality can only be explained by pointing back at solicitude, at the care for another as oneself. For Ricoeur, therefore, respect for persons means restoring reciprocity where there is a lack of it (Ricoeur, 1992, p. 225), on the basis of a plurality of persons existing as ends in themselves. What this tells us about technical practice that cultivates the virtues, is that everyone consciously affected by such a practice - both acting and suffering - ought always to be considered as a

person whose existence is an end in itself and that a balancing of agency and patiency ought to be aimed at. For instance, regarding practices of automation in R&I settings, the workers being replaced should not be regarded as mere means (replaceable human resources), but always at the same time as ends in themselves (employees who deserve respect), whose patiency should be compensated by allowing them to have a say in the way the automation process takes place and the way in which reciprocity (e.g. compensation arrangements) is ensured.

Finally, Ricoeur shows that following the norm of treating other persons with respect in a technical practice is not *enough*, and that one needs to reflect on this norm, which allows Ricoeur to arrive at "critical solicitude" (Ricoeur, 1992, p. 273). He argues that conflicts arise when the "otherness of persons, inherent in the very idea of human plurality" (Ricoeur, 1992, p. 262) in certain contexts proves to clash with "the universality of rules that underlie the idea of humanity" (ibid.). In order to question Kant's notion of humanity, Ricoeur considers the possibility of making an exception on a universal rule based on the otherness of another person. For this, he considers the important case of "keeping a promise" where a constitutive rule is invoked stating that one places oneself under the obligation of doing such and such in specific circumstances. Ricoeur argues that this constitutive rule only confronts us with a moral problem on the basis of the *reason* for keeping the promise, which he posits as the principle of fidelity. Such a moral problem involves a dyadic structure, meaning that it involves two persons who both relate to the reason for keeping the promise, which implies that next to my will to keep a promise there is another person's expectation(s). Therefore, "it is to the other that I wish to be faithful" (Ricoeur, 1992, p. 268). This shows that there is a difference between legal promises such as contracts, which require something of someone regardless of the other, and keeping a promise, which implies someone *counting* on someone else, expecting him to keep his promise.

In light of this distinction, Ricoeur shows how exceptional circumstances can cause one to consider to either act or to refrain from acting in accordance with a moral rule based on the respect for persons, which essentially derives from solicitude. For instance, he considers the case of telling the truth to the dying. In case the dying person is morally or physically too weak to hear the truth, one should be compassionate and consider refraining from telling the truth. Similarly, questions arise when the distinction between "thing" and "person" to whom one is obliged to act in a certain way is ambiguous, like in the case of embryos conceived in test tubes for scientific research. This is one of the few places in his work where Ricoeur explicitly refers to modern technology, by stating that certain technical practices such as ones

involving the use of embryos in R&I render the dichotomy between things and persons ambiguous and therefore problematic. He also refers to the case of the impacts of emerging technologies on future generations, whose personhood is not firmly established, as another context that problematizes the universal application of rules connected to the respect to persons. To tackle these problematic cases, Ricoeur posits the notion of *critical solicitude* as the refined version of respect for persons, and adhering to three features of practical reason: (1) the recognition that a conflict between a rule and its application involves different positions adhering to the same principle of respect, (2) the positioning of Aristotle's notion of the mean as guiding principles, while accepting that this might imply enforcing a dichotomy, and (3) to take guidance of the *phronimos* into account. What this tells us about technical practice, is that in cases when such a practice involves problematic contexts of personhood - notably, (1) through lack of cognitive or emotional capacities, (2) through lack of a distinction between personhood and "thinghood", and (3) through the fog of time, considering future persons - the norms governing the practice regarding respect for persons ought to be receptive to adverse positions, seek the mean, and involve council of the virtuous.

C. In just institutions

To complete the determination of the ethical aim, Ricoeur moves from the interpersonal, characterised by solicitude, to the *institutional*, characterised by justice. Ricoeur thereby notably distances himself again from MacIntyre, by repudiating MacIntyre's claim that all institutions are necessarily concerned with external goods and thereby diametrically opposed to practices (MacIntyre, 2007, p. 194). Instead, Ricoeur acknowledges the intimate tie between practices and institutions, showing how institutions mediate conflicts that can arise from political practices. Starting again at the pre-dialogic stage, he argues that an institution denotes "the structure of *living together* as this belongs to a historical community", which secures duration, cohesion, and distribution (Ricoeur, 1992, p. 194). He derives this initial ethical understanding of institution from Arendt's concept of "power in common" that is contrasted with domination ("power over") and that is realised by a plurality of people acting in concert. The notion of plurality extends the face-to-face encounter of the interpersonal relation to acting with third parties that are not directly known. Because acting in concert with unknown third parties needs time to unfold, institutions are needed. Institutions thereby provide the necessary temporal dimension for the power in common to endure, which lies at the basis of any political community. This understanding of institutions is bound up with the Aristotelian virtue of justice, which Ricoeur argues does not yet belong to the deontological

notion of procedural justice. Initially, justice – like friendship – is aimed at the good, which by extending the interpersonal towards unknown third parties becomes the common good. For this extending to unfold, people do not only partake in institutions, but also *distribute* the related "parts" (roles, tasks, effort) within the political community. The ethical sense of justice thereby denotes a demand for equality, which is linked to solicitude at the interpersonal level by extending the idea of "the same for you as for me" to "the same for each", which is achieved through distribution. What this tells us about technical practice that cultivates the virtues is that the for-the-sake-of-which of such a practice ought not only to regard the known other, but also the unknown other in a political community. For instance, the technical practice of "making an investment" not only needs to regard the known parties directly involved (making a "fair" deal for the seller and potential buyers), but also needs to *refer back* to the world of action and consider a notion of fairness for those affected by the technical practice who are unknown: for instance the families affected by the investment practices into toxic mortgages prior to the 2008 financial crisis.

Ricoeur argues that *the just* faces in two directions: towards the good, connected with the pre-dialogic, ethical sense of justice, and towards the legal, connected with the moral *rule of justice* (Ricoeur, 1992, p. 197). Our sense of justice needs a normative structure to clarify its inherent ambiguities: (1) concerning the foundation of a "just share" on the basis of separation or cooperation, (2) concerning the justification for distinguishing between arithmetic equality and proportional equality, and (3) concerning the tension between justice based on the extension of solicitude and based on a system of impersonal positive law. In other words, even though we have an initial sense of justice based on the extension of solicitude, we need to actualise this sense of justice through a normative structure of rules, which demands justification. In order to justify a particular normative system, the fiction of the social contract is posited to replace the common good with an imaginary procedural deliberation (Ricoeur, 1992, p. 228). Ricoeur argues that the social contract has a function similar to Kant's notion of autonomy but differs from autonomy because it is not a "fact of reason", but an imaginary political reality. He sees Rawls' conception of justice as fairness at the furthest development of this thought, founded on the a-historical fiction of the veil of ignorance as the original position, which is the basis for a procedure towards a shared agreement on the principles concerning what is just. For instance, it stipulates when arithmetic equality should be applied, based on the claim to basic needs (Rawls, 1971, p. 286).

Ricoeur accepts Rawls' justification of the two principles of justice[76] based on the decision on distribution of goods made in a society behind the a-historical veil of ignorance as a formalisation of the sense of justice. However, he rejects this justification as the final determination of justice because it presupposes an understanding of the just and unjust that is still founded on an ethical understanding of justice. In other words, for participants behind the veil of ignorance to distinguish a just from an unjust distribution of goods based on a procedure, they still have to appeal to a notion of "good", which opens up the problem of a diversity of goods and of incommensurable but equally valid arrangements of distributive justice (Ricoeur, 1992, p. 233). Despite this reservation, the normative structure introduced by the social contract adds to the ethical aim and informs the notion of technical practice that cultivates the virtues. That is, the sense of justice implied in the for-the-sake-of-which of a technical practice needs to be submitted to a normative test involving the principles of justice. Going back to one of our earlier examples of "making an investment", we could state that we cannot simply extend our sense of justice from our relationship with buyers and sellers to the unknown people affected by our investments. Even if it would be "just" for a seller to lose a great deal of profit because of a transaction that was nonetheless conducted fairly, the same would not apply when the same sense of justice would imply that a family would need to vacate its house and risk its livelihood. Therefore, Rawls' principles of justice should be applied to test whether the sense of justice can properly be extended or whether another notion of distributive justice is called for. For instance, in this case the particular family might set the benchmark for what is a fair arrangement such that "social and economic inequalities are to be arranged so that they are"… "reasonably expected to be in everyone's advantage" (Rawls, 1971, p. 53).

Finally, Ricoeur states that it is *not enough* to submit one's sense of justice to a normative system, but that one again needs to reflect on this system, which brings us from an initial sense of equality, through the rule of justice, to a *sense of justice*. This need springs from the inevitable conflicts that arise when uniform moral principles are applied to complex social reality. Ricoeur initially refers to Walzer's critique of Rawls (Ricoeur, 1992, p. 252), which is focused on the problem of the diversity of goods and calls for the conceptualisation of "spheres of justice" whose goods are incommensurable. Ricoeur shows how this critique

---

[76] These two principles are: (1) "each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all" (Rawls, 1971, p. 220) and (2) "each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others", while "social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all" (Rawls, 1971, p. 53).

introduces the new problem of arbitration between the spheres of justice (e.g. on what basis can we demarcate and limit the competition between the sphere of "money and merchandise" and "security and wellbeing"?). This leads him to consider Hegel's notion of *Sittlichkeit,* which denotes concrete morality in a political community culminating in the state as the final arbiter between the spheres of justice. Ricoeur retains from this notion of Sittlichkeit its determination as a "system of collective agencies" and "the gradual triumph of the organic bond between men and women over the exteriority of the juridical relation" (Ricoeur, 1992, p. 255). In other words, Sittlichkeit does *not* denote a system of individual rights and duties (such as Rawls' notion of distributive justice) but a system of arbitration between competing claims in society that by juxtaposing the application of the law with the ethical significance of particular situations moves in a continual progression towards a system of rights that is maximally non-conflictual. Importantly, Ricoeur argues that Sittlichkeit can only be realised on condition that the tie between institutions and the spirit of the people is not perverted: that institutions do not posit aims apart from those granted to them by the people. Perversion happens because the initial agency of the "power in common", which relates to people sharing the ethical aim, is *forgotten* in the process of institutionalisation. We argue that with Sittlichkeit, we arrive at the "technical practice" of politics, at that with which the political philosopher as a craftsman is concerned.

Ricoeur indeed defines "the political" as "the set of organised practices relating to the distribution of political power, better termed domination" (Ricoeur, 1992, p. 257). As such political practices are aimed at addressing conflicts vertically, between the governing and the governed, and horizontally, between competing groups in civil society. Ricoeur offers three ways in which political practices mediate conflicts, which inform our notion of technical practice that cultivates the virtues. First, he argues that the realm of everyday discussions about the order and priority of different spheres of justice ought to be open and negotiable and should not be dominated by scientific or dogmatic arguments (Ricoeur, 1992, p. 258). Second, he argues that these everyday discussions should be embedded in a longer-term discussion on the ends of good government, concerning the meanings of political concepts such as "liberty" and "security" and their anchoring in a good constitution that is open to historical re-interpretation. Third, he argues that both aforementioned types of discussions ought to be embedded in a discussion on the process of legitimation (Ricoeur, 1992, p. 260). Democracy, which denotes this process according to Ricoeur, needs to confess indeterminacy concerning the basis of power, law, and knowledge, and commit to a pluralist, public debate. To balance democracy, finally, equity - rephrased as the *sense of justice* - remedies the rule of

justice where the legislator fails us through over-simplicity. What this tells about technical practice that cultivates the virtues, is that when such a practice instigates conflicts between spheres of justice or between domination and legitimation, practices towards solving these conflicts should be (1) organised openly and subjected to negotiations, (2) based on a shared conception of good governance, (3) both of which are legitimised by a democratic process and balanced by a sense of justice. A clear example in this regard are conflicts instigated by the practice of obtaining digital evidence in criminal cases, which particularly came to light when the FBI attempted to get access to an iPhone while Apple refused to offer a "back door" into the software of the device (Hack, 2016). Such a conflict both involved spheres of justice (Apple belonging to "money and merchandise" and the FBI to "security and wellbeing") and a conflict between domination and legitimation (the FBI having "power over" the data of US citizens, without clearly having its aim aligned with the "spirit of the people"). In this particular case, Apple refused to assist the FBI and the FBI eventually managed to gain access to the device by its own means. However, when applying Ricoeur's principles to this particular technical practice, neither of the parties involved should have been left to their own judgment. Instead, because of the conflicts created (1) an open public discussion should have been organised between ordinary citizens, that should have (2) either (a) been informed by existing norms for good governance or (b) led to novel and refined norms, and (3) that should have led to a process of democratic legitimisation, which would have provided the necessary arbitration between the parties involved as well as to a balancing of this process by a sense of justice (commonly vested in the judiciary branch).

6.5 Chapter Summary

At this point, we have a comprehensive account of technical practice that cultivates the virtues that integrates the notions of virtue and technological emplotment. That is, to engage in a technical practice that cultivates the virtues involves being awake and succeeding, which depends on the nesting of finalities, of the in-order-to with regard to the for-the-sake-of-which. This nesting of finalities consists in linking basic actions, action chains and practices through life plans to more or less distant ideals. Life plans, which introduce the process of technological configuration, denote the two-fold movement of ascending complexification and descending specification in the process of technological emplotment. To understand these movements, we should consider the textuality of a technology, the literacy of the respective practitioners, and the temporality and distancing of the process of technological configuration. We thus have a set of powerful concepts that allow us to interpret any

particular technical practice to a certain extent. That is, they allow us to interpret a technical practice by providing a descriptive and narrative account of it and by providing the first determination of virtue, namely of being awake and succeeding.

To link the notion of being awake and succeeding to the necessary notion of the for-the-sake-of-which in living together in a political community, we needed to link technical practices to the ethical aim. By doing so, we argued that we could only speak of a technical practice that cultivates the virtues when the for-the-sake-of-which of this practice accords with the good life, with and for others, in just institutions:

Table 6: The nine core concepts that capture Ricoeur's movement of the ethical aim, which complete the determination of technical practice that cultivates the virtues to the furthest extent.

	Ethical aim$_1$	Sieve of the norm	Ethical aim$_2$
The good life	Self-esteem	Self-respect	Conviction
With and for others	Solicitude	Respect for persons	Critical solicitude
In just institutions	Equality	Rule of justice	Sense of justice

Finally, we need to make two important reservations regarding the completeness of our framework. First, just as was the case with phenomenology, narrative contributes to ethics by recognising its *limit* (Ricoeur, 1992, p. 245), thereby recognising it as being inevitably incomplete. According to Ricoeur, particularly tragic narrative is not to be considered a source of moral teaching, but as something that opens the path to moral conviction by meditating on the inevitability of conflict in life. In other words, narrative allows us to recognise the inevitable ethical conflicts in our lives without offering an "ethical" solution to those conflicts. Even though our framework can therefore be used to approximate a notion of the good life, with and for others, in just institutions, this good life is still affected by the inevitability of tragedy. We can design technologies responsibly, use them with care for others, and derive our self-esteem from our engagement in technical practice that cultivates the virtues, but this does not salvage us from hubris, from the inevitable accident, disaster, or tragic course of events. Second, we did not and will not engage with Ricoeur's speculative attempt to provide his ethics with ontological foundations in the tenth study of *Oneself as Another*. We will not do so for two reasons: (1) it would lead us beyond our current purposes of providing an "ethics" in this dissertation, and (2) it would call for an elaborate commentary when put in light of the accounts of technological mediation, which would lead to a research project in its own right. This has as a ramification that the "otherness" implied in our account (i.e. self-esteem in relation to the esteem of others, solicitude as the care for

others, justice as solicitude for the unknown other) remains without a firm ontological foundation and has to be assumed without being adequately justified.

In chapter 7, we will use the methodology developed in this chapter, understood as the theoretical framework that offers both the resources and the justification for the construction of a practice-oriented method (which is, in the end, a procedural enterprise). In line with the recommendations provided in chapter 3, we will draw from the methodology in terms of (1) interpretation of a technical practice, (2) the involvement of R&I practitioners, (3) the realisation of ethical design, (4) the management of conflicts, (5) the consideration of alternatives, (6) the selection of stakeholders and (7) the organisation of participation. In chapter 8, we will finally evaluate whether our novel method adequately addresses the recommendations related to these points.

6.6 References

Arendt, H. (1958). *The Human Condition*. Chicago: University of Chicago Press. http://doi.org/10.2307/2089589

Blok, V. (2013). The Power of Speech Acts: Reflections on a Performative Concept of Ethical Oaths in Economics and Business. *Review of Social Economy*, *71*(2), 187–208. http://doi.org/10.1080/00346764.2013.799965

Bobb, C. (2011). Paul Ricoeur's Hermeneutics Between Epistemology and Ontology. *Logos & Episteme*, *2*(3), 335–345.

Brogan, W. (2005). *Heidegger and Aristotle: The Twofoldness of Being* (Vol. 5). New York: SUNY Press. http://doi.org/10.1111/j.1468-2265.1964.tb00961.x

Bucher, E., Fieseler, C., & Suphan, A. (2013). the Stress Potential of Social Media in the Workplace. *Information, Communication & Society*, *16*(10), 1639–1667. http://doi.org/10.1080/1369118X.2012.710245

Coeckelbergh, M., & Reijers, W. (2016). Narrative Technologies: a philosophical investigation of the narrative capacities of technologies. *Human Studies*, *39*(3), 325–346. http://doi.org/10.1007/s10746-016-9383-7

de Goede, M., & Hoedemaeker, M. (2009). *Daadwerkelijk rijgedrag met Het Nieuwe Rijden. Program*.

Donnelly, J. (1984). Cultural Relativism and Universal Human Rights. *Human Rights Quarterly 400*, *6*(4), 400–419. http://doi.org/10.2307/762182

Fagenblat, M. (2016). "Heidegger" and the jews. In I. Farin & J. Malpas (Eds.), *Reading Heidegger's Black Notebooks 1931-1941*. Cambridge, Massachusetts.

Feenberg, A. (1999). *Questioning Technology*. New York: Routledge.

Hack, M. (2016). The implications of Apple's battle with the FBI. *Network Security*, *2016*(7), 8–10. http://doi.org/10.1016/S1353-4858(16)30068-X

Heidegger, M. (1977). *The Question Concerning Technology and Other Essays*. New York: Graland Publishing, Inc.

Heidegger, M. (1996). *Being and Time*. (J. Stambaugh, Ed.). New York: State University of New York Press.

http://doi.org/10.1017/CBO9781107415324.004

Heidegger, M. (1998). Letter on "Humanism." In *Pathmarks* (pp. 239–276). Cambridge: Cambridge University Press. http://doi.org/10.2307/2904244

Heidegger, M. (2009). *Basic Concepts of Aristotelian Philosophy*. Bloomington: Indiana University Press.

Hodge, J. (1995). *Heidegger and Ethics*. London: Routledge.

Ihde, D. (1979). *Technics and Praxis* (Vol. 33). Dordrecht: D. Reidel Publishing Company.

