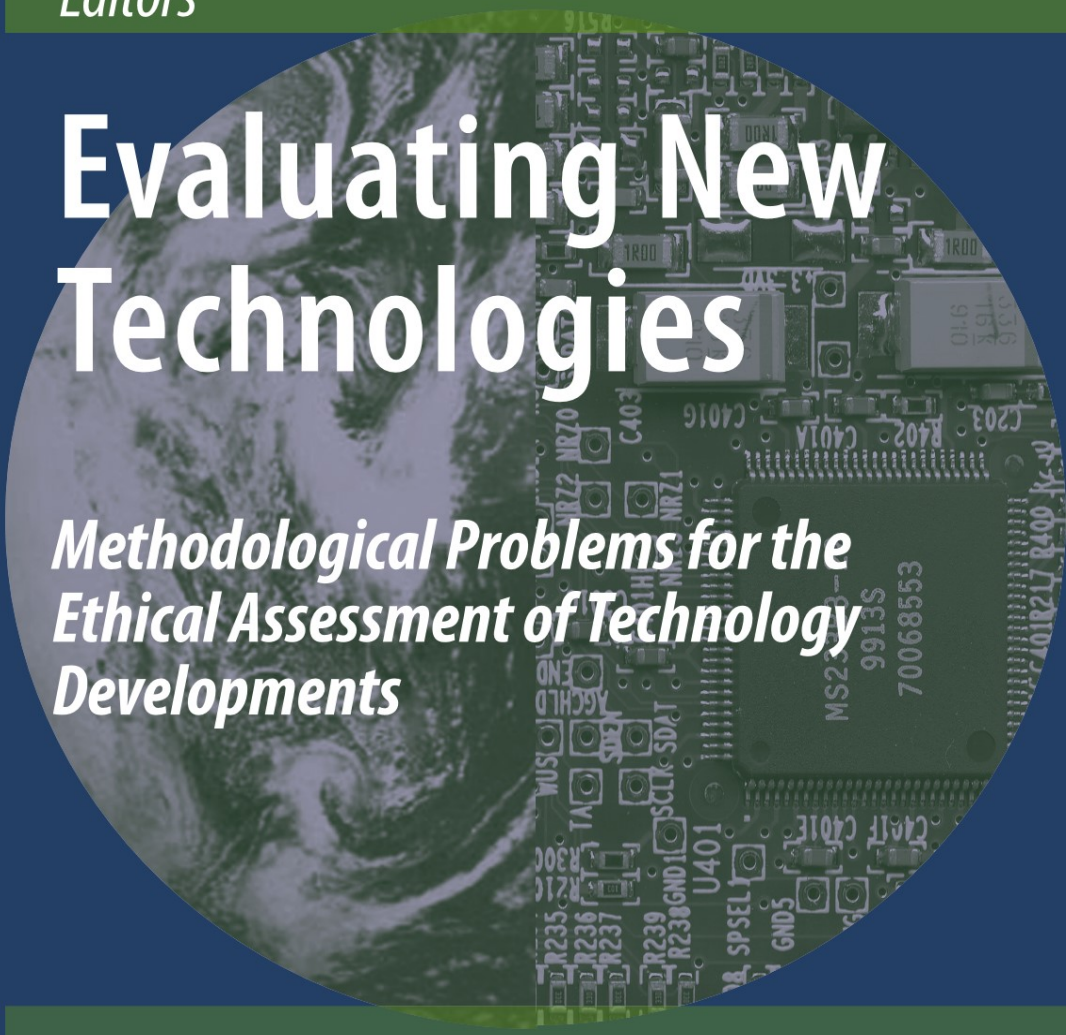
The International Library of Ethics, Law and Technology 3

Paul Sollie
Marcus Düwell
Editors

Evaluating New Technologies

Methodological Problems for the Ethical Assessment of Technology Developments



Springer

Evaluating New Technologies

The International Library of Ethics, Law and Technology

VOLUME 3

For other titles published in this series, go to
www.springer.com/series/7761

Paul Sollie · Marcus Düwell
Editors

Evaluating New Technologies

Methodological Problems for the Ethical Assessment of Technology Developments

Springer

Editors
Paul Sollie
Utrecht University
Ethics Institute
Heidelberglaan 8
3584 CS Utrecht
The Netherlands
paulsollie@gmail.com

Marcus Düwell
Utrecht University
Ethics Institute
Heidelberglaan 8
3584 CS Utrecht
The Netherlands
m.duwell@uu.nl