Kaplan, D. M. (2003). *Ricoeur's Critical Theory*. *SUNY series in the philosophy of the social sciences*. New York: SUNY Press. http://doi.org/10.1017/CBO9781107415324.004

Knight, K. (2008). After Tradition?: Heidegger or MacIntyre, Aristotle and Marx. *Analyse & Kritik*, *2008*(1), 33–52. Retrieved from http://www.analyse-und-kritik.net/en/2008-1/AK_Knight_2008.pdf

MacIntyre, A. (2007). *After Virtue: A study in moral theory* (Third Edit). Notre Dame, Indiana: University of Notre Dame Press. http://doi.org/10.1017/CBO9781107415324.004

Merritt, M. W., Doris, J. M., & Harman, G. (2010). Chapter 11: Character. In J. M. Doris (Ed.), *The Moral Psychology Handbook* (pp. 1–504). http://doi.org/10.1093/acprof:oso/9780199582143.001.0001

Partridge, E. (2006). *Origins: A Short Etymological Dictionary of Modern English* (fourth edi). London: Routledge.

Rawls, J. (1971). *A theory of Justice*. Harvard University Press.

Reijers, W., & Coeckelbergh, M. (2016). The Blockchain as a Narrative Technology : Investigating the Social Ontology and Normative Configurations of Cryptocurrencies. *Philosophy & Technology*, *7*. http://doi.org/10.1007/s13347-016-0239-x

Ricoeur, P. (1980). Narrative Time. *Critical Inquiry*, *7*(1. On Narrative), 169–190.

Ricoeur, P. (1983a). Action, Story and History: On Re-reading The Human Condition. *Salmagundi*, (60), 60–72.

Ricoeur, P. (1983b). *Time and Narrative - volume 1*. (K. McLaughlin & D. Pellauer, Eds.) (Vol. 91). Chicago: The University of Chicago. http://doi.org/10.2307/1864383

Ricoeur, P. (1988). *Time and Narrative - volume 3*. Chicago: The University of Chicago.

Ricoeur, P. (1992). *Oneself as Another*. (K. Blamey, Ed.). Chicago: University of Chicago Press.

Searle, J. R. (2010). *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press. http://doi.org/10.1093/acprof:osobl/9780195396171.001.0001

Siren, A., & Hakamies-Blomqvist, L. (2005). Sense and sensibility. a narrative study of older women's car driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(3), 213–228. http://doi.org/10.1016/j.trf.2005.04.008

Smith, A. M., & Revel, M. (2001). *The Use of Embryonic Stem Cells In Therapeutic Research* (No. BIO-7/00/GT-1/2 (Rev. 3)). Paris.

Taminiaux, J. (1992). *The Thracian Maid and the Professional Thinker: Arendt and Heidegger*. New York: State University of New York Press.

Thomson, I. (2005). *Heidegger on Ontotheology*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781107415324.004

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.

Verbeek, P.-P. (2005). *What things do; philosophical reflections on technology, agency, and design*.

Pennsylvania: Pennsylvania University Press.

Villa, D. R. (1996). *Arendt and Heidegger: the Fate of the Political*. Princeton: Princeton University Press. http://doi.org/10.1515/9781400821846

7 A Novel Method for Practising Ethics in R&I[77]

7.1 Introduction

Having traversed our philosophical investigation involving a discussion of virtue ethics of technology, the construction of the narrative technologies approach and a drawing together of these two approaches into an account of technical practice that cultivates the virtues, we arrived at a vantage point from which we can construct a method for practising ethics relevant to concrete R&I settings. As we established in chapter 1, we understand with the notion of "method" a detailed, specific problem-solving procedure. However, to this we will add that any method in light of our methodology would need to be critical.

On the one hand, we should follow Gadamer (2013) in being suspicious of any rigorous procedure that aims at solving an allegedly well-defined problem. As we discussed in chapter 6, we cannot speak of any *definite* end of a technical practice that cultivates the virtues, which is essentially open-ended with regard to its for-the-sake-of-which, and even less of a formal representation of this end that would allow it to be posited as a definite solution to a definite problem. However, on the other hand we should follow Ricoeur in not wanting to face the false choice between "truth *or* method" (Kaplan, 2003, p. 34), which would place all considerations of method outside of the scope of our endeavour. Instead, we intend to engage with method in a productive way, making use of its capacities to structure R&I activities while at the same time keeping a critical distance – weary of its power to turn into *dogma*. In order to do so, the procedural steps that we propose will remain open-ended, meaning that they cannot be brought to a definite end but require iterative revisiting, and they will contain explicit elements of critique.

What, we should ask initially, would be the end of a critical method such as the one we intend to construct? First, we need to acknowledge that the end of any method that concerns itself with ethics is praxis: that its end lies in the practical reality of R&I practitioners and more broadly of people working, playing, building and designing with

---

[77] Parts of this chapter have been adopted from the conference papers: Reijers, W., Gordijn, B., & O'Sullivan, D. (2016). Narrative ethics of personalisation technologies. In D. Kreps, G. Fletcher, & M. Griffiths (Eds.), IFIP Advances in Information and Communication Technology (Vol. 474, pp. 130–140). Cham: Springer International Publishing **&** Reijers, W., Koidl, K., Lewis, D., Harshvardhan, J. P., & Gordijn, B. (2018). Discussing Ethics Impacts in Research and Innovation: The Ethics Canvas. In D. Kreps, G. Fletcher, & M. Griffiths (Eds.), IFIP Advances in Information and Communication Technology (Forthcoming). Cham: Springer International Publishing.

technologies in their everyday lives. Second, since its end lies in praxis, we need to acknowledge that it aims at a particular type of praxis, namely at technical practice that cultivates the virtues for the sake of the good life, with and for others, in just institutions.

In what will be presented below, we need to take into account the distinction between the *method* that is presented and the *tools* that it might bring forth, of which one will be fully worked out in this dissertation, namely the Ethics Canvas. To consider the difference between our method and a related tool, one can consider the difference between the method of building a house and the tools used for the building process. The method of building a house contains procedural steps that can or should be followed, such as a step stating that when placing stones on the cement one should ensure the stone to be placed horizontally. In order to follow this step, a builder can use the tool of a spirit level while engaging in the technical practice of building a house. Similarly, we will present procedural steps in our method, for instance one stating that one should gather narratives concerning a particular technology. Accordingly, we propose a tool for someone to actually do this in practice, which will be the Ethics Canvas.

Our method will be an attempt at formulating an answer to the question of *how* to ethically improve technical practices so that they become better attuned to the cultivation of the virtues. The methodology that we developed in the last three chapters offers the conceptual framework and the resources that will guide us in this formulation. In what follows, we will first present the procedural steps that together form our critical method for practising ethics in R&I. Second, we will show how this method can be applied in concrete R&I settings by being translated into a tool that R&I practitioners can use in their day-to-day activities: the Ethics Canvas. Third, we will evaluate our effort to construct a novel method for practising ethics in R&I by revisiting the recommendations presented in chapter 3.

7.1 A Novel Method: Designing Practice for Virtue

In this section, we construct our method, starting at the most general, descriptive level, which explicates the central phases involved in the application of the method, and moving to the prescriptive level, which provides more specific guidance regarding how technical practices should be cultivated. Our method will be called "designing practice for virtue", alluding to the notion of "value sensitive design", but also distinguishing it from the latter approach: namely, by not focusing on values embedded in the design of a technology, but on virtue(s) cultivated by technical practices in which a technology is involved. First, we present a case of a general technology and a specific application of that technology that will serve as an

illustration that enables us to connect the procedural steps we present with concrete R&I practices. Second, we work out the three phases of our tripartite method: from description to interpretation, interpretation to evaluation and from evaluation to prescription. Below (table 7), a schematic overview of the method is presented.

Table 7: Overview of the stages and procedural steps, which together form our novel methods for practising ethics in R&I.

Stage:		Procedural steps:
Phase 1: Description < > Interpretation		
I.	Map technical practices	Iteratively map technical practices in which a technology is involved
		Consider the different technical practices according to the points of view of making, using, and governing
II.	Gather narratives	Select stakeholders, using the categories of users, makers, governors
		Gather second-order narratives
		Gather first-order narratives
III.	Undertake hermeneutic analysis	Interpret the technology's textuality
		Interpret literacy by considering those using and making the technology
		Investigate the temporality of the process of technological configuration
		Investigate distancing of the process of technological configuration
Phase 1: Interpretation < > Evaluation		
IV.	Uncover standards of excellence	Uncover constitutive rules of a technical practice
		Keep these rules open to critique and revisions
V.	Uncover life plans	Uncover life plans that explain a technical practice
		Explicate life plans according to the movements of ascending complexification and descending specification
VI.	Evaluate narrative unity of a life	Uncover elements that constitute the narrative unity of life
		Establish whether technical practices link to a narrative unity of life or whether one deals with the mere application of a technique
Phase 1: Evaluation < > Prescription		
VII.	The good life	Prescribe forms of mentorship that promote self-esteem
		Prescribe forms of an ethical oath, allowing practitioners to gain self-respect through giving themselves a norm
		Prescribe forms of structured dialogue to mediate conflicts and arrive at considered convictions
VIII.	With and for others	Prescribe forms of community building that promote reciprocity
		Prescribe forms of codes of conduct that ensure respect for persons
		Prescribe forms of expert guidance to mediate conflicts in dealing with vulnerable persons and arrive at critical solicitude
IX.	In just institutions	Prescribe forms of civic education that promote equality

Stage:	Procedural steps:
	Prescribe forms of legal regulations that ensure respect for the rule of justice
	Prescribe forms of democratic decision-making that mediate conflicts between spheres of justice and arrive at a sense of justice

7.1.1 Illustrative case: personalisation technology

Throughout this chapter, we will continually link the further development of the novel method with concrete R&I practices in order to illustrate how its procedural aspects relate to the world of action. To this end, we will introduce a specific technology as a practical case: personalisation technology. This technology is involved in an innumerable variety of different technical practices, from detecting fraudulent insurance holders to tailoring learning experiences for children in primary schools. To also consider a more particular technology, as in a type of technical object that involves multiple but not innumerable technical practices, we will additionally focus on the case of smart border controls that incorporate the technical process of personalisation.

Personalisation can be applied to a wide range of different technologies, but it generally revolves around notions of "adapting", "fitting" or "tailoring" digital content to the human being(s) interacting with it. A central term in personalisation research is said to be the "adaptability" of a system (Lycett, 2013). Adaptive systems include three basic models in their design: a "user model" that is a structured model for the collection and categorisation of personal data belonging to a user, "an application model", which is a description of relevant features of the application, and an "interaction model" that is meant to structure the organisation of interactions between a user and a system (Lycett, 2013). Asif and Krogstie argue that multiple personalisation approaches exist that can be based on "machine-learning algorithms, agent technology and ubiquitous and context-aware computing" (Asif & Krogstie, 2012, p. 346). They identify a "basic level" of personalisation, at which a user selects a certain configuration of a computing device or interface, which subsequently remains the same. This basic level corresponds with the conventional, instrumental view of a personalisation technology: of the human user determining the technology's settings and behaviour. Then there is a "second level", at which the configuration of a system is based on a user's "profile", and a "third level" at which both the profile of the user and his or her "context" (mostly comprised of meta-data such as location, time of the activity, type of activity) are used as the basis for the configuration of the system. Bouzeghoub and Kostadinov (2007) make a distinction between profiles and queries: a profile being a "user

model" "defined by a set of attributes" and a query being an "on-demand user need". Roosendaal (2014) offers an additional distinction, namely one between "digital personae", which are representations of known individuals in the real world and "digital profiles" which are sets of characteristics about persons that can be used as inputs for algorithmic decision making.

The process of personalisation invariably integrates underlying assumptions about *what* a person is and – more importantly – about what we can understand as the *right* information for the *right* person. It is an automated process of categorisation, and therefore of inclusion and exclusion of both digital content and of digital personae or profiles from the information people access. As a result, personalisation processes can for instance influence what type of information we retrieve from our search engine, what kinds of products are recommended to us when browsing online, and what kind of feedback we receive about our daily activities. Also, personalisation processes are used by intelligence agencies to identify suspicious individuals and by insurance companies to establish people's personal credit ratings. Accordingly, these increasingly ubiquitous computational processes that directly influence many aspects of people's everyday lives can have significant ethical implications. For instance, de Vries (2010) argues that ethical concerns arise from cases of implicit discrimination based on profiling. Moreover, Bohn et al. (2005) argue that the practice of matching a digital persona with profiles that present a public security concern has ethical implications because it can lead to mass-surveillance practices. In a different vein, Schubert (2015) argues that the use of personalised technologies that nudge their users in certain directions can lead to a reduction of personal autonomy and agency.

A clear example of personalisation based on a digital persona and profiles is an automated passport check, or smart border control, at an airport (Juels et al., 2005). When someone's passport chip is scanned, the retrieved data is compared with a data entry containing the document number belonging to the respective person, accompanied by her picture, biometric information and information about country of origin, age, and so on. Based on an algorithmic assessment of this personal data, the person can either pass through or will be held and interrogated by the border police. In this process, the digital persona can be compared with and transformed into a digital profile, for instance by linking it with a certain profile containing passport features that are deemed "suspicious", or by using it to add to the digital profile of people originating from the same country. What the process of personalisation in smart border control systems has in common with personalisation processes in general is that it uses interactions with humans to gain knowledge about them

and create a representation that changes the behaviour of a system in order to fit, adapt to or be tailored to this representation. The purpose of many instances of personalisation research is to make sure that the behavioural changes of the system approximate the expectations, wishes and/or needs of the human interacting with it. Terms like tailoring suggest that the user – the human agent – is seen as a given, as a static point to which the personalisation processes need to adjust. Just like a tailor adjusts the sizes and shapes of a piece of clothing to the human body that remains the same (or rather, that is defined by fixed measurements), personalisation processes are supposed to be tailored to users who are presumed to remain the same. Accordingly, user needs or preferences are supposed to be fixed. For instance, it might be assumed that the user of a weather application has a fixed need of knowing what the weather will be the next day in her city, and on request a personalisation process will link the data of her location and weather forecasts to provide for the desired information.

In what follows, we will usually refer to personalisation technology in general, and smart border controls in particular to illustrate the use of our method. That is, we will relate the technical practices that we need to consider in our method to personalisation technologies, and use them to explicate relevant stakeholders, basic actions, narrative structures, and so forth. We will present a number of fictional accounts of narratives related to technical practices in which smart border control systems are involved. Occasionally, however, we will also illustrate some points with reference to familiar technologies different from smart border control systems (e.g. smart phones). The fictional accounts we provide are to be taken solely as illustrations, explicating how the method could apply to specific technical practices, and are not to be taken as demonstrations that could be part of an actual case study.

7.1.2 An iterative, tripartite method

To start, the outline of our method needs to be made explicit. At a very general level, our method follows Aristotle's explication of method in the *Nicomachean Ethics,* which Irwin (Aristotle, 1999, p. 256) claims to be most clearly stated in Book 7, Chapter 1:

> "We must set out the appearances, and first of all go through the puzzles. In this way we must prove the common beliefs about these ways of being affected – ideally, all the common beliefs, but if not all, most of them, and the most important. For if the objections are solved, and the common beliefs are left, it will be an adequate proof" (NE, 7.1, 1145b4-8, trans. Irwin).

Aristotle argues that applying the *method* of ethics means firstly to understand the common beliefs about a particular issue, secondly to subject these common beliefs to a philosophical critique, and thirdly to return to these common beliefs by targeting the world of action – wherein the *telos* of any *ethics* lies. Similarly, in line with our discussion of the ethical aim, we argue that any reflection on new ethical concerns that we face due to technological change *starts* from convictions (i.e. common beliefs), which raise certain puzzles. These puzzles pertain to technical practices, for it is the standards of excellence, life plans, and the narrative unity of life linked to practices that are questioned in our ethical framework. In questioning these, we can respond to the puzzles by either confirming or refining the pre-given convictions, which then are turned into considered convictions.

We thereby "map" our method on the three-fold movement developed by Ricoeur: of (1) mimesis$_1$-practices, (2) mimesis$_2$-life plan, and (3) mimesis$_3$-narrative unity of a life, and accordingly, of the move from description, through narration, to prescription. By assuming this tripartite method, we adopt a structure similar to the one proposed in value sensitive design (VSD). However, VSD proposes an order different from ours insofar as it starts with the conceptualisation of values, proceeds with empirical investigations of these values in context, and integrates them in technical investigations that aim at the embedding of values in design (Friedman & Kahn, 2002). In contrast to VSD, we do not start with conceptual investigations but with the "empirical investigations", namely with the search for what Aristotle designates as "common beliefs" and what Ricoeur calls "convictions". Only when we have gained an understanding of the narrative structures by which practitioners make sense of a technology and its respective technical practices can we proceed to the conceptual work of criticising these structures. In our method, we additionally distinguish three structural levels: (1) "phase", which refers to the gradual carrying out of a particular task, (2) "stage", which refers to a distinct standpoint in the process of carrying out such a task and (3) "step", which refers to a concrete procedural element of the method.

In its general outline, our method incorporates three over-arching phases, each phase denoting a transition:

1. **Description < > interpretation:** gaining a thorough understanding of a network of technical practices in which a technology is involved.

2. **Interpretation < > evaluation:** interpreting these technical practices according to the puzzles that they raise regarding standards of excellence, life plans and narrative unity of life.

3. **Evaluation < > prescription:** evaluating how these technical practices relate to the good life, with and for others, in just institutions and prescribe ways to improve them.

7.1.3 Phase I: Description < > Interpretation

The first general phase of our method concerns the move from a description of technical practices to their interpretation. Suppose one is a philosopher who is charged with applying our novel method to a particular R&I project that aims at developing a personalisation technology for smart border control systems. In accordance with our methodology, the first step would be to uncover the different technical practices in which such a technology is involved, which can be categorised according to the making, using, and governing of a technology. The second step to take would be to uncover the narrative structures that characterise one or more technical practices. This can both be done by (1) gathering narratives *about* a technology, and by (2) investigating the relevant process of technological configuration (i.e. focusing on the design of the technology). These narrative structures can consequently be used as the basis for the hermeneutic analysis that is informed by the narrative technologies approach, and which aims at explicating the textuality and literacy concerning the process of technological emplotment and the temporality and distancing in the process of technological configuration.

1. Mapping technical practices

The **first stage** of our method is concerned with the mapping of technical practices to uncover a **network of technical practices** in which a particular technology is involved. The uncovering of such a network is in a way similar to uncovering the cluster of actants in actor network theory (Latour, 2013), but instead of focusing on actants and the semiotic relations between them it focuses on a cluster of technical practices and their hierarchic organisation according to constitutive rules, life plans and the narrative unity of a life. We need to emphasise what sets our approach grounded in virtue ethics apart from most other methods for practising ethics in R&I: namely, the focus on *technical practice*. That is, our methodology shows that what is at stake in any ethical appraisal of a technology is neither merely the thing in question – the smartphone, test tube, self-driving car, and so forth – nor its potential consequences – existential risks, climate change, the death of a patient, and so forth – but foremost the technical practices in which the technology is involved. This is not to say that the things or their consequences do not matter, but simply that these only matter *with regard to* a technical practice and not considered in isolation. Thus, if we for instance want to

evaluate the ethical impacts of a smartphone, we should not focus on the thing, but on the technical practices in which it is involved, such as "having an online conversation".

Now, we need to address two important objections to such starting point for our method. First, one could object that not every interaction with a technology constitutes a technical practice as we earlier conceptualised it. For instance, we could hardly speak of the ethics of "swiping" or "checking the weather" on a smartphone, because these activities do not relate to standards of excellence and life plans. Yet, these activities should not be considered ethically "neutral", for they could bring about benefits or harms, so we should include them in our considerations. As a response, we reply that indeed what is at stake when considering these activities would need to be extended to a technical practice in order to be ethically significant. As such, we cannot speak of "swiping" on a smartphone as being either good or bad or bringing about good or bad character, even when we could establish that excessive swiping could for instance lead to health problems. When the latter is the case, we ought not to focus on the swiping as such, but on the technical practice in which it is involved, such as "playing video games". Such a technical practice links to standards of excellence and to life plans (of leisure time). Addressing the basic activity of swiping means therefore primarily addressing technical practices such as playing video games on a smartphone, which can explain *why* a smartphone user engages in excessive swiping.

Second, one could object that a technology can be involved in innumerable technical practices, which makes it impracticable or even impossible to focus on technical practices instead of on the thing in question, or on its consequences. This brings us back to the problem of uncertainty: the further a technology is removed from any concrete application the less we can reliably speak of a technical practice in which it is involved. For instance, R&I of quantum computing is still in a very experimental phase and we cannot yet say anything reliable about the way in which it will play a role in technical practices of those who will use it. We can therefore speak of a strong indeterminacy of technical practices when considering R&I of quantum computing – and for that matter of most emerging technologies. As a response, we argue that indeed we face indeterminacy with regard to technical practices of the *use* of quantum computing. In the words of Ihde (1979), we could say that the *use context* of quantum computers is thus far indeterminate. However, we face no indeterminacy regarding the technical practices concerned with *making* quantum computers (designing hardware components, experimenting with software implementations, etc.), because these clearly link with standards of excellence and life plans. Similarly, we face significantly less indeterminacy with regard to technical practices concerning the *governing* of quantum

computing R&I (e.g. curating research results, securing research facilities, etc.). We therefore argue that though we need to embrace the indeterminacy of technical practices of use when dealing with emerging technologies, and that we should focus on technical practices related to the *making* and *governing* of these technologies. In other words, not yet having sufficient knowledge concerning the use of emerging technologies does not preclude efforts to shape technical practices that cultivates the virtues of making and governing these technologies.

How could the mapping of technical practices be done? First, we need to stress that this mapping is an *iterative* task: that new relevant technical practices can be uncovered on the way, as new narrative structures are taken into account. For instance, a researcher might at some point provide an account of a new type of use of a technology that was not thought of before, and can thereby add this technical practice to the network. Second, each technology should be considered regarding technical practices of *making, using,* and *governing,* and not merely with regard to use. Even though these types of technical practices all have the same structure (for instance, both the *making* of software for a smartphone and the *use* of a smartphone to have a conversation have their related standards of excellence and life plans), they relate differently to the technical object in question. Making and using refers to technical practices that involve the first-order technological configuration of a particular technology. *Making* refers to technical practices concerned with the being-completed of the technical object, for instance designing the camera for a smart border control system. *Using* refers to technical practices concerned with the in-order-to of a technical object, for instance putting one's passport in the smart border control system in-order-to cross the border control. *Governing,* then, refers to the broad category of technical practices involved in the second-order technological configuration of a particular technology (i.e. concerned with narratives *about* the technology), for instance technical practices that regulate the collection and curating of personal data at the smart border control. Governing thereby does not merely refer to technical practices of domination (e.g. making laws to govern the use of a technology), but also to practices of power-in-common (e.g. online protests against laws that aim to govern the use of a technology). Even though the types of *making, using* and *governing* often overlap (e.g. in *making* a camera, an R&I practitioner *uses* a 3D printer), these types of technical practices need to be distinguished because they enable one to uncover technical practices in which a technology is involved from three distinct **points of view**.

We can formulate the first set of procedural steps of our method, aimed at gathering together a network of technical practices:

- Iteratively **map** technical practices in which a technology is involved.

- Consider possible technical practices according to the **points of view** of *making, using,* and *governing* this technology.

2. Gathering narratives

The **second stage** of our method is concerned with **gathering narratives** of particular technical practices. Once different technical practices have been identified, they need to be made explicit. That is, in order to *understand* the technical practice, one needs to know the narrative structures through which it can be explained. These narrative structures need to be grounded in what Aristotle calls common beliefs. This implies that a **selection of stakeholders** has to be made, which might include any practitioner who engages in the technical practices making, using, or governing a technology. For instance, in the case of smart border controls, users are people passing through these passport controls in airports and other places where border controls are implemented. Makers can be engineers contributing to the design of the system, the software for the system, or building the infrastructure needed for the system. The people engaging in technical practices of governing can be government officials in ministries of foreign affairs, or members of parliament who propose legislation specifically targeted at regulating smart border controls. When considering the category of governing, also peripheral stakeholders should be taken into account, meaning those stakeholders that have little to say about a particular technology but that do have a "stake" in the way it is made and used. For instance, people who don't use social media (i.e. "non-users") have a stake in the use of social media because of the so-called network-effect, meaning that by not using social media they can be left out of significant social communication (Hargittai, 2007). As with the mapping of technical practices, the selection of stakeholders is an iterative process, meaning that new stakeholders can be encountered along the way.

In order to uncover the *prefigured time* of technological emplotment with regard to particular practices, which implies **gathering second-order narratives** *about* a technology, one can look at different established sources, which can come from the media, academic opinion pieces, literature, or other significant spaces where stories are expressed such as the "blogosphere" on the Internet (Adamic & Glance, 2005). Arguably, philosophy cannot provide any strict in- or exclusion criteria for the narratives to be considered, because the much-contested criterion of relevance would need to be the differentiating factor (e.g.

relevance according to popularity, expert opinion, and so forth). We will therefore not determine any particular way of establishing relevance, and argue that practitioners should venture outside of philosophy to find proper tools to do so (e.g. trend analysis or bibliometric analysis). Again, however, it makes sense to look for technical practices of making, using and governing, amongst other reasons because narratives concerning technical practices of using are often unavailable when considering an emerging technology. In such a case, it is more valuable to look for instance at technical practices of making, such as the ones explored in an article about the working environment in Silicon Valley tech companies (Weigel, 2017). In the case of smart border controls, which are already being implemented, we can for instance consider the comment piece in Nature about automated borders as a source of second-order narratives (Clavell, 2017). This piece offers accounts of technical practices of using (e.g. "traveling") and of governing (e.g. "monitoring"), and also provides certain standards of excellence, for instance for monitoring that it should involve "a rapid and straightforward verification" and that it "should not discriminate" (Clavell, 2017, p. 35).

In order to subsequently uncover the *configured time* of technological emplotment, one needs to gather narratives that recount the engagement in a technical practice, for instance narratives of someone recounting queuing for a passport control, walking into the booth of a smart border control system, sliding the passport into the system, waiting for one's face to be recognised, and so forth. In other words, one should **gather first-order narrative**s, coming directly from those engaging in a technical practice. In the case of smart border controls, one would want to gather narratives concerned with the process of technological emplotment from makers (e.g. the people assembling the technology or designing its infrastructural support), users (e.g. travellers moving through international airports) and governors (e.g. border police officers operating the systems). Because the process of technological emplotment relates to the *design* of a technology (i.e. design features that mediate the temporality and distancing in the process of emplotment), first-order narratives should involve aspects of design, which implies that the person gathering them should ideally have a certain level of knowledge of a particular technology design. When considering emerging technologies, it is especially crucial to involve makers (i.e. R&I practitioners), because they have the earliest and most well developed access to the first-order narratives in the process of technological emplotment (cf. Brey, 2000).

At this point, we can formulate the second set of procedural steps, concerned with gathering the narratives that will allow for the hermeneutic analysis:

- **Select** stakeholders, iteratively, and according to the three overall categories of *users, makers,* and *governors.*

- **Gather second-order narratives** from sources that present stories *about* a technology.

- **Gather first-order narratives** from stakeholders engaging with the process of technological emplotment.

3. Hermeneutic analysis

Once the relevant narratives concerning technical practices have been gathered, they should in the **third stage** be analysed according to the narrative technologies approach. This practice of analysing a technology incorporates the four hermeneutic concepts of *textuality, literacy, temporality,* and *distancing.* The interpretation of a technology according to each of these concepts depends on the human interpreter, but can be structured according to a number of steps. First, concerning **textuality**, an interpreter should look at the differences between prefigured and configured narrative structures. For instance, if a narrative about a bridge (e.g. recounting that the bridge enables the trade from one side to the other side of the river) is similar to the recounting of the actual use of the bridge (e.g. "I cross the bridge every day to deliver my goods to the company at the other side"), the interpreter could conclude that the process of emplotment involves a low degree of textuality. Conversely, if the narrative about a technology (e.g. recounting that driving a car implies following certain traffic rules in particular situations) and the narrative about the use of a technology (e.g. "I needed to change gears and drive in a lower speed because of bad visibility at that particular corner"), are distinct, the interpreter could conclude that the process of emplotment involves a higher degree of textuality. This difference can also be established through inquiries into cultural references that imply a distinction between users and non-users, such as those implying the distinction between digital natives and non-natives, which points at a high level of textuality. The degree of textuality shows the interpreter *where to look:* whether to merely investigate narratives *about* a technology or also narratives concerning the technological emplotment *by* a technology.

Second, concerning **literacy**, an interpreter should consider the types of stakeholders, relating to the making, using and governing of a technology. For instance, in the case of nuclear weapons, only few makers (e.g. nuclear scientists) and users (e.g. military personnel with high-level clearance) can be considered, but many people can be considered engaging in

practices related to the governance of these technologies (e.g. members of the International Atomic Energy Agency, people in protest groups, members of parliament, etc.). From this, one could conclude that the degree of literacy concerning nuclear weapons is low. Conversely, when considering smart border controls, many users (i.e. travellers) and a considerable number of makers can be considered, which suggests a high degree of literacy. The degree of literacy shows the interpreter *to whom to look:* whether one should engage with experts or rather more with non-expert stakeholders. In order to further investigate the process of technological *configuration* (i.e. the mediating stage between prefigured and refigured time), an interpreter should focus on its dimensions of temporality and distancing. These dimensions relate to technical practices in each case. For instance, in chapters 5 and 6 we mentioned aspects of the design of a car as contributing to the temporality of the technical practice of driving. Similarly, we argued in chapter 5 that design features of electronic derivative trading systems contribute to the abstraction from the world of action in the technical practice of derivative trading.

Concerning the **temporality** of the plot, an interpreter should focus on the way in which design features organise human action either according to a chronological sequence of actions or according to a non-chronological ordering of actions. As we saw in chapter 5, she can thereby focus on the succession of events or on the presence of connectors (i.e. dating mechanisms, version control mechanisms, tracing mechanisms). In the case of smart border control systems, for instance, an interpreter could focus on the following account provided by a traveller: "I first had to walk through the little glass door, stand on the two footprints on the floor, and place my passport in the machine. I then had to look at a certain point to have my picture taken". Similarly, a border control officer could recount: "each passing traveller gets a unique ID connected to the time-stamp of border crossing". These accounts indicate that regarding the technical practice of "border crossing", the smart border control system predominantly configures the chronological dimension of time, enforcing a strict sequence of actions.

Concerning the **distancing** of the plot, an interpreter should aim at uncovering crucial entities and events that are either close to the world of action or abstracting from it. As we argued in chapter 5, abstraction leads to a process of *forgetting*, for instance forgetting that through algorithmic trading of a certain type of electronic bond, house owners might be affected. To revert this process, one has to start from the abstracted level at which quasi-entities and quasi-events operate. In the case of smart border controls, quasi-entities are for instance digital profiles that result from the process of abstraction of concrete individual

travellers into aggregates of demographic profiles of these travellers. By investigating the origin of these quasi-entities (i.e. referring-back to the world of action), the interpreter will be able to uncover how these entities affect actual people engaging in technical practices. For instance, when a traveller with a certain physical appearance is stopped at the border, this could be explained by referring-back from the technological configuration involving an abstracted profile (e.g. representing the aggregate of properties of a "suspicious individual") to the world of action in which an actual person is affected.

The third set of procedural steps applies the narrative technologies approach, and offers the interpretative framework supporting the evaluation of technical practices:

- Determine where to look by interpreting the **textuality** of a technology, focusing on the difference between narratives about and by technological emplotment.
- Determine to whom to look by interpreting the **literacy** of a technology, focusing on the difference between stakeholders relating to the process of technological emplotment.
- Investigate the **temporality** of the process of technological configuration, focusing on the organisation of events and the presence of connectors.
- Investigate **distancing** in the process of technological configuration, referring-back from quasi-entities and quasi-events to the world of action.

7.1.4 Phase II: Interpretation < > Evaluation

The second phase of our novel method concerns the transition from interpretation of technical practices in accordance with the narrative technologies approach to their evaluation in accordance with the narrative unity of life. In line with the theory of technical practice that cultivates the virtues as laid down in chapter 6, this first of all involves interpreting the move from basic actions to technical practices, according to standards of excellence. Secondly, it involves the interpretation of life plans, as those movements that make technical practices intelligible. Thirdly, it involves a consideration of the narrative unity of life to which technical practices and life plans are related.

4. Standards of excellence

The **fourth stage** in our method concerns the **uncovering** of **standards of excellence**, which again is an iterative exercise because new standards of excellence can be uncovered on the way. For this exercise, the hermeneutic analysis should be used. First, the degree of textuality

indicates whether to take design into account in considering the standards of excellence or whether to merely take narratives into account *about* a technology. Considering smart border control systems, we will have to inquire into the process of technological configuration to uncover certain standards of excellence. Second, the degree of literacy indicates which stakeholders can provide valuable insight into the existence and contents of standards of excellence. For instance, it indicates that travellers going through airports are important stakeholders to consider with regard to smart border control systems. Third, temporality indicates the degree to which standards of excellence imply a strict succession of events or a course of action that can be freely chosen. In the case of smart border controls, most standards of excellence of the use of such systems relate to the strict succession of events. Fourth, distancing indicates the degree to which standards of excellence might be forgotten. Considering smart border control systems, this might imply that by referring back from the abstract profiles that are used by border police officials to concrete individuals, decisions based on these profiles could imply a forgetting of the standard of excellence of treating each passing traveller according to her personal circumstances.

We need to recall that standards of excellence are initially considered to be constitutive rules, which are not yet strictly speaking "ethical" and therefore merely provide a basis for the way technical practices can be evaluated. For instance, a constitutive rule for the technical practice of "crossing the border" might be "walking into the booth of the smart border control at a regular pace counts as a valid way to initiate the process of border crossing". To uncover constitutive rules related to practices, relevant stakeholders should be asked about the way in which they perform certain technical practices. Recounts of these standards of excellence can already be found in the narratives that were gathered in the first phase, but could additionally be gained from focus groups (Bloor et al., 2001) or narrative interviewing (Jovchelovitch & Bauer, 2000). Channelled questioning of the way of the technical practice of border crossing could for instance lead to an account like the following: "in crossing the border, I try to keep calm and move through the border control quickly, not to keep up other people in line". From this, one could derive "keeping calm" as part of a constitutive rule that mediates the technical practice of border crossing.

It is important to note that one should not yet take constitutive rules as properly "ethical" standards of excellence. In many cases, constitutive rules can actually obstruct the cultivation of virtues, for instance a constitutive rule such as: "looking away counts as a valid response when witnessing an assault in a public space". However, constitutive rules add to the understanding of "common beliefs": to the everyday understanding of a technical

practice. For this reason, one should keep the constitutive rules *as* standards of excellence **open to critique and revision**. For instance, later considerations of the ethical aim could lead to a transformation of the above-mentioned standard of excellence into: "gathering support and attending to a victim counts as a good response when witnessing an assault in a public space".

The procedural steps that can be derived from this initial consideration of standards of excellence are the following:

- **Uncover constitutive rules** of technical practices in question, by asking relevant stakeholders about the way they engage in these practices.
- **Keep** constitutive rules as standards of excellence **open to critique and revision**.

5. Life plans

The **fifth stage** of our method concerns the **uncovering of life plans**, or the movements of emplotment through which technical practices are made intelligible. As discussed in chapter 6, life plans are characterised by the movements of ascending complexification and descending specification. These provide explanations of *why* agents engage in certain technical practices. It is important to note that new standards of excellence can be uncovered in the consideration of life plans. This can be illustrated by discussing the notion of integrity that comes along with being a border police officer, which might be expressed in a narrative such as: "I'm working at the border control because I chose to be a police officer and serve the public interest". Related to this life plan, standard of excellence for the configuration of smart border controls could emerge, such as: "the system should not discriminate between people of different ethnic backgrounds".

To uncover life plans, the movements of ascending complexification and descending specification need to be made explicit. To do so, an investigator should inquire how narrative structures that constitute life plans include accounts of how basic actions explain more or less remote ideals and vice-versa, how more or less remote ideals explain basic actions. For instance, a person going through the border control could recount: "I had to keep standing in place and look at the camera while it tried to capture my face without success. This experience violated my sense of justice". Such a statement would indicate a link between basic actions (i.e. standing, looking) with the remote ideal of justice. Conversely, an engineer

might explain: "I recently had a discussion with my colleague about privacy, after which we decided to change some settings of the smart border control, giving border control officers less insight into irrelevant details of travellers once they touch the screen to prompt a traveller's profile". Such an account links the remote ideal of privacy with basic actions, of touching the screen.

We derive a fifth set of procedural steps from the discussion of life plans, focusing on the movement of emplotment:

- **Uncover life plans** of technical practices in question, by asking relevant stakeholders to explain why they engage in these practices.
- Explicate life plans through the movements of **ascending complexification** and **descending specification**, focusing on basic actions that are explained through more or less remote ideals, and vice-versa.

6. Narrative unity of life

Engaging in a first **evaluation of the narrative unity of life** constitutes the **sixth stage** of our method. This evaluation is primarily aimed at understanding whether technical practices nest their in-order-to with regard to their for-the-sake-of-which. In other words, it establishes whether a technical practice can be linked with a notion of living together in a political community, which is necessary for it to be qualified as virtuous. Looking back at our discussion of Heidegger in chapter 6, we state that this step allows us to differentiate between the mere application of a technique and a technical practice. More strongly, we could state that this difference denotes the distinction introduced by Arendt in her later work, between the "banality of evil" (Arendt, 1964) and "acting viciously". That is, even before the possibility for acting virtuously or viciously is opened up, there is an "inability to think" (Ezra, 2007, p. 141) in the mere application of a technique. This notion coincides with what Heidegger calls the "supreme danger" of technology (Heidegger, 1977, p. 26), and for both Arendt and Heidegger this implies that the in-order-to of a technical activity is indefinitely extended. In other words, whenever a practitioner is unable to link the in-order-to of her technical practice with its for-the-sake-of-which, she is unable to act virtuously or viciously but is able to engage in banality, which is dangerous considering the growing power of technology. Having a narrative unity of life that supplies one with the for-the-sake-of of a technical practice is therefore necessary for such a practice to be called "virtuous".

To **uncover** the "unstable mixture of fabulation and actual experience" (Ricoeur, 1992, p. 162) that constitutes the **narrative unity of life** to which a practitioner can appeal when recounting technical practices and life plans, one should focus on elements of the narrative structures that recount how one wants to live his or her life. As such, this narrative unity is the basis for a naïve understanding of the good life. As can be derived from Ricoeur's characterisation of the narrative unity of life, it is likely to appear as a fragmentary structure that draws from different subordinate elements, which might be standards of excellence or life plans. For instance, an employee of a civil society organisation that criticises the use of smart border controls might state the following: "I was visiting my sick grandmother when the system kept me waiting in the airport for no apparent reason, which caused me to arrive too late. This experience led me to oppose the use of smart borders, because they violate my idea of how people should treat one another in this country. I believe everyone should be able to appeal to a person's judgement when having to pass an arbitrary border." Such an account contains the recounting of a life plan and constitutes what might be called a fragment of a narrative unity of life, for it supports one's practice by appealing to a notion of living together in a political community. Even though this initial notion of the narrative unity of a life does not yet offer any grounds for saying *how* a technical practice contributes to the good life, it does indicate *whether* a technical practice relates to a notion of the good life, or whether we are dealing with a mere application of a technique.