ISBN 978-90-481-2228-8 e-ISBN 978-90-481-2229-5

DOI 10.1007/978-90-481-2229-5

Library of Congress Control Number: 2009921731

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

Contributors

Deryck Beyleveld Durham Law School, Durham University, Durham, United Kingdom; Ethics Institute, Utrecht University, Utrecht, The Netherlands deryck.beyleveld@durham.ac.uk

Marianne Boenink Faculty of Behavioural Sciences, University of Twente, Enschede, The Netherlands m.boenink@utwente.nl

Roger Brownsword School of Law, King's College London, London, United Kingdom roger.brownsword@kcl.ac.uk

Steve Clarke Faculty of Philosophy, University of Oxford, Oxford, United Kingdom stephen.clarke@philosophy.ox.ac.uk

Bart Custers Tilburg Institute for Law, Technology and Society, Tilburg University, Tilburg, The Netherlands; Capgemini Consulting Services, Utrecht, The Netherlands b.h.m.custers@uvt.nl

Alfred Driessen MESA+ Research Institute, University of Twente, Enschede, The Netherlands A.Driessen@ewi.utwente.nl

Marcus Düwell Ethics Institute, Utrecht University, Utrecht, Netherlands m.duwell@uu.nl

Anke van Gorp Innovation Policy group, TNO Quality of Life, Leiden, The Netherlands anke.vangorp@tno.nl

Jeroen van den Hoven Faculty of Technology, Policy, and Management, Delft University of Technology, Delft, The Netherlands m.j.vandenhoven@tudelft.nl

Nicole C. Karafyllis College of Humanities and Social Sciences, United Arab Emirates University, Al Ain (Abu Dhabi), United Arab Emirates karafyllis@t-online.de

Noëmi Manders-Huits Faculty of Technology, Policy, and Management, Delft University of Technology, Delft, The Netherlands n.l.j.l.manders-huits@tudelft.nl

Niels Nijsingh Ethics Institute, Utrecht University, Utrecht, The Netherlands n.m.nijsingh@uu.nl

Paul Sollie Ethics Institute, Utrecht University, Utrecht, The Netherlands paulsollie@gmail.com

Dirk Stemerding Rathenau Institute, The Hague, The Netherlands d.stemerding@rathenau.nl

Tsjalling Swierstra Faculty of Behavioural Sciences, University of Twente, Enschede, The Netherlands t.e.swierstra@utwente.nl

Anton Vedder Tilburg Institute of Law, Technology, and Society, Tilburg University, Tilburg, The Netherlands anton.vedder@uvt.nl

Peter-Paul Verbeek Department of Philosophy, University of Twente, Enschede, The Netherlands p.p.c.c.verbeek@gw.utwente.nl

Chapter 1
Evaluating New Technologies: An Introduction

Paul Sollie and Marcus Düwell

1.1 Introduction

During a Department of Defence news briefing in February 2002, Donald Rumsfeld was confronted with the question about reports that stated that there was no evidence of a direct link between Iraq and terrorist organisations. He subsequently answered:

> Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones. (Rumsfeld, 2002)

Many received this comment with a chuckle, but Rumsfeld's now famous reply carries a lot of truth. Many decisions in different spheres of life take place under conditions of risk and uncertainty, the known unknowns and unknown unknowns. Modern, complex technology developments in the field of, for instance, nanotechnology, synthetic biology, or photonics are paradigm cases of these categories.

Before further introducing the issue of complexity and uncertainty in ethical technology assessment let us commence with briefly elucidating the concept of technology. In this book we adopt a generic definition of technology that not only includes technological artefacts, such as iPhones, remote controls, airplanes, or batteries, but also technological processes or technological knowledge. Technology is considered as a multifaceted concept that is best echoed in Carl Mitcham's taxonomy of technology. For Mitcham technology is a four dimensional activity. (See, 1994, 161–266) First of all, technology is an object, which is the most common and readily to mind connotation of technology. Technological artefacts comprise tools, manufactured objects and the like. 'Technology as object is the most immediate, not to say the simplest, mode in which technology is found manifest, and it can include all humanly fabricated material artifacts whose function depends on a specific materiality as such.' (Mitcham, 1994, 161) Second, technology signifies

P. Sollie (✉)
Ethics Institute, Utrecht University, Utrecht, The Netherlands
e-mail: paulsollie@gmail.com

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_1,

technological knowledge. This knowledge from science, engineering, social and physical science is technological 'how to' knowledge, namely the knowledge or know-how to understand technology and to be able to apply it into technological artefacts. Third, technology not only includes artefacts and technological knowledge, but it also refers to the technological activity of making and using technological artefacts. 'Technology as activity is that pivotal event in which knowledge and volition unite to bring artifacts into existence or to use them', Mitcham asserts. (1994, 209) In this procedural meaning it concerns the process of problem solving, designing, research and development, invention, or innovation in technology developments. Fourth, technology is volition, by which Mitcham means that technology is a social construction or force. In this connotation technology is perceived as what one does or wants with technology and, moreover, how it influences human beings and their behaviour.