The discussion of the narrative unity of life provides the method with a sixth set of procedural steps, which focuses on the mixture of fabulation and actual experience that makes up the narrative unity:

- **Uncover** elements that constitute the **narrative unity of life**, by inquiring into accounts of living together in a political community.
- **Establish** whether technical practices link to a narrative unity of life, or whether one deals with the mere application of a technique.

7.1.5   Phase III: Evaluation < > Prescription

The third general phase of our method is concerned with the move from an evaluation of the narrative unity of life to *prescription* in accordance with the ethical aim, which establishes a *stable* for-the-sake-of-which of technical practices. This exercise leads to a basis for prescription because it allows for the establishment of grounds (i.e. reasons) for imputation of a person with regard to her actions, of responsibility with regard to others, and of recognition

with regard to unknown others. In other words, after this final phase we can formulate prescriptions for an agent to engage in technical practice that cultivates the virtues concerning the extent to which this practice supports the good life, with and for others, in just institutions. In this phase, we explicitly reconnect with the heuristic of virtues that we presented in chapter 4. We will do so by connecting the self-regarding virtues (e.g. courage) in that heuristic to an understanding of the good life, other-regarding virtues (e.g. friendship) to an understanding of "with and for others" and virtues concerning unknown others (e.g. justice) to an understanding of "in just institutions". Practical forms of prescription based on considerations of the ethical aim might take many different forms and do not only refer merely to deontic categories of obligation, permission, or prohibition. In addition, prescription might take the form of educational guidelines, design specifications, constitutional norms, procedures for just deliberation, and so forth.

In accordance with the explication of the ethical aim in chapter 6, each stage of phase 3 will contain a threefold structure: starting with the naïve conception of the good life, moving through the sieve of the norm, and reflecting on the norm on the basis of a refined notion of the ethical aim. The first segment of each stage will be concerned with the "naïve" conception of the good life that serves as the narrative unity of life in relation to technical practices. As such, it makes sure that we are not dealing with a mere application of a technique but with a genuine technical practice that incorporates standards of excellence and life plans in relation to a narrative unity of life. The second segment of each stage will be concerned with the formulation of the basis for a normative structure that might govern a particular technical practice. This normative structure would ensure that the possibility for evil is minimised. The third segment of each stage will be concerned with the formulation of forms of arbitration and will therefore underlie grounds for shaping *participation* in practising ethics in R&I. As such, it presents a basis for (1) *how* participation should take place, (2) *who* should participate, and (3) *how* participation can be legitimised. As such, participation is the corner stone of the method to arrive at considered conviction, critical solicitude and a sense of justice.

We need to emphasise two issues related to the recommendations that were formulated in chapter 3. First, "participatory design" in the context of our method does not merely mean that people design a technology together, but more importantly that people have a recognised voice in how a technology should be designed. This implies that participatory design cannot be accomplished without democratic governance that offers participants a basis from which to question design decisions, to search proper guidance in making design

decisions, and to appeal to just, democratic decision-making procedures. Second, even though we respond to the issue of "value conflicts", the conflicts that we deal with – between autonomy, respect, institutions, and the refined ethical aim – are explicitly *not* value conflicts, but conflicts between conceptions of the good and obligations, or between norms and the good life. As such, it is not values – which Ricoeur designated as being "compromise terms" – that conflict, but actualisations of conceptions of the good life in norms, laws and institutions and the tension between these and a reflection back on a better informed notion of the good life that is the basis for arbitration to solve conflicts.

7. The good life

The **seventh stage** of our method offers ways for reflecting on technical practices that link these practices to a notion of *the good life*. At this point, we reconnect with the heuristic of virtues offered in chapter 4, because they offer points of entry for thinking about self-esteem. We will consider prescription of different aspects of technical practice, according to Ricoeur's notion of the good life in terms of self-esteem, self-respect, and considered conviction.

i. Self-esteem

Self-esteem refers to the extent to which a practitioner can evaluate her actions in relation to a naïve conception of the good life. This is where the virtues of character come into play, which pertain to individual conduct. In line with Vallor's (2016) heuristic of virtues, we argue that a practitioner needs to possess the self-regarded virtues of *humility, courage, perspective, self-control* and *magnanimity* in order to gain self-esteem. Our understanding of a technical practice that was developed in phases 1 and 2 assists in understanding how a technology either obstructs or supports the cultivation of these virtues. For instance, regarding magnanimity, we could argue that because of the strict enforcement of the sequence of events, a traveller is obstructed in showing leadership when this might be needed while engaging in the technical practice of border crossing. For instance, she will be less inclined to encourage bystanders to object to the treatment of a fellow traveller, both because she faces a strict succession of events that she cannot reasonably interfere with and because this precludes her to appeal to another person (i.e. a member of the border police), whose leadership she might call into question. As such, smart border controls could obstruct her sense of self-esteem while engaging in the technical practice of crossing the border.

The primary forms of prescription in relation to self-esteem are forms of personal development or *mentorship*. For instance, as we already alluded to in chapter 6, human resources departments, which might better be called human development departments, play an important role in making sure R&I practitioners can relate the technical practices they engage in to a naïve conception of the good life. For instance, they aim to ensure that people enjoy the work they do for the good reasons (i.e. not merely for monetary gain but also for the conviction that doing the work is the right thing to do). In common words, one could say that mentorship should lead practitioners to have a sense of purpose in what they do, which is the bare minimum for engaging in a technical practice that cultivates the virtues.

ii.   Self-respect

Self-respect refers to the extent to which a practitioner can evaluate her actions in relation to a norm that she gives to herself. This norm derives from the necessity to limit one's actions in accordance with the facticity of evil. Consider for instance the movement of descending specification of a border police officer having a conversation with his colleague about privacy, leading him to change the settings of the smart border control system. By doing so, he gives himself a norm according to which he could evaluate the standards of excellence of acting while being a police officer, which in relation to the naïve conception of the good life relates to protecting the public. In this context, "privacy" can therefore be understood as the norm that limits "security" as a public good.

The primary form of prescription of the norm is that of a *code*. As such, the norm is for example reflected in what used to be Google's first rule in its corporate code of conduct, "don't be evil" (Miller, 2005). Even though our method provides a good basis for actually criticising this rule for not being *sufficient* for ensuring technical practices that cultivate the virtues, the formulation of this rule does illustrate the basis for the norm leading to self-respect. However, in opposition to the Google code of conduct, we need to emphasise that the norm has to be *given* to oneself in order to ensure its enacting in technical practices. This is why we referred to the importance of an *ethical oath* (Blok, 2013) as the proper means for ensuring a practitioner to give the norm to herself. As such, an ethical oath functions as what Ricoeur designates as a "test" (Ricoeur, 1992, p. 224), a means for a practitioner to test whether her naïve conception of the good life invoked in a technical practice does not violate the norm.

iii.   Considered conviction

Considered conviction refers to the extent to which a practitioner is capable of dealing with conflicts between the naïve conception of the good life and the norm she gives herself. Responding to these conflicts requires engaging in finding a reflective equilibrium between an ethics of argumentation and considered convictions that are historically situated. Consider for instance a smart border system that not only lets people halt, wait or pass, but also engages in a close up investigation in search of dangerous or illegal items on a person's body. An engineer might implement a rule that reflects a norm stating that such an automated investigation should be carried out on the grounds of automatically detected suspicious aspects of a person that in no case link up with a person's ethnicity or gender, which should therefore not be recorded. The norm reflected by this rule would therefore limit the technical practice of "securing the border" by implementing a restriction of non-discrimination. However, in a particular case a woman from a certain ethnic background might be harmed by the implementation of this norm because she assigns the male gender to robotic entities and because of her cultural background sees it as immoral to be intimately investigated by a male. In such a case, a conflict arises between the norm and historically situated convictions.

The most important form of prescription following from the consideration of conviction is one of *structured dialogue*. Since above-mentioned conflict happens in public life, its resolution depends on the capacity and inclination of different practitioners to engage with one-another. On the one hand, considered convictions need to be taken into account, which for instance means that the design of a smart border control system should allow a user to override its automated sequence of events and request a dialogue with a human. The person responding (i.e. a member of the border police) could structure the dialogue according to principles as laid out in discourse ethics, which would allow her to consider the traveller's conviction while also upholding the interest of the public; ensuring a sufficient level of security. In the case described above, a structured dialogue might for instance lead to the decision to have a female agent perform the security check.

The above-mentioned discussion dealing with the good life provides us with the following, seventh set of procedural steps:

- **Prescribe** forms of **mentorship** that promote the relevant virtues of humility, courage, perspective, self-control and magnanimity, which incline practitioners to have self-esteem while engaging in a technical practice.

- **Prescribe** norms in the form of an **ethical oath**, which allows practitioners to gain self-respect by giving themselves those norms in a public setting.
- **Prescribe** principles for **structured dialogue,** which should mediate conflicts between norms and historically situated convictions, to arrive at considered convictions.

8. With and for others

The **eight stage** of our method offers ways to reflect on the way in which technical practices link up with a notion of the good life, not merely for the self being worthy of esteem but also with and for others. Again, we start by considering the naïve conception of "with and for others" in the notion of reciprocity, by integrating the other-regarding virtues developed by Vallor. From this initial notion of reciprocity, we move through the sieve of the norm of respect for persons towards the reflective notion of critical solicitude.

i. Reciprocity

Reciprocity refers to the extent to which a practitioner interacts with others (e.g. with colleagues, customers, and so forth) in relation to a naïve conception of the good life *with and for others*. The other-regarding virtues of character allow us to evaluate a technical practice in this regard, which in Vallor's approach are *care, friendship* and *honesty*. We encounter these virtues in Ricoeur's work as well, where the virtue of friendship initiates the discussion concerning the other-regarding character of the ethical aim, where honesty comes into play while considering the virtue of "keeping one's word" (Ricoeur, 1992, p. 118) as being constitutive of selfhood, and where care is reconceptualised as solicitude. We can relate these other-regarding virtues of character with an analysis of technical practice. For instance, we could argue that when a border policy officer reconfigures the technical practice of "crossing the border" after having had a discussion with his colleague about privacy, the other-regarding virtue of friendship is at stake. That is, the friendship that binds him with his colleague inclines him to engage with the for-the-sake-of-which of the technical practice that they together engage in, which are linked together in one or multiple life plans.

The principal form of prescription to establish the most basic form of reciprocity is what in management literature is usually referred to as "team building" (Liebowitz & de Meuse, 1982), but which in broader society can be captured instead with the notion of "community building" (Lichterman, 1995). Both these types of prescribed activities are

focused on cultivating the other-regarding virtues. They thereby contribute to the bare minimum for people engaging in technical practices together to do so with and for other people.

ii.   *Respect for persons*

Respect for persons refers to the extent to which practitioners inter-act in a technical practice while treating each other never merely as a means, balancing agency and patiency. This places a limit on technical practices, by making sure through the inauguration of a norm that respect for persons will be upheld. We can understand how this actualisation of respect through the norm comes about by considering our understanding of a technical practice. Considering smart border control systems, we could argue that when faced with the question of automation, engineers ought to take into account the involvement of human border police officers in securing the border. Balancing their agency and patiency would mean that their voices ought to count in making a decision on automation instead of reducing such a decision to mere efficiency considerations, which would reduce practitioners to mere means. This would lead to a clear choice between technological alternatives, where a symbiosis between man and machine would be preferred over full automation, while considering a particular technical practice of securing the border.

Since we again deal with a *test* when considering the sieve of the norm, we encounter the test contained in categorical imperative as the primary form of prescription following from respect for persons. That is, to incorporate respect for persons in the ethical aim, practitioners ought to consider the categorical imperative when engaging in a certain technical practice. In accordance with Ricoeur (1992, p. 222), it is the second formulation of the categorical imperative that functions as the test: "act in such as way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end" (Kant, 2002, p. 96). Such a test could be integrated in a code of conduct, or in a protocol that would govern the engagement in a certain technical practice. "Following the protocol" would in such a case include concerns expressed by the categorical imperative in order to ensure respect for persons.

iii.   *Critical solicitude*

Critical solicitude refers to the extent to which a practitioner engaging in a technical practice is aware of and responsive to the consideration of vulnerable persons, whose personhood is subject to ambiguity. Vulnerable persons are important to take into account explicitly because

they can be affected by conflicts between the norm (e.g. the above-mentioned protocol) and the ethical aim. That is, in some cases the norm might prescribe a certain course of action that conflicts with a practitioner's conception of the good life with and for others. When considering smart border control systems, we argue that such a situation might arise when a traveller is kept at an airport for being stateless (i.e. not having a passport), whose reason for traveling is prosecution by a violent regime. A standard of excellence of the technical practice of securing the border could be that people who are able to make a credible case for being under threat if they are not let through should be let through, even if this means acting against the protocol. A potential problem of smart border control systems is that they do not allow for such a decision based on ambiguity for they simply keep a person from moving when no passport is presented.

The primary form of prescription that follows from these considerations of critical solicitude is the involvement of the *phronimoi* in making a prudent decision regarding ambiguous cases. This means that *expert guidance* is called for in such cases, but expertise should explicitly not be understood as having familiarity in a technique. Rather, expertise should imply a proper level of experience in a practice (e.g. having dealt with refugees trying to cross the border for a long time), and relevant knowledge (e.g. having knowledge of applicable laws and human rights declarations). As argued in chapter 6, experts should acknowledge that the dilemma they face concerns the principle of respect for persons and that they should aim at finding the mean, which might entail enforcing a dichotomy (e.g. deciding on letting someone pass versus keeping someone at the border). This consideration points at ways in which a proper symbiosis between man and machine might be achieved, for the efficiency gained with installing smart border control systems could be complemented by a greater attention for ways to interrupt the process when ambiguity arises and employing proper experts to deal with ambiguous cases on the spot.

Having considered the good life, with and for others in relation to technical practices, we are able to formulate the eighth set of procedural steps for our method:
- **Prescribe** forms of **community building** that promote the relevant virtues of care, friendship and honesty, which incline practitioners to attend to each other reciprocally.
- **Prescribe** forms of **codes** or protocols of **conduct** that ensure practitioners to respect persons while engaging in a technical practice.

- **Prescribe** forms of **expert guidance**, which should mediate conflicts arising from ambiguity in dealing with vulnerable persons, arriving at critical solicitude.

9. In just institutions

The **ninth** and final **stage** of our method is to the furthest extent concerned with the political dimension, with the realm of institutions. As we argued in chapter 6, institutions give a temporal dimension to power in common, but also at all times risk being perverted. Therefore, technical practices that cultivate the virtues have to link to the good life *in just institutions*. The institutional dimension cannot be ignored in ethics, because in the form of *domination* it impacts technical practices, be-it regarding the design of technologies involved in those practices, the education provided to promote virtues of practitioners, and so forth. We start again with a naïve understanding of just institutions, considered in relation to the virtues of character that shape it, and move through the sieve of the norm of distributive justice to the sense of justice, which demands democratic legitimation.

i.   Equality

Equality refers to the extent to which practitioners engaging in technical practices relate to unknown others (i.e. the stranger) in accordance with a naïve conception of the good life in just institutions. The relevant *political* virtues of character are, in line with Vallor's heuristic, *justice*, *civility*, and *flexibility*. We can relate these virtues to Ricoeur's theory, first because civility can be considered as the politicised version of the "ethics of argumentation", or the ethics of civil discourse that we encountered before while considering the good life. Justice, in Vallor, connotes with equality in Ricoeur, for it is concerned with distribution based on the extension of the good towards to common good, and flexibility in Vallor connotes with Ricoeur's notion of sense of justice, because it is focuses on arbitration between spheres of justice that are historically situated. We can relate these political virtues to an analysis of technical practice. Concerning smart border controls, we argue that in the related technical practice of crossing the border the virtue of civility, or "making common cause" (Vallor, 2016, p. 141) is at stake because the automated system physically separates a person from other people waiting in line, making it difficult to engage in discourse to make common cause when the situation calls for it.

The most prominent form of prescription related to equality in accordance with the virtues of character of civility, justice, and flexibility, is that of *civic education*. Civic education is one of the primary ways of instilling practitioners with political virtues relevant

for establishing the for-the-sake-of-which of the technical practices they engage in (cf. Gutmann, 1995). This means that practitioners in R&I contexts should receive schooling that explicitly takes into account the political virtues and their relation to particular technical practices, and should have the means for being vigilant in civil life – for instance by being members of trade unions.

ii.   Rule of justice

The rule of justice refers to the extent to which practitioners take principles of distributive justice into account when establishing the for-the-sake-of-which of a technical practice. This places a limit on a technical practice, by ensuring that it does not involve technological configurations that lead to inequalities that are unjustifiable. This aspect of the ethical aim is of particular importance when considering the design of smart border controls. For instance, regarding the practice of "data controlling" of data gathered by smart border control systems, a practitioner might question the abstractions brought about in the process of technological configuration. That is, the system might be triggered to react on certain personalised profiles, thereby affecting the world of action by means of abstract entities. This process would constitute a movement of descending specification, of remote ideals incorporated in an ontology underlying a profiling system (i.e. system bias) towards basic actions (e.g. a traveller having to stay in place). This might violate the rule of justice concerning the technical practice of data controlling, because those affected by the practice could be subject to an arrangement of social and economic inequalities that is not in everyone's advantage (e.g. be halted because of their ethnic profile).

This is the first time we arrive at *legal regulations* as the primary form of prescription in accordance with the rule of justice. This does not imply that only policy changes can follow from considerations of the rule of justice, because it can equally well translate into design requirements, such as regulations regarding material use, privacy settings in software design, minimum safety requirements, and so forth. However, the focus on regulations does imply that considerations of the rule of justice cannot be merely left to a practitioner's own devices and decisions. Rather, it is the political community that is supposed to formulate and implement regulations that accord with the rule of justice. The role of the practitioner, in this regard, is to (1) pay reasonable respect to the rule of the law that is meant to accord with the rule of justice and to (2) be vigilant in recognising particular moments when the rule of justice is not sufficiently respected and regulations are absent. For this reason, regarding the practitioners, legal regulations should not merely be about the upholding the rule of justice,

but also about the empowerment of practitioners in contributing to its enforcement and refinement. As such, it should for instance include regulations for whistle-blowers, protecting their position and encouraging them to voice their concerns.

iii.   Sense of justice

The sense of justice refers to the extent to which practitioners are capable of arbitrating between competing forms of domination, between the governed and the governing and between groups in civil society, in establishing the for-the-sake-of-which of a technical practice. As such, it brings us closest to the significance of the political in the ethics of technology. Again, it does not merely refer to considerations of decision-making in terms of constitutional or legal procedures, but also relates to design choices, such as the channelling of political speech on Internet forums, the possibility of users to flag digital content and co-decide on what is done with it, and so forth. Design is an increasingly pressing issue in particular because digital technologies are increasingly developed for political purposes: for voting, organising public debates, contracting and decision-making. Considering smart border controls, we argue that arbitration is needed with regard to the domination of states over the people travelling through airports. For instance, smart border controls could arbitrarily halt people from a particular nationality, thereby arbitrating between the powers of the governed (the travellers) vis-à-vis the power of the governing (the state). As such, smart border controls would compromise the ability of civilians to challenge the domination embedded in the technical practice of securing the border, as for instance happened during protests in the United States at airports, where border police officers in some cases refused to carry out the federal executive order signed by the president that ordered them to halt people from certain nationalities (Gambino & Siddiqui, 2017).

The primary form of prescription following from considerations of the rule of justice is captured under the notion of *democratic decision-making*. Democratic decision-making is a complex and contested issue that perhaps will never be fully resolved, but practitioners are able to at least take it into account in technical practices to the furthest extent. That is, by tentatively agreeing on ways in which democratic decision-making can be organised, relevant stakeholders could implement agreed-upon democratic principles in formal policies and procedures as well as in requirements for technology design. Such principles could for instance guide decisions with regard to user-agency in creating digital contents on public forums on the Internet, provide the governed (i.e. users) with the means to challenge the governing (i.e. Internet companies), and put conflicts forward to public debate (i.e. to the

relevant legislators). Such a public debate should be organised in a two-fold way: both (1) by questioning the particular technical practice, informed either by norms of good governance or novel or refined norms, and by (2) questioning the very process shaping the decision-making.

Having considered the good life in just institutions related to technical practices, we have come to the point that we can formulate a ninth set of procedural steps:

- **Prescribe** forms of **civic education** that promote the relevant political virtues of justice, civility and flexibility, which incline practitioners to attend to the establishment of equality.

- **Prescribe** forms of **legal regulations** that ensure practitioners to respect the rule of justice and enhance their ability to challenge violations of this rule.

- **Prescribe** forms of **democratic decision-making** that mediate conflicts by moving them forwards to public debate, according to received and novel norms of good governance while at all times being critical of and questioning the process of decision-making itself, thereby establishing a sense of justice.

The above-mentioned nine steps together form our novel method for practising ethics in R&I. What they make explicit is that practising ethics is not limited to evaluating the "pros" and "cons" of a particular technology, or advising on the desirability of a technology. These aspects are important, but far from sufficient for striving for technical practices that cultivate the virtues for the sake of the good life, with and for others, in just institutions. To cultivate such technical practices the well being, sense of community, and civic education of practitioners needs to be taken into account. For instance, an engineer without any sense of justice cannot be expected to engage in the technical practice of "designing a military drone" in a virtuous manner. Furthermore, and to some extent as a counter-balance to Aristotelian virtue ethics, technical practices that cultivate the virtues cannot persist without personal norms (e.g. through an ethical oath), communal norms (e.g. through codes of conduct) and political norms (e.g. through regulations), imposing limits on those practices. Finally, the inevitability of conflicts between norms and a refined understanding of the good life calls for the necessity of structured dialogue, expert guidance and democratic decision-making as prudent means to establish what one might call "best practices".

The critical reader might comment that the notion of "design" in the novel method is somewhat misleading, because no specific guidance for technology design is provided. In

response to this objection, we reply that by changing the focus of the method from "technology" to "technical practice", the focus and meaning of "design" is also changed by encompassing a technical practice as a whole. That is, "ethical design" of for instance a smart border control system does not *merely* refer to changing the specifications of the actual thing, for instance its sensors, its camera, the software it runs, and so forth. More importantly, it refers to the design of the related technical practices of for instance "crossing the border" or "securing the border". As such, it never suffices to merely change the design of the particular technology to make it "ethical", but it always also requires attending to the standards of excellence of the practices in question, the norms that shape these and the notion of the good life that practitioners can relate to in doing their day-to-day work.

One could moreover object that the procedural steps are still very general and do not yet sufficiently answer the exact question of "how" they can be implemented in practice. For instance, the procedural step concerned with decision-making does not indicate what *kind* of decision-making should be used in technical practices (i.e. representative democracy, direct democracy, citizen councils, and so forth). We provide two answers as a response. The first answer concerns the role of the philosopher in formulating any type of method. We argue that in order for procedural steps to be responsive to a particular technical practice, they require the input from those engaging in and knowledgeable of that practice. This implies that in developing procedural steps further, to make them responsive to for instance the technical practice of designing military drones, a philosopher would need to engage with military personnel, with legislators that have experience in drafting defence-related legislation, and so forth. In other words, further development and refinement of the procedural steps would require an inter- and multidisciplinary effort and cannot solely be a philosophical effort. The second answer concerns the character of more specific procedural steps, which we argue would take us from an understanding of method to an understanding of "tools", such as the Ethics Canvas that we present below. A philosophical method, we argue, should present a general way of doing things that is backed up by a philosophical theory, but should not present definite answers to the "how" question in the form of a tool, understood as a "thing" that enables the in-order-to relation (e.g. "this tool can be used in-order-to formulate a code of ethics"). This is not to say that such tools are unimportant. On the contrary, they are a necessary component of practising ethics in R&I, *but* at the same time should always be subordinate to the philosophical theory and method from which they originate. In other words, tools for practising ethics should not become ends in themselves, and should always be challenged on the basis of philosophical critique and convictions.

7.2 From Method to a Practical Tool: The Ethics Canvas

In this section, we present our practical efforts to construct a tool for practising ethics in R&I and tentatively evaluate the usefulness of this tool in relevant settings. The method presented in the previous section contains a comprehensive[78] set of procedural steps that together could guide the entire process of practising ethics in R&I. However, these steps do not yet offer proper practical guidance for those having to bring the method into practice. For instance, an engineer might understand *what* to do by interpreting these procedural steps, but not necessarily *how* to bring these into practice. Here, we hit upon the difference between method and a tool*, or a thing that allows one to put a procedure into practice. A great variety of tools could be proposed for the different procedural steps, such as focus groups for the uncovering of constitutive rules, citizen councils to democratically deal with conflicts arising from technical practices, and so forth. Devising a complete "tool box" covering each and every procedural step in our method would vastly outstrip the scope of this dissertation. However, in order not to leave the reader completely wanting and not to invoke the idea that none of the procedural steps could actually be practically realised, we propose a practical tool for a particular procedural step in our method: namely for the gathering of narratives (procedural step 2).

To start, we searched for an extensively adopted tool that facilitates accessible and clear processes of discussion and reflection amongst people with little to no expertise in ethics. Consequently, we transformed a business-modelling tool that is widely used in business development practices, the Business Model Canvas (BMC) (Osterwalder & Pigneur, 2010), into the "Ethics Canvas". The Ethics Canvas is a collaborative brainstorming tool that has two distinct aims: (1) to have R&I practitioners come up with and discuss narratives outlying possible *ethical impacts* of the technologies they develop, and (2) to have R&I practitioners consider *pivots* in their technology design or business model to avoid or mitigate the negative ethical impacts. We argue that by engaging in this exercise different narrative structures come to the fore that can be valuable input for the hermeneutic analysis of technical practices. To gain an indication of whether the Ethics Canvas actually works in practice, we empirically assessed its perceived usefulness by putting it into practice in a classroom situation of business & IT students who were developing novel technological

---

[78] With "comprehensive" in this context, we mean that (1) the method is holistic, in that it provides a relatively complete account of technical practice that cultivates the virtues and (2) that procedural steps and their elements link up with one-another.

applications. A follow-up questionnaire that the students filled in provided a number of initial suggestions regarding the usefulness of the Ethics Canvas for practising ethics in R&I.[79]

In what follows, we first discuss methods used to create business models that can offer tools for putting the activity of gathering narratives into practice. Second, we propose the design of the Ethics Canvas that is inspired by the Business Model Canvas (BMC) approach in business modelling research. Third, we explore the merits of the Ethics Canvas in a practical setting by evaluating the outcomes of a questionnaire that students filled in after having engaged in the Ethics Canvas exercise.

7.2.1 The Business Model Canvas

In chapter 3, we selected and justified our choice for the BMC as a useful tool to appropriate for practising ethics in R&I. We stated that our tool should align with three requirements: that it should be (1) closely aligned with concrete research practices in the workplace, (2) easy-to-use in order to be capable of iterative deployment in these practices and (3) as much as possible related to our theory of technology that is rooted in the notion of narrative. In line with these requirements, we have to make sure that the practical threshold for using our tool will be low and that it will be possible to use it without thorough background knowledge of ethical theories or conceptual discussions of goods or obligations. Additionally, the tool should facilitate an open-ended process of interpretation in a collaborative fashion to identify potential ethical impacts of a technology by discussing related technical practices in the narrative mode. Different relevant narrative structures should be brought to the table by means of a collaborative process in which multiple people involved in an R&I process express their expectations of potential ethical impacts in the form of narratives (i.e. for this group of stakeholders, such-and-such feature of our technology can have such-and-such ethical implications). These considerations regarding requirements brought us to the field of business development, or more precisely the field of business model development.

Some discussions of ethics and responsible R&I have already emerged in the field of business model development. For instance, Henriksen et al. (2012) discuss business models that promote sustainable ways of production, or "green business model innovation". In a similar vein, Bocken et al. (2013) explain how a re-definition of the notion of "value" in business models can help to render businesses more sustainable, for instance by focusing on value through the entire supply-chain. However, these approaches primarily focus on the design of the business case (in terms of resources, customers, etc.) and not on technologies

---

[79] This study has been approved by the Trinity College Dublin Research Ethics Committee.

that are developed in R&I processes and the technical practices to which they relate. To change this focus, we will investigate how to transform an existing business model development approach to align it with our aim of constructing a tool for gathering narratives concerned with the technical practices in which a technology is involved.

Before we do so, however, we have to ask to what extent tools in business model development methods incorporate the two requirements we introduced in the previous section: (1) engaging non-ethicists with the disclosure of ethical impacts and (2) facilitating this as a collaborative process of interpretation. We can observe how business models are defined in the literature. As pointed out by Zott et al. (2011, p. 1024), a business model can be understood as an "architecture", a "heuristic logic", a "concise representation" and also a collection of "stories", aimed at describing and explaining how a particular enterprise functions or operates. A spectrum of business models can be identified, with on the one hand business models that are meant to offer a strict representation of both internal and external processes of an existing corporation, and on the other hand models that use stories to give an account of these processes – possibly of businesses that do not already exist (i.e. a model for a start-up). The latter type of business modelling approaches is particularly interesting for our purposes, since it appears to focus on an understanding of business processes in terms of narratives that are constructed through social interaction.

Lucassen et al. (2012, p. 13) use two indicators to capture the two aspects of the above-mentioned spectrum between models that are strictly representational and those that are the result of people's interpretations, in order to review and compare different visual business modelling approaches. They use the notion of "capturing" to indicate to what extent a business modelling method accurately represents a business process, and the notion of "communicating" to indicate to what extent a business-modelling tool is accessible and generates understandable outcomes. They argue that the BMC is most successful regarding the indicator of "communicating", compared to two other established models[80] "because it effectively models explicit information of both tangible and intangible aspects of a business and communicates this information in a highly accessible manner to parties unfamiliar with the modelling technique" (Lucassen et al., 2012, p. 14). As Kuparinen argues, the BMC can be classified as a "narrative business model" (Kuparinen, 2012, p. i), because it enables "participant narratives". The BMC provides a visual-linguistic tool (see figure 4) that can be

---

[80] The two traditional business model approaches that the business model canvas was compared with are the "software ecosystem model" approach and the "board of innovation" approach (Lucassen et al. 2012: 6).

used in a collaborative process in which participants generate ideas by offering and discussing certain narratives that are related to the thematic boxes displayed on the canvas.

Figure 4: The Business Model Canvas (Osterwalder & Pigneur, 2010).
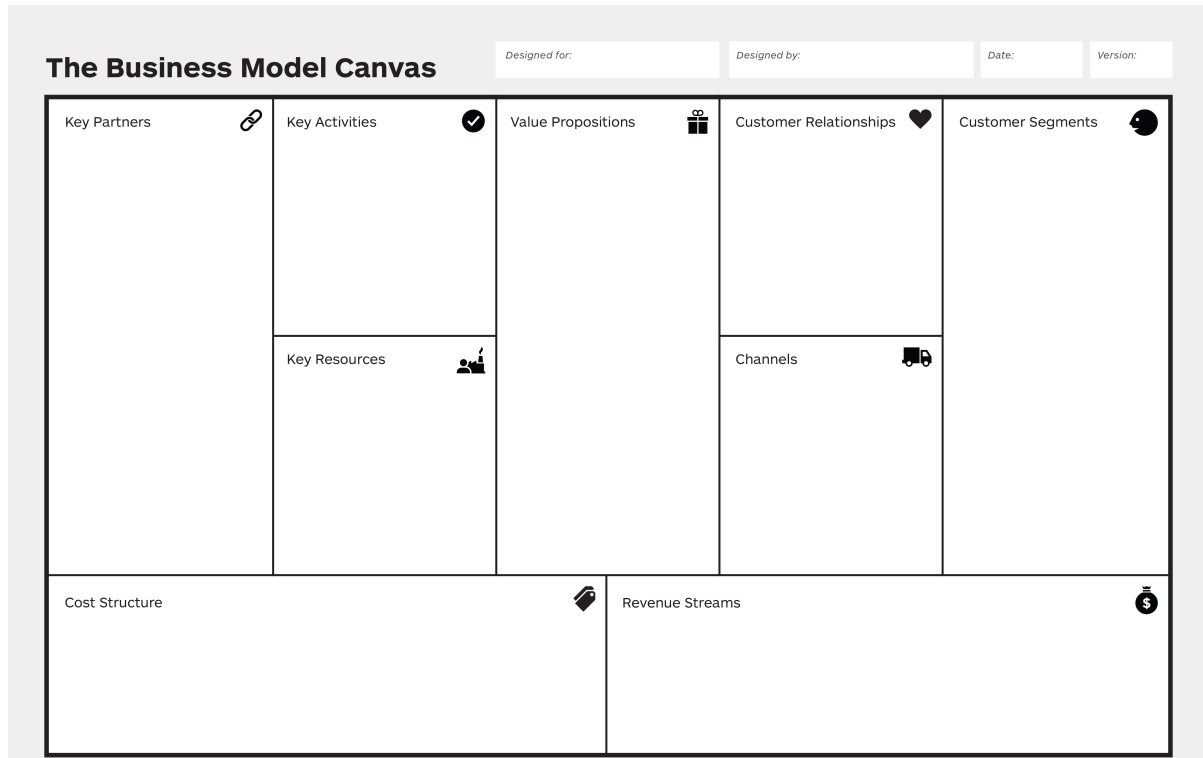
We argue that of the existing business modelling approaches, the BMC best fulfils the requirements we formulated. First, as Lucassen et al. (2012) argue, it is highly accessible and understandable to people without specific knowledge of the field. If the structure of the BMC can be incorporated in a tool for disclosing ethical issues, it would be an answer to Brey's (2000) concern regarding the disclosure level for it allows researchers to engage with ethical reflection in an accessible manner without them having to have thorough knowledge of the field of applied ethics. Second, since the BMC relies on the collaborative generation of participant narratives, it seems to satisfy the second demand to a large extent. It enables participants to engage in a collaborative process of interpreting and discussing business processes. Considering the foregoing arguments, turning towards the BMC to find a novel tool for disclosing ethical impacts in R&I processes seems justified. However, we need to transform the BMC, which is clearly focused on discussing business processes and has little to do with ethics, into a tool that can be used in the context of practicing ethics in R&I. Furthermore, we need to strongly emphasise the pragmatic character of this undertaking.

Because we are drawing from different disciplines and thematic foci, we should not expect the tool to elegantly align with the method we presented in the previous section. Rather, aligning method and tools will likely be a continuing process of reflection and iterative revisions.

7.2.2 Designing the Ethics Canvas

To explain the process of designing the Ethics Canvas, we first briefly describe the BMC and discuss its aims as well as the way in which it is used in a collaborative setting. The BMC was developed by business theorists Osterwalder and Pigneur (2010) as a visual-textual plane that is divided up into nine "building blocks" through which a business model can be described in a holistic manner. It is argued that a business model can be defined as a model that "describes the rationale of how an organisation creates, delivers, and captures value" (Osterwalder & Pigneur, 2010, p. 14), and that this definition can be captured by participants discussing all the "building blocks" of a business model. By engaging in a collaborative discussion about the different building blocks of a business model, such as key partners, channels and revenue streams, participants working with the BMC are able to arrive at a comprehensive understanding of the way in which their organisation is supposed to create, deliver and capture value. In its original form, the BMC is printed on paper and used as the core instrument for a collaborative workshop. In addition to the canvas, Osterwalder and Pigneur (2010) wrote a handbook that provides guidance for the workshop participants in understanding the meaning of the different building blocks and presents use-cases of the canvas as well as techniques for designing better business models. The printed canvas is used as the focal point of a collaborative workshop, with participants discussing and writing down ideas for each of the building blocks. Next to the original BMC, online applications have been developed that offer digital versions[81] of the canvas, through which teams of different sizes can create multiple business models and save them on their accounts.

Taking the BMC as a starting point, we aimed to transform it in a way that would enable its users to discuss how a technology might bring about ethical impacts related to technical practices for different stakeholders instead of discussing a business model. To achieve this, we considered different building blocks that could amount to a holistic discussion of a certain technological application. The building blocks were constructed in a two-way process: (1) by considering literature in Science, Technology and Society studies (STS) and philosophy of technology, and (2) simultaneously engaging in an evaluation and

---

[81] See for instance https://strategyzer.com/ and https://canvanizer.com/new/business-model-canvas

217

design exercise of using the Ethics Canvas to improve its user-friendliness. Each building block consists of a central term and a number of core questions that can guide the discussion concerning a term. Because the development process of the Ethics Canvas happened in parallel with the development of the narrative virtue ethics of technology approach, the heuristic we used for the Ethics Canvas does unfortunately not align with our approach but instead draws from different normative theories of technology. That is, at the time when the most current version of the Ethics Canvas was finalised to be submitted to the practical evaluation exercise our philosophical approach was not yet finished and could therefore not directly inform the design of the Canvas.

Thus, instead of using our narrative virtue ethics of technology approach, we consulted literature containing normative theories of technology that explicates essential aspects of impacts of technologies on individuals, groups and society as a whole. We need to stress that this consultation was mostly aimed at pragmatically gathering different vantage points to consider ethical impacts of technology, and not at providing a coherent theoretical framework underpinning the Ethics Canvas. STS literature offers useful accounts of the ways in which technologies are embedded in relationships between different "relevant social groups" (Pinch & Bijker, 1984, p. 35), which can be types of individuals (e.g. producers, technology users, women, elderly) or institutional, collective actors (e.g. government, companies, labour unions). Akrich (1992) discusses the Actor-Network Theory (ANT) approach and shows how technologies can have impacts on actors that are not directly connected to its design, production or use such as non-users but also non-humans (understood as e.g. the impact of a mobile phone on a supply chain for raw materials). She argues that technologies can politicise social and material relations, which can for instance be made explicit by considering how non-users of social media applications can become marginalised.