To return to the complexity of technological innovations, it is a common feature of complex technologies that, whilst under design, one is often ignorant and uncertain of the possible applications and consequences, which result from a multifaceted heterogeneous field of political, social, moral, economical, and scientific forces. Ethical aspects, impacts, and future consequences of technology developments are often not apparent at the outset. Moreover, these future consequences may be unanticipated, unintended, and unforeseen. This pertains to what David Collingridge has coined the control dilemma:

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

The control dilemma highlights two problems of technology assessment represented by the two horns of the dilemma—the prediction horn and the control horn. The prediction horn of the dilemma states that, although in the early stages of the design of new technologies control is in principle very well possible, it is in fact meaningless due to a lack of relevant information and a subsequent inability to forecast the future. While designing technologies, one is often ignorant of possible adverse consequences and side effects that may harm society at large. When, after a long and complex process, technologies are introduced to and adopted by society, one has a variety of instruments for assessing the impact of new technologies on society via e.g. social scientific, economic, and ethical research. Conversely, the control horn of the dilemma is connected with the control of a stabilised technology. In this phase the technology is in use and the consequences and impacts gradually can become apparent. Whereas in the design phase it may be considered illogical to speak of control of technology due to the shortage of information to guide the assessment, the phase of application and stabilisation is characterised by the fact that this information becomes available. At this point, however, adjustments are costly and slow. Hence, the question of how and who should control emerging technologies.

This dilemma gives rise to different and intriguing questions. How will new technologies materialise and what is their impact and influence? How do they transform

human practices? How are we to morally evaluate technology developments that have open horizons, encompass uncertainties, and lack control? Technology is influential on society; technological innovations act upon the perception of ourselves, the world, and our relation with fellow humans and other objects. Technology is changing everything we do by creating new entities (such as software, nanoparticles, or Internet), by changing the scale of activities (e.g. vast amounts of data about people can be stored and analysed, and not infrequently without people being aware of this), by generating new kinds of knowledge (for instance about illnesses, the human genome and so on). Technologies, as a consequence, impinge upon our morality and for this reason an ethics of technology should not wait passively until moral problems arise and not only focus on identified and existing moral problems, but contemplate technology developments and possible impacts proactively. However, this is easier said than done, because a prospective and proactive evaluation of technology developments is complicated by complexity and uncertainty.

The uncertainty of technology development is closely related to one of the striking features of technology, namely what Jim Moor has coined logical malleability. (1985, 269) Technological devices are logically malleable in that they can be shaped to do any activity that can be characterised in terms of logical operations. Computers, according to Moor, are generic, general purpose machines that have now intruded every sphere of life. We belief that we may extend this notion to technology at large and advocate that the malleability of technologies allows them to be used in new and unforeseen ways, ways for which we frequently do not have policies regarding the control of applications and their effects. This, in fact, resembles Don Ihde's notion of multistability. (See, e.g., 2002, 106) Ihde argues that technologies are multistable, because they can be used for a variety of purposes and, therefore, be conceived of differently according to specific contexts of application. Technologies can replace and simplify existing devices, processes, actions, but also allow for unforeseen applications and impacts. For the same reason it is maintained that technology developments are morally non-transparent or opaque of nature (See, e.g., Brey, 2001, 52–53). These technological practices are not (yet) morally controversial, but nevertheless have moral import. They may be morally opaque for two reasons: they are unknown or they have a false appearance of neutrality. Many of the practices and social consequences related to complex new technologies are (yet) unidentified, because they are not revealed or invisible to people. For example, many, if not most, people use cell phones while being ignorant of the fact that it is their cell phones they might be tracked and traced by. Furthermore, over the course of the past decades, technology is more and more considered, especially and notably with the advent of ICT and medical technologies, as a practice that has moral and political impact. (See, e.g., Winner, 2004; or Verbeek, 2006) Although technologies were often thought of as being morally neutral, many technological artefacts have false neutrality and, hence, are not morally neutral. Search engines, for instance, are developed to help internet users find information quickly, but the particular algorithms underlying these engines are often far from neutral. (See, e.g., Introna and Nissenbaum, 2000).