In order to subsequently understand *how* technologies impact relevant individuals or groups, we turn to writings in philosophy and technology. Ihde (1990) and Verbeek (2005) show how individuals can change their behaviour or relationships by engaging with technologies. For instance, Verbeek shows how ultrasound technology has transformed the relationship between parents and their unborn child (Verbeek, 2005, p. 140), and how technologies, such as traffic lights and speed bumps, mediate the behaviour of car drivers (Verbeek, 2005, p. 159). These scholars accentuate that ethical impacts are not simply consequences of technological change, but should be understood as impactful relations between human beings and technologies. Feenberg (1999) goes beyond this focus on the technological mediation at the level of the individual, by arguing how technologies can

impact relations *between* people and collectives, for instance between workers and their companies, between governments and labour unions. In line with this, he shows how technologies can impact the public sphere, in which "everyday communicative interactions" take place (Feenberg, 1999, p. 167), in which ideologies are formed and social struggles arise. To consider ethical impacts that are more directly related to the material aspect of technologies, we consulted scholars discussing "constructive technology assessment". These show that technology assessment should take impacts of technologies on the environment and production processes into account (Schot & Rip, 1997). Finally, to provide a bridging step in the move from description (i.e. what *are* the ethical impacts?) to prescription (i.e. what *should* be done?), we turned once more to VSD and included the notion of technical choices driven by value considerations (Friedman et al., 2006) as the end-point of the Ethics Canvas. However, we broadened up the choices to be considered, asking participations to think beyond the technical by also considering organisational changes or changes in policies. To summarise, we inferred the following characterisations of technology impacts from the literature:

- Ethical impacts occur as relations between technologies and different types of actors, which can be types of individuals and types of collectives, or groups.

- Technologies can mediate the behaviour of practitioners, but also the relations that practitioners have with one-another.

- Technologies can mediate collective worldviews of practitioners and can bring about conflicts between social groups.

- Technologies impact the material network in which they are designed, produced and used, including for instance the supply chain it constitutes and the natural resources it needs.

While taking the above-mentioned characterisations of technology impacts as a guideline, we entered into an evaluation and design process of the Ethics Canvas. This design process was aimed at making sure that the rationale of the Ethics Canvas design would not only be grounded in the relevant literature, but that it would also be user-friendly and intuitive. Based on our literature study, we designed nine different versions of the Ethics Canvas, all with different building blocks and layouts. These designs were iterated through a series of analysis exercises conducted by the Ethics Canvas design team[82]. In addition, versions of the Ethics Canvas were trialled in teaching and training settings with over 260 undergraduate and

---

[82] The Ethics Canvas has been developed by Wessel Reijers, David Lewis, Harshvardhan Pandit, Arturo Calvo, Killian Levacher, Andreas Burburdan, and Farnoosh Mohri.

postgraduate students in computer science, engineering, business studies and working on groups on pre-assigned digital application designs. This provided a further source of insight into improving the usability of the Ethics Canvas design. The criteria for success we used during these meetings were that participants (1) should be able to complete the entire Canvas within a reasonable amount of time (a maximum of 1.5 hours) and (2) should be able to address each building block without having to consult any external source.

As a result of this evaluation exercise, some important changes were made concerning the wordings of each box, because some terms in the consulted literature (e.g. actor, human-technology-world relations, ideology) were not intuitive for the users and needed to be translated into concepts that are more easily usable (e.g. group, behaviour, worldview). The table below (table 1) provides an overview of the conceptual framework of the Ethics Canvas, displaying sources in the academic literature and the corresponding approaches that each building block is based on and explicating what changes in terminology were applied to ensure the usability of the Ethics Canvas.

Table 8: Overview of (1) the central questions of the Ethics Canvas, (2) authors consulted to address these, (3) the approaches used by these authors, (4) the changes in wordings applied during the evaluation and design sessions and (5) the final boxes for the Ethics Canvas.

Central question	Literature consulted	Approach	Change in wording	Boxes
Who are affected?	Pinch and Bijker (1984), Akrich (1992)	Actor Network Theory	Relevant social group / actor / actant => individual / group	1) Individuals affected 2) Groups affected
How are stakeholders affected?	Ihde (2009), Verbeek (2005)	Postpheno-menology	"Human-technology-world" relation => behaviour / relations	3) Behaviour 4) Relations
	Feenberg (1999)	Critical Theory of Technology	Ideology => worldviews Struggles => social conflicts	5) Worldviews 6) Social Conflicts
	Schot and Rip (1997)	Constructive Technology Assessment	Risks of products and processes => product or service failure Environmental aspects => Problematic use of resources	7) Product or service failure 8) Problematic use of resources

Central question	Literature consulted	Approach	Change in wording	Boxes
What can be done?	Friedman, Kahn, and Borning (2006)	Value Sensitive Design	Technical choices driven by value-considerations => What can we do?	9) What can we do?

As can be seen in the table, the evaluation exercise led to some important changes concerning the wordings of each box, because some terms use in the consulted literature (e.g. actor, human-technology-world relations, ideology) were not intuitive for the participants and needed to be translated into concepts that are more easily usable (e.g. group, behaviour, worldview). Eventually, the design process brought us to the current design of the Ethics Canvas (figure 5). The Ethics Canvas is organised according to nine thematic blocks that are grouped together according to four different stages of completing the canvas. The first stage (blocks 1 and 2) challenges the participants to consider which types of individuals and groups are relevant stakeholders when considering a specific technology. The second stage (blocks 3 to 6) asks the participants to discuss potential ethical impacts, considering the technical practices of different stakeholders that were identified. The third stage (blocks 7 and 8) asks the participants to consider potential ethical impacts that are not stakeholder specific, pertaining to product or service failure or any problematic use of resources. The fourth stage (block 9) challenges participants to think beyond the potential ethical impacts they discussed and discuss some initial ideas for overcoming these ethical impacts. To complete the Ethics Canvas exercise in a physical space, participants write down their ideas on a printed Canvas, and consult the Ethics Canvas Manual that provides guidance on how to conduct the exercise. An online version of the Ethics Canvas[83] has also been developed. On this platform, people can collaborate to complete a particular Ethics Canvas online while being in different physical places.

---

[83] https://ethicscanvas.org/index.html

Figure 5: The Ethics Canvas, version 1.9.

7.2.3 Assessing the usefulness of the Ethics Canvas

The BMC is a widely used tool for business model development and has been positively assessed in multiple studies. We wanted to similarly assess the Ethics Canvas and its usefulness as a tool that supports practising ethics in R&I settings. Comparing the Ethics Canvas with other tools for practising ethics in R&I is not possible due to lack of similar tools that are used in day-to-day activities of R&I practitioners. Therefore, we assessed the Ethics Canvas by evaluating its perceived usefulness amongst its users and its anticipated effects related to follow-up activities. This assessment and evaluation exercise has an explorative character and aims at providing a tentative indication as to whether the Ethics Canvas is considered as a useful tool for practising ethics in R&I by R&I practitioners. It should therefore not be considered as a empirical test or experiment aiming at gathering conclusive evidence for the usefulness of the Ethics Canvas.

We organised an Ethics Canvas pilot with students who were required to create a new ICT application as part of their coursework. The pilot was part of a course on innovation in the computer science programme at Trinity College Dublin. In this course, students were expected to develop an ICT application and work on a business model for this application.

The students attended a one-hour lecture at which the content of the Ethics Canvas Manual was presented. After this, they were given the assignment to complete the Ethics Canvas in groups, using the online version for their particular R&I project in approximately one hour. Students were free to meet up in a physical space or to hold a conference call for completing the exercise. A total of 109 students participated in the Ethics Canvas exercise, organised into groups that each comprised of 3 or 4 students. After the groups had completed the exercise, all participating students were asked to fill in a questionnaire that asked them about their perception of the usefulness of the Ethics Canvas to practise ethics in their respective R&I projects. Filling in the questionnaire was voluntary. The feedback questionnaire was filled in by 31 students, which represented 28% of the total number of students who worked on the Ethics Canvas exercise. The questionnaire followed a 5-point Likert scale, with a 1-point assessment indicating strong disagreement and a 5-points assessment indicating strong agreement. Statements were formulated in the affirmative mode and as negations to be able to assess whether participants paid attention to the statements. The participants were asked about (1) the perceived usefulness of the Ethics Canvas (e.g. did the Ethics Canvas add to the overall understanding of ethical considerations?) and (2) the anticipated effect of the Ethics Canvas (e.g. did the exercise influence the business model and or technology design?). In what follows, these two aspects are discussed based on reflections on the questionnaire results.

The *perceived usefulness* of the Ethics Canvas was evaluated extensively in the questionnaire. Generally, 56% of the participants agreed and additionally 28% strongly agreed that the exercise improved their understanding of the potential ethical impacts of their R&I projects. Participants were asked whether the Ethics Canvas exercise widened their understanding of different individuals or groups affected by their project, to which 44% of the participants replied that they agreed and 29% that they strongly agreed. On being asked whether the exercise helped to create a broad overview of potential ethical impacts of their project, 42% of the participants stated to agree and 35% to strongly agree. To further the scope of the assessment, the participants were asked whether the ethical impacts they discussed in the task sufficiently fitted the structure of the Ethics Canvas. 40% of the participants agreed that it sufficiently fitted and 21% strongly agreed. To assess the value of the Ethics Canvas in stimulating productive discussions, participants were asked whether they considered any ethical impacts that were not known to them or unclear beforehand. Only 21% of the participants disagreed or disagreed strongly with this question, indicating that the majority of the participants discussed ethical impacts that were new to them. These outcomes

suggest that the Ethics Canvas can be a useful tool to guide participants into discussing ethical impacts that group members didn't know or didn't clearly think about beforehand.

The second theme of the survey focused on the assessing the anticipated *effect* the Ethics Canvas has the business model and technology design of the ICT application that the students are working on. First the participants were asked whether the exercise would have any impact on their project's technology design, with which 32% of the participants agreed and 16% strongly agreed. A similar question was asked in relation to the impact of the canvas on the business model. 52% of the participants agreed that the Ethics Canvas led them to reconsider their business models and 5% strongly agreed. Finally, the participants were asked whether the exercise was useful in promoting the group's ethical behaviour. 35% of the participants agreed that the exercise promoted ethical behaviour and 40% strongly agreed. Even though these outcomes do not directly indicate that follow-up actions have been taken or will be taken, they at least indicate an intention amongst the students to use the outcomes of the Ethics Canvas exercise to adjust their business models or technology designs.

Overall, the results suggest that it is reasonable to state that the Ethics Canvas is perceived as a useful tool to guide participants in discussing a broad range of ethical impacts as well as the identification of relevant stakeholders. Moreover, the results indicate that it is reasonable to assume that the Ethics Canvas can lead to the inclination based on ethical concerns of participants to reconsider their business models or technology designs. Nevertheless, results also indicate that the structure of the Ethics Canvas will need to be improved to be more inclusive of potential ethical impacts. Moreover, our study is limited due to the limited participation rate (28% of all the students who worked on the Ethics Canvas exercise). This might possible have led to biased results, because the cohort of students that voluntarily filled in the questionnaire could have coincided with the cohort of students that was most positively engaged during the Ethics Canvas exercise. Hence, even though these initial results positively suggest that the Ethics Canvas is a useful tool for practising ethics in R&I, further development of the Ethics Canvas and additional ways of assessing its usefulness will be needed for future studies.

7.3   Evaluating the Novel Approach

At this point, we still need to provide an answer to the third question that guided our work in this dissertation: *To what extent* can insights from virtue ethics and philosophical hermeneutics be synthesised to construct the theoretical foundation for a new method for practising ethics in research and innovation that incorporates the recommendations

advanced? Most of the work towards answering this question has already been done in the course of constructing the novel method. However, we have not yet made the reflective turn back towards the recommendations that motivated our efforts. That is, we did not evaluate the "*extent*" mentioned in our research question, which will follow from considering the limitations of the work we did. In this section, we will therefore evaluate our work in constructing a methodology in accordance with the recommendations that we put forward in chapter 3.

Before we do this, we need to present two reservations regarding this task. First, we need to acknowledge that the aim of our evaluation cannot be the appraisal of the work we ourselves did, but only the providing of guidance for the critical reader to come to his or her own estimations. For this reason, we will only explain the ways in which we responded to the recommendations and subsequently present what we believe are important limitations of these ways of responding – limitations of this dissertation. Second, we need to emphasise that the scope of our evaluation is limited by the scope of the recommendations offered in chapter 3. Recommendations that we failed to formulate at that stage of the investigation will consequently not be covered in this section, which we can therefore not claim to be complete. In what follows, we will restate the recommendations presented in chapter 3 and for each recommendation briefly reflect on the extent to which we believe it has been met and to what extent it calls for further development and refinement; leading to some suggestions for future work.

7.3.1   Uncertainty of technological change

We offered the following two recommendations regarding the aim of dealing with uncertainty of technological change in practising ethics in R&I:

- Methods for practising ethics in R&I that make use of methodological constructs to imagine or foresee possible futures pertaining to the development and use of emerging technologies should more thoroughly engage in an epistemological discussion of the limits of knowledge pertaining to such foresight.
- Whenever future development and use of emerging technologies cannot be meaningfully foreseen, methods for practising ethics in R&I should take appropriate approaches into account that divert from action-guidance based on speculative knowledge about the future such as approaches for the analysis of present promises and expectations concerning emerging technologies and approaches in virtue ethics.

Because our approach does not contain elements of foresight based on prediction of possible futures, we did not respond to the first recommendation. Our approach did, however, address the second recommendation by adopting virtue ethics as the basis for our methodology. Since we already addressed the merits of the virtue ethics tradition regarding the issue of uncertainty in chapter 3, we will not discuss those in detail here. Instead, we will shortly reflect on the way in which our approach, being based on virtue ethics, responds to the challenge of uncertainty.

First, we argued in our approach that instead of focusing on the uncertain technical practices of the use of an emerging technology, we should focus on the much more certain technical practices involved in the making and governing of such a technology. For instance, even though the consequences of the use of certain nanotechnologies that are to be introduced in society are often uncertain, we can already pay attention to the technical practices R&I practitioners engage in to make and govern them. Doing so, we could for instance formulate a particular ethical oath, enhance the education and training of R&I practitioners, and formulate policies regulating the marketing and dissemination of nanotechnologies. In fact, the medical sciences have developed many standards of excellence in their long history that relate to technical practices of making (e.g. the Hippocratic oath, standards for medical trials, etc.), that arguably all in a certain sense respond to the challenge of uncertainty, namely the uncertainty of impacts of medicinal applications on the health of human beings. Instead of trying to "foresee" these impacts, practitioners are expected to respect certain standards of excellence that enables them act in accordance with the virtues. To illustrate this difference, consider for example a scenario in which technical practices in the medical sciences would have standards of excellence similar to those for designers in ICT start-ups in Silicon Valley. Medical appliances would be disseminated according to standards of excellence such as "release early, release often" (Raymond, 2005), which would probably both lead to patients receiving medical solutions they need earlier than would be normally the case, but also to hazardous health risks and possible deaths. Could such an ethically problematic situation really be attributed to the problem of "uncertainty", because one indeed would not have engaged in any foresight of the consequences, or would the lack of attention to the cultivation of the respective technical practices be a better explanation? It seems that the latter would be the case rather than the former, which has led us to take the notion of technical practice. In a way, this implies that our approach has altogether reduced the significance of uncertainty as a factor in practising ethics in R&I.

Second, we responded to the challenge of uncertainty by engaging with narrative theory in constructing our method. Narratives, such as those contained in literary fiction, can confront people with imaginative variations that invoke them to think differently about possible worlds, possible times, and possible technical practices. However, we should not be misled into thinking that such narratives are "about" the future. For instance, even though the narrative of Asimov's *I, Robot* is "set" in the future that fact does not determine the imaginative variations the novel brings about. It could well have been set in an imaginative past, and still lead to the anticipatory insights it is famed for, including its three laws of robotics that crucially regulate human practices. Even though narrative can therefore help us to understand possible worlds, possible times, and possible technical practices (cf. Milojević & Inayatullah, 2015), it does not give us insight into "the future". Instead, it makes explicit how practitioners can cultivate technical practices in accordance with the virtues, because it allows them to *explain* these practices. As such, it is aligned with Ricoeur's two important theses: that (1) "to explain more is to narrate better" (Ricoeur, 1983, p. 171) *and* (2) "to explain more is to understand better" (Ricoeur, 1983, p. x). Integrating narrative in a method for practising ethics in R&I therefore does not enable us to somehow look into the future, but rather to better understand technical practices as they currently are and as they could be.

There are, however, limitations to the ways in which we addressed the recommendation concerning uncertainty. First, we should acknowledge that even though we disregarded foresight methods in our approach, we should not conclude that predictions of future states of affairs are not important. On the contrary, technical practices of prediction are vital for many types of R&I, in the forms of for instance risk analyses, trend analyses, and so forth. R&I practitioners will need to gain insight into the potential risks of what they are doing and making, and need to be able to respond to those risks. However, we should not subsume these technical practices under the notion of practising ethics. Technical practices such as risk analysis are concerned with identifying and explicating potential risks in R&I projects, and *not* with the ethical appraisal of those risks. Nevertheless, while not integrating technical practices such as risk analysis into our method for practising ethics, it would have been prudent to explain the relation between these activities and show how one could benefit the other.

Second, our response has limitations that can be derived from a number of criticisms of virtue ethics that we did not address in this dissertation (cf. Louden, 1984). One important criticism that our approach does not address concerns the distinction between first-person deliberation and third-person evaluation. As Williams argues, the virtues are invoked to

evaluate a person's character and actions (e.g. "she acted courageously"), but rarely or perhaps never play a role in a person's actual deliberations that lead to action (e.g. "I'm going to act courageously in this situation" will rarely be the basis of someone's deliberation) (Williams, 2006, p. 10). As such, thinking about possible states of character in terms of the virtues does not imply that one thinks about one's own actions, but about ways in which others might characterise these actions (Williams, 2006). Accordingly, our approach is limited in that it provides a third person account of the ways in which we might improve technical practices through education, through mentorship, through the formulation of a code of conduct, and so forth, but does *not* provide any guidance for first-person deliberation. For instance, a practitioner using the method will not be able to use it while acting to for instance link the technical practices he engages in with a conception of living together in a political community. The link between the third-person and the first-person perspective within our approach is one that could therefore be paid more attention to.

Third, another potential problem of virtue ethics that is yet to be fully explored is consists in the tension between prudence and technological innovation. We already argued in chapter 4 that the setting in which human action takes place is co-constitutive of its virtuous character, but how should we consider action in a technical setting that is pushed to the limits of human comprehension? In other words, how can R&I practitioners engage in their practices prudently when these practices are aimed at bringing forth something radically open-ended, such as certain forms of genetic manipulation or development of artificial intelligence? As Blok (2019) argues, Aristotle already pointed at the idea that private efforts to innovate has the tendency to negate the stability of a political community, and thereby the virtues it promotes. Certainly, we have to acknowledge that innovation in the modern world is very different from innovation in Ancient Greece, but in a certain sense the danger identified by Aristotle has only become more present, with the growing power of private actors (e.g. Google, Amazon, Facebook) that through innovations have a profound disruptive impact on society that at some point might surpass the ability of human beings to counter through promotion and exercise of prudent activities. Our method is limited insofar as it does not take this tension into account.

7.3.2 Ethical technology design

We offered the following two recommendations regarding the aim of ethical technology design in practising ethics in R&I:

- Approaches dealing with ethical technology design could focus more on the integration of ethics in the day-to-day work of R&I practitioners, especially with regard to the disclosure of ethical issues in design.

- Considerations of methodological aspects of ethical technology design could be based on a normative theoretical framework that explicates how certain technology design choices can be identified as ethical, or how "ethics" is mediated by technology design.

We argue that our work has offered two responses to the first recommendation. First, our approach has broadened its scope compared to other methods for practising ethics in R&I, by not merely focusing on the technical practices internal to the R&I process (e.g. design), but also on those outside of it. By doing so, it has accordingly conceptualised the practitioner not merely as an anonymous doer who happens to write the code for a computer program, design a medical appliance, or use a smartphone to talk with friends, but also as a person who receives a certain type of education, who has to make certain life choices and reflect on those, who can take and ethical oath, and so forth. By thus giving a "face" to the practitioner, our method has become responsive to her day-to-day work or rather perhaps rather to her day-to-day technical practices, as those relate to standards of excellence and life plans. Second, our approach has produced a concrete tool, the Ethics Canvas, that enables ethical reflection in the day-to-day work of R&I practitioners. This tool has been based on a tool that is already widely used in practice, namely the business model canvas (BMC), and our empirical evaluating of its perceived usefulness indicated that it is of value for use in R&I practices. We addressed the second recommendation, concerning technological mediation, by constructing the narrative technologies approach and integrating it in our account of technical practice that cultivates the virtues. Thereby, we responded to our criticism of Vallor, in which we stated that her virtue ethics of technology approach was in need of an account of technological mediation. Taking Ricoeur's narrative theory as our point of departure, we also responded to two central problems of contemporary philosophy of technology, being the inadequate treatment of the role of language and the social.

The limitations of the way our approach addresses ethical technology design derive from two issues. First, limitations arise from the transition made in our method from philosophical concerns to practical reality. That is, we departed from philosophical concerns in constructing the narrative technologies approach, but eventually engaged in an effort of constructing a concrete tool that seemed far removed from philosophical reflection, which

included iterative evaluation exercises and an empirical study to assess people's attitudes. It is questionable to what extent the latter part of our work falls within philosophy, and to what extent it has turned into a form of positive social science. This ambiguity, it seems, is a challenge that both our approach and approaches in applied ethics in general have not yet come to terms with. This indicates that one would need to transition from consideration in philosophy to a *positive* discipline (e.g. social science or medicine) in order to really *apply* ethics in practice. This transition, however, has remained underdeveloped in this dissertation and has largely been taken for granted.

Second, our narrative technologies approach finds a limitation in not being responsive in its current form to philosophical critiques of narrative. For instance, Strawson has heavily criticised narrative as an ontological and as a normative concept (Strawson, 2004). He argues that both the claims that our understanding and experience of the world are mediated through narrative and that we ought to live our lives according to narratives, are misguided. We have not engaged with such critiques in our approach, and have largely relied on MacIntyre's and Ricoeur's justifications for the role of narrative, which do not take contemporary criticisms into account. Some scholars have responded to the criticisms raised by Strawson and others (cf. Roth, 2017), in defence of narrative, and incorporating those might lead to an improved framework for the account of narrative technologies.

7.3.3   Identifying, analysing, and resolving ethical impacts

We offered the following two recommendations regarding the aim of identifying, analysing, and resolving ethical impacts in practising ethics in R&I:

- Researchers and assessors should use a convincing methodological solution for the problem of value conflicts, when they occur. This could be done by including procedures for reasoned balancing of ethics principles whenever no fixed and justified ranking of principles is provided.
- Methods that analyse ethical impacts of technologies should offer procedural guidance that would allow for using the analysis to choose between certain sociotechnical alternatives.

Regarding the first recommendation, our approach has essentially re-drawn the dilemma from a conflict between values to a conflict between goods and obligations. The conflicts drawn from Ricoeur's "little ethics" (1992) do not take "value" as their starting point, but instead

reflect on the unavoidable struggle between the necessity to bring about a normative system based on the notion of obligation and the application of such a system to particular situations which can invoke a challenge to the notion of the good life. In this light, for instance, the famous conflict between the "values" of privacy and security can be reformulated. Security is not simply to be understood as "being free from interference", but as being protected, which finds its origin in solicitude, understood as care for one-another. Social contract theorists have wrestled with this issue for centuries, to justify positing protection as the primary good provided by the state. Privacy puts necessary limitations on technical practices of protection, in order to respond to the problem of the possibility of evil arising from domination implied in the imbalance between agency (the protector, such as the state) and patiency (the protected, such as consumers of ICTs). Conflicts between privacy and security, then, are to be understood as conflicts between the obligations arising from a conception of privacy and the notion of the good life in a particular situation in accordance with security. Accordingly, our response to the first recommendation is to be found in the responses to the conflicts between goods and obligations: in (1) the reflective equilibrium between the ethics of argumentation and considered convictions, (2) the involvement of experts in prudent decision-making, and (3) the necessity of democratic ways for conflict resolution.

However, the ways in which we addressed conflicts between goods and obligations are limited by the complexity of social practices people engage in to solve conflicts in concrete situations. We can illustrate this limitation by considering the third way of solving conflicts: of establishing democratic ways to arbitrate between claims coming from different spheres of justice and from the governing vis-à-vis the governed. On the one hand, our approach could lead to further philosophical inquiry into democratic theory that would aim to establish certain principles of democratic governance. On the other hand, however, democracy is not achieved through theory but through concrete historical struggles and negotiations between people with different backgrounds, interests and points of view. When "democratic governance" is to be implemented in a certain context (e.g. for arbitration between corporate and state interests regarding ICT security), democratic theory is limited and a lot of the work to be done will depend on what Deuze calls the "bricolage", the incorporation of "bits and pieces" (Deuze, 2006, p. 70) in concrete political reality. This limitation thereby implies a consideration of the role of "procedure" that can be derived from democratic theory in the response to conflicts between goods and obligations. Such a consideration could be accommodated by a reflection on the actualisation of normative theory by anthropologists or ethnologists.

Regarding the second recommendation, our explicit responses to the issue of sociotechnical alternatives have been limited. One response can be found in our explication of respect for persons. We mentioned that when respect for persons is at stake, sociotechnical arrangements in which there is a symbiosis between man and machine are to be preferred over full automation in order to assure reciprocity between agency and patiency. Additionally, the plea for democratic means for arbitration between spheres of justice is a response to the issue of choice between sociotechnical alternatives, because it explicates ways in which societies are to choose as we illustrated in chapter 6 with the example of the attempt by the FBI to gain access to personal data on an iPhone. In other words, our response has been that certain norms for choosing between sociotechnical alternatives can be agreed upon, but that we should simultaneously attend to the democratic ways in which decisions on sociotechnical alternatives are made. However, what is as yet missing in our approach is a translation of our responses into *procedural* guidance, as was suggested by the recommendation. This shows the importance for methods for practising ethics in R&I to engage with legal theory or jurisprudence, to formulate legal procedures that could accommodate the respect for persons, and with political theory, to formulate procedures that would guide the process of arbitration.

7.3.4 Appropriate participation of stakeholders

We offered the following two recommendations regarding the aim of appropriate participation of stakeholders in practising ethics in R&I:

- For methods that deal with stakeholder identification, we recommend that they should include considerations of justified stakeholder selection. These considerations could be based on justified criteria or a mapping-framework for stakeholder identification, or could gain from collaborative approaches or approaches guided by democratic principles.
- For methods that facilitate stakeholder participation in the process of practising ethics in R&I, we recommend that they should include considerations that negate a top-down approach. They could do so by shaping the framework for ethical analysis according to a participatory process or by integrating insights from participatory design in process of practising ethics in R&I.

In our method, we paid heed to the requirements for stakeholder participation that we formulated in section 3.3.3. With regard to the governance of R&I practices, (1) the aspect of structured dialogue enables societal actors to work together, while acknowledging the existence of different power-positions, (2) the aspect of expert guidance allows for the public debate to be opened up without losing the important distinction between stakeholders and lay people, and (3) the aspect of democratic decision-making guarantees democratic governance that does not only provide room for consensus building but also for contestation. Furthermore, with regard to the principles of participatory design, (a) the aspect of community building enhances the capacity of actors to regard each other as equals; (b) the aspect of mentorship promotes the capacity of R&I practitioners to connect their practices with the common good; and (c) the aspect of civic education enables R&I practitioners to have different perspectives regarding the technological innovation they work on. The other three principles of participatory design, that (d) our method needs to be useable in everyday settings, (e) should lead to mutual learning, and (f) should be oriented not towards theory but towards practice, are addressed by the development of the Ethics Canvas.

Notwithstanding these aspects of our method that respond to the challenge of stakeholder participation, our approach is limited in the ways in which it responds to the two recommendations above. First, we have neither discussed nor provided proper guidance for the justification of stakeholder selection, which therefore largely remains an open challenge. A legitimate question to ask is whether a philosophical method can provide proper guidance for the justification of stakeholder selection at all. For instance, we could consider whether stakeholder selection could best be based on a philosophical argument, stating who should be included and who should be excluded based on what reasons, or whether a much better practice would be one in which stakeholders could engage in self-selection (e.g. applying *as* a stakeholder in a particular R&I project), or one in which there is no selection of stakeholders but rather an election of stakeholders or other democratic process resulting in stakeholder involvement (e.g. a citizen's council). Regardless of the potential answers to these questions, they have not been addressed in this dissertation and call for future work to be done that would compare different ways of stakeholder selection and argue what best practices might be put forward.

Second, we did to a certain extent respond to the second recommendation, by considering a practical way to achieve a bottom up rather than a top down manner of practising ethics in R&I, namely in developing the Ethics Canvas. Contrary to common practices in R&I settings such as applying for ethics clearance in accordance with regulations

that are enforced in a top down manner, from the level of the ethics committee on the individual researcher, the Ethics Canvas allows researchers to have agency in considering particular ethical impacts that should be considered. We therefore showed that practising ethics in R&I in a bottom up manner is possible and our empirical research suggested that practitioners also consider it as valuable. However, a lot of work is yet to be done in order to meet the recommendation we ourselves put forward. Our work is limited in that we only constructed a single tool for one of the nine stages of our method, which implies that more tools would be needed to realise participatory practices for each stage. This would require a considerable amount of work, and would require collaboration between philosophers, R&I practitioners and other stakeholders. Furthermore, our work has been limited with regard to our method for evaluating the usefulness of the Ethics Canvas. Our population sample (students working on R&I projects) was adequate, but different other population samples should be considered (e.g. professional R&I teams) in order to strengthen the findings. Also, our means of assessment, a questionnaire, constituted a limitation because it gave us merely an on the spot assessment and no insight into the long-term effects of the use of the Ethics Canvas.

7.4 Chapter Summary

In this chapter, we developed our novel method for practising ethics in R&I comprising nine procedural steps and we developed a specific tool, the Ethics Canvas, which allows one of the procedural steps, namely the gathering of narratives, to be put into practice. The method consists of three phases, the first focusing on mapping technical practices, gathering narratives, and analysing them according to the narrative technologies framework, the second focusing on the interpretations of the technical practices according to their related standards of excellence, life plans, and narrative unity of life, and the third focusing on prescription of aspects of technical practice according to the good life, with and for others, in just institutions. The practical tool that we developed, the Ethics Canvas, resulted from a consultation of business model development literature. We adopted the BMC, which is a collaborative brainstorming tool for creating a business model, and transformed it into the Ethics Canvas, which enables practitioners to discuss the ethical impacts of relevant technical practices. Eventually, we evaluated the extent to which our approach has responded to the recommendations in chapter 3. In the next chapter (chapter 8), we will provide an overview of the work done in this dissertation; discuss some of its limitations and present potential avenues for future research.

7.5 References

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election. *Proceedings of the 3rd International Workshop on Link Discovery - LinkKDD '05*, 36–43. http://doi.org/10.1145/1134271.1134277

Akrich, M. (1992). The De-scription of Technical Objects. In W. Bijker & J. Law (Eds.), *Shaping technology/ Building society*. Camebridge, Massachusetts: MIT Press.

Arendt, H. (1964). *Eichmann in Jerusalem: a Report on the Banality of Evil*. New York: Viking Press. http://doi.org/10.1525/california/9780520220560.001.0001

Aristotle. (1999). *Nicomachean Ethics*. (T. Irwin, Ed.). Indianapolis: Hackett Publishing Company.

Asif, M., & Krogstie, J. (2012). Taxonomy of Personalization in Mobile Services. *Proceedings of the 10th IADIS International ...*, (January), 343–350. Retrieved from http://www.researchgate.net/publication/258847751_Taxonomy_of_Personalization_in_Mobile_Services/file/60b7d5293d9f45c89a.pdf

Blok, V. (2013). The Power of Speech Acts: Reflections on a Performative Concept of Ethical Oaths in Economics and Business. *Review of Social Economy*, *71*(2), 187–208. http://doi.org/10.1080/00346764.2013.799965

Blok, V. (2019). Towards an Ontology of Innovation: On the New, the Political-Economic Dimension and the Intrinsic Risks Involved in Innovation Processes. In *Routledge Handbook of Philosophy of Engineering*, London: Routledge.

Bloor, M., Frankland, J., Thomas, M., & Robson, K. (2001). *Focus Groups in Social Research*. London: SAGE Publications. http://doi.org/10.4135/9781849209175

Bocken, N., Short, S., Rana, P., & Evans, S. (2013). A value mapping tool for sustainable business modelling. *Corporate Governance: The International Journal of Effective Board Performance*, *13*(5), 482–497. Retrieved from 10.1108/CG-06-2013-0078%5Cnhttp://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=90610939&site=ehost-live

Bohn, J., Coroama, V., Langheinrich, M., & Mattern, M. (2005). Social, economic, and ethical implications of ambient intelligence and ubiquitous computing. *Ambient Intelligence*, *10*(5), 5–29. http://doi.org/10.1007/3-540-27139-2_2

Bouzeghoub, M., & Kostadinov, D. (2007). *Data Personalization : a Taxonomy of User Profiles Knowledge and a Profile Management Tool. Rapports de recherche du laboratoire PRiSM,[en ligne]*.

Brey, P. (2000). Disclosive Computer Ethics: The Exposure and Evaluation of Embedded Normativity in Computer Technology. *Computers and Society*, *30*(4), 10–16.

Clavell, G. G. (2017). Protect rights at automated borders. *Nature*, *543*(7643), 34–36. http://doi.org/10.1038/543034a

Deuze, M. (2006). Participation, remediation, bricolage: Considering principal components of a digital culture. *Information Society*, *22*(2), 63–75. http://doi.org/10.1080/01972240600567170

De Vries, K. (2010). Identity, profiling algorithms and a world of ambient intelligence. *Ethics and Information Technology*, *12*(1), 71–85. http://doi.org/10.1007/s10676-009-9215-9

Ezra, M. (2007). The Eichmann Polemics : Hannah Arendt and Her Critics. *Democratiya*, *9*(Summer 2007), 141–165.

Feenberg, A. (1999). *Questioning Technology*. New York: Routledge.

Friedman, B., Kahn Jr., P. H., & Borning, A. (2006). Value Sensitive Design and Information Systems. In K. E. Himma & H. T. Tavani (Eds.), *Human-Computer Interaction and Management Information Systems: Foundations* (pp. 1–27). John Wiley & Sons, Inc. http://doi.org/10.1145/242485.242493

Friedman, B., & Kahn, P. (2002). Value sensitive design: Theory and methods. *University of Washington Technical*, (December), 1–8. http://doi.org/10.1016/j.neuropharm.2007.08.009

Gadamer, H.-G. (2013). *Truth and Method*. (J. Weinsheimer & D. G. Marshall, Eds.). London: Bloomsbury.

Gambino, L., & Siddiqui, S. (2017). Thousands protest against Trump travel ban in cities and airports nationwide. Retrieved May 13, 2018, from https://www.theguardian.com/us-news/2017/jan/29/protest-trump-travel-ban-muslims-airports

Gutmann, A. (1995). Civic Education and Social Diversity. *Ethics*, *105*(3), 557–579.

Hargittai, E. (2007). Whose space? differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, *13*(1), 276–297. http://doi.org/10.1111/j.1083-6101.2007.00396.x

Heidegger, M. (1977). *The Question Concerning Technology*. New York: Garland Publishing.

Henriksen, K., Bjerre, M., Bisgaard, T., Almasi, A. M., & Damgaard-Grann, E. (2012). Green Business Model Innovation : Empirical and Literature Studies. *Nordic Innovation Report*, (October). Retrieved from www.nordicinnovation.org/Publications/green-business-model-innovation-empirical-and-literature-studies/

Ihde, D. (1979). *Technics and Praxis* (Vol. 33). Dordrecht: D. Reidel Publishing Company.

Ihde, D. (1990). *Technology and the Lifeworld. The Indiana Series in the Philosophie of Technologie* (Vol. 1). http://doi.org/10.1049/et:20060114

Jovchelovitch, S., & Bauer, M. W. (2000). Narrative Interviewing. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image and sound: a practical handbook*. London: SAGE Publications.

Juels, a, Molnar, a, & Wagner, D. (2005). Security and Privacy Issues in ePassports. *Security and Privacy for Emerging Areas in Communications Networks, 2005.*, 74–88. http://doi.org/10.1109/SECURECOMM.2005.59

Kant, I. (2002). *Groundwork for the Metaphysics of Morals*. (J. B. Schneewind, M. Baron, S. Kagan, & A. W. Wood, Eds.) (Vol. 47). New Haven: Yale University Press.

Kaplan, D. M. (2003). *Ricoeur's Critical Theory. SUNY series in the philosophy of the social sciences*. New York: SUNY Press. http://doi.org/10.1017/CBO9781107415324.004

Kuparinen, P. (2012). Business Model Renewal and its Networking Aspects in a Telecom Service Company. *Master Thesis*.

Latour, B. (2013). *An Inquiry into the Modes of Existence: An Anthropology of the Moderns*. (C. Porter, Ed.). Cambridge, Massachusetts: Harvard University Press.

Lichterman, P. (1995). Piecing Together Multicultural Community: Cultural Differences in Community Building among Grass-Roots Environmentalists. *Social Problems*, *42*(4), 513–534. http://doi.org/10.1525/sp.1995.42.4.03x0130z

Liebowitz, S. J., & de Meuse, K. P. (1982). The Application of Team Building. *Human Relations*, *35*(1), 1–18. http://doi.org/10.1177/001872678203500102

Louden, R. B. (1984). On Some Vices of Virtue Ethics. *American Philosophical Quarterly*, *21*(3), 227–236.

http://doi.org/Journal Article

Lucassen, G., Brinkkemper, S., Jansen, S., & Handoyo, E. (2012). Comparison of Visual Business Modeling Techniques for Software Companies. In M. A. Cusumano, B. Iyer, & N. Venkatraman (Eds.), *Software Business: Third International Conference, ICSOB 2012, Cambridge, MA, USA, June 18-20, 2012. Proceedings* (pp. 79–93). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-30746-1_7

Lycett, M. (2013). "Datafication": making sense of (Big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386. Retrieved from http://v-scheiner.brunel.ac.uk/handle/2438/8110

Miller, J. I. (2005). Don't be Evil: Gmail's Relevant Text Advertisements Violate Google's Own Motto and Your E-Mail Privacy Rights. *Hofstra Law Review*, *33*.

Milojević, I., & Inayatullah, S. (2015). Narrative foresight. *Futures*, *73*, 151–162. http://doi.org/10.1016/j.futures.2015.08.007

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. BusinessModelgeneration.com.

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artifacts: or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, *14*(3), 221–232. http://doi.org/10.1177/030631284014003004

Raymond, E. (2005). The cathedral and the bazaar. *First Monday*, *2*(SPEC), 23–49. http://doi.org/10.1007/s12130-999-1026-0

Ricoeur, P. (1992). *Oneself as Another*. (K. Blamey, Ed.). Chicago: University of Chicago Press.

Roosendaal, A. (2014). Digital Personae and Profiles as Representations of Individuals. In M. Bezzi, P. Duquenoy, S. Fischer-Hubner, M. Hansen, & G. Zhang (Eds.), *Privacy and Identity Management for Life* (pp. 226–236). Springer.