The uncertainty of future consequences not only relates to the applications of these technologies, but also to the pervasive impact of technologies on society. Although technology is easily one of the most permeating and consequential features of modern society, surprisingly, an ethics of technology is still in its infancy. Important reasons for this 'underdevelopment' of a methodology for morally evaluating technology development are related to its complex, uncertain, dynamic, and large-scale character that seems to resist human control. The uncertainty surrounding technology development is one of the problems ethics of technology must deal with. The Dutch technological project *Towards Ultrafast Communication* (TUC) is a paradigm case of a complex, uncertain technology development. In part I of the book a number of authors will take up the example of TUC as the centre of their analysis. In the ensuing paragraph—and more extensively by Alfred Driessen in Chapter 2—the project of TUC is elaborated on.

1.2 Project 'Towards Ultrafast Communication' (TUC)[1]

TUC is a technological initiative that aims at bringing photonics at the centre of information and communication technologies. 'Without doubt, light has become the dominant medium for transmitting information. In fact, photonics is considered to be the most important key technology of this century, to such an extent that one might refer to the present century as that of the photon, just like last century was that of the electron.' (Vedder and Lenstra, 2006, 3)

The past decades, electronics have dominated information and communication technologies. The speed of developments within this area is unprecedented. In 1965, Intel co-founder Gordon E. Moore predicted—popularly known as 'Moore's Law'—that the number of transistors on a chip (i.e. the processor speed) and size of memory doubles roughly every two years.

> The complexity for minimum component costs has increased at a rate of roughly a factor of two per year . . . Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer. (Moore, 1965, 2)

Contrary to Moore's prediction in the seventies, his 'law' still seem to hold, and some experts, like those engaged in the TUC project, are convinced that it may hold for a decade or more. However, new technological developments in the field of computing and telecommunication technologies, such as the Internet, will challenge electronic technology.

---

[1] A project funded by Netherlands' Organization for Scientific Research (NWO) (http://www.nwo.nl/projecten.nsf/pages/1900117719?opendocument&nav=EOB_15_NL) and the Technology Foundation STW (http://www.freeband.nl/kennisimpuls/projecten/tuc/ENindex.html).

The speed of communication on the Internet has increased tremendously over the past years. The speed is largely determined by factors such as the speed of the client, the capacity of the data line, the use of the data line, the routing, and the speed of the server. With current technologies, such as optical fibre cable, sending packets of information over the Internet at very high speeds is rather unproblematic. However, two trends might compromise future internet traffic by causing congestion and necessitate more capacity of the Internet network. First, the amount of data transmitted over the Internet will continue to increase when new possibilities open up for people, for instance the streaming of 3D movies. Second, the number of people entering and using the Internet is still growing significantly.[2] These trends will ultimately impact on the capacity of the network. While sending packets of information at very high speed is not a problem with current technologies, restraining factors are the nodes where information is processed on its way from client to user. This suboptimal situation might be improved on when the processing speed at nodes is increased and congestion is avoided. In the current situation these nodes are fully electronically operating, which is significantly slower than using photonic technology. As these computers are unable to process information optically, they are the delaying factor in the network slowing down the transfer or ultimately causing congestions. At these Internet nodes information needs to be converted from optical information into electronic information. In order to overcome future problems of congestion, a consortium of engineers led by Daan Lenstra have engaged in the technology program of TUC trying to increase the power of processors to prevent anticipated problems of congestion. Compared to the all optical switching, electronic transistors produce enormous amounts of heat that subsequently necessitates the cooling of the electronic devices that in turn involves increasing use of energy. Photonics does not suffer from this problem and entails a decrease in energy. If they succeed—and the first results are promising (See Vedder and Lenstra, 2006, 5)—then the speed of sending information over the Internet might even be increased as the processing speed at nodes is boosted and congestion is avoided while at the same time reducing the consumption of energy. The TUC research project aims at just this aspect of communication, namely developing a prototype optical chip that can form the basis for ultra fast telecommunication nodes with a network capacity of 1 Terabit per second or higher leaving conventional electronics far behind.