Roth, B. (2017). Reading from the middle : Heidegger and the narrative self. *European Journal of Philosophy*, (October), 1–17. http://doi.org/10.1111/ejop.12314

Schot, J., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change*, *54*(1996), 251–268. http://doi.org/10.1016/S0040-1625(96)00180-1

Schubert, C. (2015). *Christian Schubert On the ethics of public nudging : Autonomy and Agency On the ethics of public nudging : Autonomy and Agency* (No. Joint Discussion Paper Series in Economics No. 33-2015). University of Kassel.

Strawson, G. (2004). Against Narrativity. *Ratio*, *17*(4).

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.

Verbeek, P.-P. (2005). *What things do; philosophical reflections on technology, agency, and design*. Pennsylvania: Pennsylvania University Press.

Weigel, M. (2017). Coders of the world, unite: can Silicon Valley workers curb the power of Big Tech?

Zott, C., Amit, R., & Massa, L. (2011). The business model: Recent developments and future research. *Journal of Management*, *37*(4), 1019–1042. http://doi.org/10.1177/0149206311406265

8 Conclusion

8.1 Introduction

This eighth and final chapter presents the core findings of the work done in this dissertation. Subsequently, it discusses this dissertation's main achievements and some of its main limitations. Finally, it presents some avenues for future work.

8.2 Summary

The first research question and point of departure of this dissertation has been: what methods have been developed for practising ethics in R&I? The aim of this question was to gain a comprehensive overview and create an analysis of the state of the art of academic literature on this topic. To answer it, we engaged in a systematic literature review, which brought us over a hundred useful sources that discuss methods for practising ethics in R&I. We observed that these were often field-specific, and that in particular the fields of health technologies, information systems and computer science attempted to incorporate ways of practising ethics in R&I. In order to make sense of the great variety of methods developed for different fields of R&I, we categorised them according to whether they aimed (1) to anticipate impacts of emerging technologies (ex ante), (2) to practise ethics in the design process (intra) or to ethically evaluate technologies after they had been developed and introduced in society (ex post). For each of these categories, we identified procedural steps that were prevalent or shared between methods. For instance, for ex ante methods we identified four procedural steps: (1) identify potential emerging technologies, (2) construct scenarios about future impacts, (3) evaluate potential ethical impacts and (4) assess the status of uncertain normative claims.

We subsequently engaged in answering the second research question: what are the main shortcomings of these methods and what recommendations for a novel method follow from these shortcomings? To answer this question, we first engaged in a critique of the existing methods, which focused on questioning their over-arching aims. Ex ante methods, aiming at dealing with uncertainty of technological change, face the problems of unreliability of speculations about future states of affairs and unjustifiability of the assumption that sufficient knowledge about the future for action guidance can be gained. Intra methods, aiming at ethical technology design, face the problems of not being responsive to the day-to-day work of R&I practitioners and of not adequately showing how ethical considerations can

be incorporated in design. Ex post methods, aiming at identifying, analysing and resolving ethical impacts, face the problems of inadequately dealing with value conflicts and of providing inadequate guidance for choosing between sociotechnical alternatives. Furthermore, we identified a common challenge for all methods, namely of stakeholder participation, which invoked the problems of justifying selections of relevant stakeholders, and of offering overtly top-down approaches. Several recommendations were presented for each of the critiques of the different types of methods. Based on these, a framework for a novel method was constructed, which stipulated and defended that (1) virtue ethics would provide a fruitful basis for our method, (2) that a theory of technological mediation based on the notion of narrative was needed, and (3) that the principles of participatory design and more specifically the business model canvas (BMC) should be taken as points of departure for the organisation of participation in our novel method.

In line with the answers provided to the second research question, we turned to the third research question, which encompassed the effort to construct a novel method and provide it with a methodological basis: to what extent can insights from virtue ethics and philosophical hermeneutics be synthesised to construct the theoretical foundation for a new method for practising ethics in research and innovation that incorporates the recommendations advanced? We started working on this question by investigating virtue ethics, and its contribution to ethics of technology. We provided a short overview of the core questions that the virtue ethics tradition aims to answer, and discussed MacIntyre's (2007) work on virtue ethics and its conception of the cultivation of practices that cultivate the virtues in relation to life plans and moral traditions. We also invoked Vallor's (2016) recent work on virtue ethics of technology, in particular because it provided us with a heuristic of virtues that most convincingly fits our contemporary technological age. However, we also criticised Vallor's work and showed that virtue ethics of technology cannot do without an account of technological mediation. Additionally, we criticised MacIntyre's notion of "practice" for being both opaque and idealised, putting forward the initial reasons for preferring Ricoeur's (1992) notion of practice as the one on which we would eventually base our novel method.

In order to complement our virtue ethics approach with an account of technological mediation, we turned to Ricoeur's narrative theory to construct the narrative technologies approach. To start, we provided a short overview of Ricoeur's work and overall philosophy, which showed Ricoeur's particular preoccupation with the human understanding of the self as being indirect, mediated by language and by social relations in public life. We then took a

deep dive into his narrative theory developed in *Time and Narrative* (Ricoeur, 1983, 1985, 1988). We first explicated Ricoeur's model of emplotment, which denoted a movement from prefigured time, through configured time, to refigured time. In order to justify the use of Ricoeur's model, we argued for the continuity between textual mediation on the one hand and technological mediation on the other, showing how technologies can configure characters and events in a meaningful whole in a similar way as texts do. We then formulated a model for technological mediation that is based on four central concepts, derived from a re-interpretation of Ricoeur's account of emplotment: 1) Textuality refers to the extent to which a technology is similar to the paradigm of the text by being capable of bringing about an active process of emplotment. 2) Literacy refers to the extent to which different people are attuned to engage with the process of technological emplotment. 3) Temporality refers to the dimension of time configured by a technology, which can be chronological or non-chronological. 4) Distancing refers to the extent to which technological configuration would bring about imaginative variations or abstraction from the world of action through strict representation.

Having outlined both the virtue ethics of technology approach and our philosophical account of technological mediation, we had to find a way to bring them together. We did so by using the notion of "technical practice". We first engaged in a discussion of Heidegger's work on technology (1996, 2009), which brought us a determination of technical practice that cultivates the virtues as acting awake and succeeding for the sake of living together in a political community. We then argued that we needed a complementary account of authenticity to distinguish between "good" and "bad" practices, and that instead of Heidegger's notion of authenticity we would need one grounded in narrative, which we found in Ricoeur. In order to arrive at an account of technical practice that cultivates the virtues, we first outlined Ricoeur's (1992) theory of practice, which integrated an account of human activities (basic actions, practices) with an account of the structures by means of which we make sense of those activities (constitutive rules, life plans, narrative unity of life). To finalise our account, we linked technical practice with the ethical aim, understood as the good life, with and for others, in just institutions. For each stage of the ethical aim, we made initial suggestions as to what it could contribute to the cultivation of technical practices in accordance with the virtues.

Together with a number of other considerations developed throughout this dissertation, the findings in chapter 6 were taken up to construct a method for practising ethics in R&I. The method we constructed consists of three phases, the first focusing on the

gathering and interpretation narratives, the second focusing on the interpretation and evaluation of technical practices, and the third focusing on the evaluation of these practices according to the ethical aim, leading to different forms of prescription. The three phases contain nine stages, each of which contains specific procedural steps. A particular stage is for instance the consideration of a technical practice regarding the good life, containing procedural steps that aim to prescribe forms of mentorship, of an ethical oath, and of structured dialogue. We argued that the method as such did not yet contain concrete resources for R&I practitioners to integrate ethics into their day-to-day work, and that this would need a translation from procedural steps into practical tools. To show that this can be done, we presented an effort to construct a tool for "gathering narratives" that is based on the collaborative brainstorm tool for business modelling called the business model canvas. By considering relevant ethical impacts of technical practices that we derived from literature in STS and philosophy of technology, we designed the Ethics Canvas, which was subsequently refined and redesigned on the basis of an iterative evaluation and design process. To estimate whether the Ethics Canvas could actually work in a practical setting, we assessed its perceived usefulness amongst groups of students that used it as an exercise in their course work. This assessment suggested that the Ethics Canvas could be a valuable tool for practising ethics in R&I.

8.3   Achievements and Limitations

To conclude, we first shortly reflect on what has been achieved in this dissertation before moving on to the limitations of our work. The main achievement of this dissertation lies in a contribution to philosophy, and consists in bringing together and re-interpreting philosophical perspectives that have until now been disconnected. First, we have reinterpreted Ricoeur as a philosopher of technology and formulated an approach to practise ethics in R&I inspired by his philosophical hermeneutics. Second, we have brought theories of technological mediation in dialogue with the tradition of virtue ethics. Third, we have synthesised Ricoeur's contribution to our understanding of technology with his contribution to virtue ethics. We expect that this effort will open up a new strand of philosophical inquiry that uses philosophical hermeneutics not only as a way to understand technological mediation, but also as a point of departure for ethics of technology. Additionally, this dissertation made two other, more minor achievements. First, it managed to construct a comprehensive overview and analysis of the myriad of methods for practising ethics in R&I. Second, it presented one of the first efforts to integrate a rigorous philosophical approach with a highly pragmatic

interdisciplinary approach to produce a tool that can be used to practise ethics in R&I settings.

Notwithstanding the merits of this dissertation, it also faces some significant limitations. Detailed limitations were already discussed in chapter 7, and here we will offer a number of general limitations that are linked to the main achievements. Regarding our philosophical contribution, two general limitations need to be mentioned. First, our relatively narrow focus on Ricoeur's work has produced a number of blind spots regarding other theorists of narrative, notably those engaging in STS. On the one hand, we have been blind for some philosophical perspectives, most notably of Latour's (1994), whose material semiotics also explicitly departs from narrative theory, namely from the one developed by Greimas. It remains unclear in our work how the appropriation of narrative theory by Latour relates to Ricoeur's. This limitation could be overcome by constructing a critique of Latour's Actor Network Theory by way of Ricoeur's explicit engagement with Greimas. Other notable philosophical perspectives on narrative to which this dissertation has been blind are Gallagher's (2011), whose recent work on "narrative competency" has been highly influential, and Kearney's (1996) work on narrative and ethics. On the other hand, our work has been limited in its engagement with theorists of narrative outside of philosophy. A notable scholar in this regard is Czarniawsma (1998), whose work on narrative has been highly influential in organisation studies and engages with method much more explicitly than Ricoeur has done. A second general limitation of our philosophical contribution pertains to our limited engagement with Ricoeur's work. As we indicated in chapter 6, we refrained from engaging with Ricoeur's (1992) speculative work on developing an ontology in his tenth study of *Oneself as Another*. Our approach therefore lacks the ontological grounding that Ricoeur argued was necessary, and is inherently limited thereby.

Regarding our contribution to the description and analysis of existing methods, our work is limited due to decisions we made regarding scope and limitations of the literature review. We chose to limit our review to methods that explicitly deal with ethics, which made sense in terms of our limited time and resources but precluded us from gaining relevant insights from methods that do not explicitly deal with ethics but are closely related and might have presented us with superior methodological frameworks. We should in particular think of methods in technology assessment (TA) or environmental impacts assessment, which not infrequently deal with ethical issues. Methods belonging to these fields know a longer history than methods for practising ethics in R&I, having been developed since the 1970s (Schot & Rip, 1997; Suter, 2008), and have much more often been translated into practical policies,

242

tools, and R&I settings. It might therefore have been the case that some of the shortcomings we encountered in methods for practising ethics in R&I could have been remedied by means of methodological considerations coming from the fields of TA and environmental impact assessment. Furthermore, we should think of methods in business and management studies that deal with the issue of stakeholder participation and stakeholder theory. Methods belonging to this field are often clearly linked to empirical practices, for instance by studying the formation of public-private partnerships in which stakeholder participation is shaped (cf. Blok et al., 2015). They also seem to provide a more comprehensive overview of the participation mechanisms that are applied in the governance of R&I practices and the benefits and drawbacks that these have (Joss & Bellucci, 2002; Gould, 2012).

Concerning our contribution to the practical application of ethics in R&I settings, which culminated in the Ethics Canvas, our work is limited in three ways. First, the Ethics Canvas is not yet a proper translation of the method we developed because of the discrepancy between the design process of the Canvas in a multidisciplinary setting and the development of the philosophical approach. Design-related decisions for the Ethics Canvas had to be made at a point at which the philosophical approach was still incomplete, which caused its heuristic to be based on existing and not necessarily related theories. Second, the evaluation exercise we did for the Ethics Canvas has been far from sufficient for establishing it as a proper tool for practising ethics in R&I. Evaluation efforts will need to be expanded to a greater number of stakeholders, including professional R&I teams, user-groups of technologies, and so forth, to provide a conclusive picture of the feasibility of the Ethics Canvas. Third, our research has not managed to make clear how the Ethics Canvas would fit and be interoperable with a number of other tools for practising ethics in R&I. At the moment it is merely a single tool, used for a single particular stage in our method.

8.4 Future Research

Based on the abovementioned general limitations and the more detailed limitations presented in chapter 7, we can provide some potential avenues for future research. Regarding the philosophical work done in this dissertation three avenues seem most relevant. The most important one seems to be research that will embed the philosophy and ethics of technology narrowly based on Ricoeur's work within a larger academic discussion on the role of narrative. Such research could for instance juxtapose Ricoeur's narrative theory with those developed by Latour and Czarniawsma, and integrate insights from such encounters in a more refined philosophical approach. A second strand of research could explicitly engage

with the critics of narrative, and engage in a defence of the relevance of narrative theory for our thinking about technology. Such research could engage with the critical work of scholars such as Strawson, and also acknowledge the potential limitations of any theory of narrative. As such, the narrative technologies approach could aim to incorporate both contemporary criticisms of narrative as an ontological and normative concept, and rebuttals of those criticisms. Third, a more speculative avenue for future research would lie in a dialogue between the approach developed in this dissertation and political theory. Some of the prescriptive aspects of our method explicitly engage with notions from political theory, such as democratic decision-making. An engagement with political theory would provide R&I practitioners with more insight about ways in which they could organise such aspects of practising ethics in R&I.

Concerning the analysis of existing methods, we can provide two potential avenues for future research. The first would focus on creating a more unified account of all (or at least, significantly more) types of methods that have been developed to assess technologies in some shape or form. Assessment should be understood more broadly in this case, as not only pertaining to ethical assessment, but assessment of "social impacts", "environmental impacts", and so forth. A more unified perspective would enable researchers to draw parallels between different methods (e.g. between methods for practising ethics and methods in environmental impacts assessment) and to show general gaps or inadequacies that would otherwise remain undiscovered because of the fragmentation between the different fields. Also, a unified perspective would strengthen conceptions of concerns that cut across disciplines, such as stakeholder participation. In this regard, it could more adequately discuss the limitations of stakeholder participation in certain contexts (e.g. regarding public or private R&I processes) and prescribe mechanisms for shaping it institutionally (e.g. through public-private partnerships). A second strand of research could be concerned with developing a more philosophical critique of the idea to "apply ethics" to R&I settings. "Ethics" is witnessing an unprecedented rise in popularity at the current moment (anno 2018), being incorporated in numerous R&I funding schemes (cf. European Commission, 2013) and governmental strategies (cf. Shelley-egan & Rodrigues, 2015). Hence, there is a significant institutional effort to apply ethics to R&I. However, there is a general lack of reflection on the extent to which this effort is effective and whether it is adequately designated as ethics or should better be understood as "governance", "policy making", and so forth. In other words, a philosophical critique of applied ethics could clarify the transition from philosophical

debate to positive science, from normative arguments to bureaucratic tools, and thereby criticise and redefine the limits of applied ethics.

Finally, future research could focus on the development and testing of tools for practising ethics in R&I. We can provide two different avenues for research in this regard. First, a widespread effort to implement and evaluate tools such as the Ethics Canvas could be developed. Because of the variety of different stakeholders for which such tools would be relevant, such an effort should not exclusively be an academic one. Instead, it should engage stakeholders from industry, civil society, and government, in a similar way as Osterwalder and Pigneur (2010) did in developing the BMC. They tested the BMC with stakeholders around the world, in different settings (online, and offline), and engaging more than 400 experts. It seems that research in applied ethics aiming to translate philosophical approaches into practical tools cannot but follow such an way of testing and implementing. Second, research could focus not on the implementation of tools for practising ethics in R&I, but in the actual practices they mediate in R&I labs, offices, and so forth. Such a type of research would be of a anthropological and ethnographic character and would ask how R&I practitioners really "practise" ethics in their day-to-day activities, once they have received the means, procedures, regulations, brainstorming tools, and so forth, to achieve this.

8.5 References

Blok, V., Hoffmans, L. and Wubben, E. F. M. (2015). Stakeholder engagement for responsible innovation in the private sector: critical issues and management practices, *Journal on Chain and Network Science*, *15*(2), 147–164. doi: 10.3920/JCNS2015.x003.

Czarniawsma, B. (1998). *A Narrative Approach to Organization Studies*. London: SAGE Publications.

European Commission. (2013). *Ethics for researchers*. Luxembourg: Publications Office of the European Union. Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/ethics-for-researchers_en.pdf

Gallagher, S. (2011). Narrative competency and the massive hermeneutical background. In Fairfield, P. *Hermeneutics in Education*, New York: Continuum.

Gould, R. W. (2012). Open innovation and stakeholder engagement, *Journal of Technology Management and Innovation*, *7*(3), 1–11.

Heidegger, M. (1996). *Being and Time*. (J. Stambaugh, Ed.). New York: State University of New York Press. http://doi.org/10.1017/CBO9781107415324.004

Heidegger, M. (2009). *Basic Concepts of Aristotelian Philosophy*. Bloomington: Indiana University Press.

Joss, S. and Bellucci, S. (2002). *Participatory Technology Assessment: European Perspectives*. London: University of Westminster.

Kearney, R. (1996). Narrative and Ethics. *Aristotelian Society* (Vol. 70, pp. 29–61). http://doi.org/10.2307/4107002

Latour, B. (1994). On Technical Mediation - Philosophy, Sociology, Genealogy. *Common Knowledge*, 3(2), 29–64. http://doi.org/10.1111/j.1365-294X.2010.04747.x

Osterwalder, A., & Pigneur, Y. (2010). Business model generation: a handbook for visionaries, game changers, and challengers. Retrieved from BusinessModelgeneration.com.

MacIntyre, A. (2007). *After Virtue: A study in moral theory* (Third Edit). Notre Dame, Indiana: University of Notre Dame Press. http://doi.org/10.1017/CBO9781107415324.004

Ricoeur, P. (1983). *Time and Narrative - volume 1*. (K. McLaughlin & D. Pellauer, Eds.) (Vol. 91). Chicago: The University of Chicago. http://doi.org/10.2307/1864383

Ricoeur, P. (1985). *Time and Narrative - volume 2*. (K. McLaughlin & D. Pellauer, Eds.). Chicago: The University of Chicago.

Ricoeur, P. (1988). *Time and Narrative - volume 3*. Chicago: The University of Chicago.

Ricoeur, P. (1992). *Oneself as Another*. (K. Blamey, Ed.). Chicago: University of Chicago Press.

Schot, J., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change*, 54(2–3), 251–268. http://doi.org/10.1016/S0040-1625(96)00180-1

Shelley-egan, A. C., & Rodrigues, R. (2015). Ethics Assessment and Guidance at the European Union Level, (June), 1–25. SATORI Project.

Suter, G. W. (2008). Ecological risk assessment in the united states environmental protection agency: A historical overview. *Integrated Environmental Assessment and Management*, 4(3), 285–289. http://doi.org/10.1897/IEAM_2007-062.1

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.