To sum up, TUC aims at creating faster and more effective use of information and communication networks, and at creating fast access to a broad range of sources of information and instruments for communication. Improving means of communication does not seem very risky in itself, let alone to raise any concrete moral dilemmas. Nevertheless, we are uncertain about future consequences and applications of complex technology developments such as TUC. The same uncertainty also accompanies developments in the field of, for instance, nanotechnology and synthetic biology. One of the interesting issues that arises from the complex, uncertain technology development is how we should deal with new technologies. How

[2] See, e.g., http://www.internetworldstats.com/stats.htm.

and when should new technologies be ethically evaluated and controlled? It is for this reason that we address the question of how to formulate a framework for an ethics of technology that is able to deal with complex technology developments. These are not only interesting questions for ethicists, but also for engineers, politicians, or policymakers. Although the articles in this book are predominantly written from an ethical perspective, it is however our conviction that the articles prove to be useful and insightful beyond the field of ethics.

1.3 Thematic Structure

This book will take up methodological issues that accompany the ethical assessment of complex technological developments. The central theme of the book is disentangled in three parts:

1. The technological programme of 'Towards Ultrafast Communication' (Part I).
2. Methodological issues of the ethical assessment of new technologies (Part II),
3. Uncertainty and precaution as central aspects of complex technology developments (Part III).

Part I—A Case Study: Towards Ultrafast Communication (TUC)

The first part of the book elaborates on the technological programme of TUC and highlights a variety of ethical aspects that accompany TUC. First, TUC and its ethical aspects are further introduced by Alfred Driessen (Chapter 2). Thereafter, TUC is discussed in relation to specific issues of its ethical assessment. Anton Vedder and Bart Custers (Chapter 3) attend to the problem of responsibility in the development of information and communication networks like TUC—in which stage should ethical issues, such as quality of information, privacy, or security, be identified and addressed, and by whom? Next, Anke van Gorp (Chapter 4) develops a checklist for engineers and researchers to help them identify ethical issues in technology development. Finally, Noëmi Manders-Huits and Jeroen van den Hoven (Chapter 5) introduce and discuss value-sensitive design as an approach for reflecting upon developments in ultrafast communication.

Part II—Evaluating New technologies—Methodological Issues

In Part II of the book the focus is on a variety of methodological issues in ethical technology assessment that stem from the complexity and uncertainty of technology developments. First, the relation between human beings and technology and, more specifically, the moral significance of technology and the implications thereof for ethics of technology is investigated by Peter-Paul Verbeek (Chapter 6). Niels Nijsingh and Marcus Düwell (Chapter 7) scrutinise the relation between ethics and empirical science with regard to the evaluation of technology developments. Nicole Karafyllis (Chapter 8) provides a critical reflection on the recent tendency to transform technology assessment into vision assessment by scrutinising the methodological, anthropological, and ethical weaknesses of this type of assessment. Lastly, Tsjalling Swierstra, Dirk Stemerding, and Marianne Boenink (Chapter 9)

investigate scenario analysis for anticipating the moral consequences of technology developments, which is a popular method of preparing for the future while at the same time acknowledging its essential uncertainty and openness.

Part III—Evaluating New technologies—Uncertainty and Precaution

Part III takes up the issue of uncertainty and precaution. Each of the contributions investigates the role uncertainty and precaution play and should play in ethical technology assessment. To begin with, Paul Sollie (Chapter 10) investigates the concept of uncertainty in relation to technology developments. A three-dimensional typology of uncertainty is introduced that sheds light on and has implications for the ethical assessment of complex technologies that are characterised by uncertainty. Next, Steve Clarke (Chapter 11) puts the precautionary principle to scrutiny by addressing a number of problems that affect the influential versions of the precautionary principle. Recognising the many criticism that precautionary principles have been attacked with, Deryck Beyleveld and Roger Brownsword (Chapter 12) investigate the role of precautionary reasoning. They argue that there are conditions under which precautionary reasoning is valid by formulating a general principle for the limiting case. Finally, Paul Sollie (Chapter 13) presents a principlistic approach based on Gewirthian ethics for dealing with complex technology developments that are surrounded by uncertainty.

Acknowledgment First of all, we would like to express our gratitude to the project partners of the TUC-project at the Netherlands' Organization for Scientific Research (NWO) and the Technology Foundation STW: Daan Lenstra, Alfred Driessen, Anton Vedder, Bart Custers, Anke van Gorp, and Louis Logister. We would like to thank collaborators at the Ethics Institute (Utrecht University) and two anonymous reviewers from Springer for their critical comments and suggestions that helped to improve this book.

References

Brey, P. (2001). Disclosive Computer Ethics. In: Spinello R.A. and Tavani, H.T. (Eds.), *Readings in Cyberethics*. Sudbury, MA, Jones and Bartlett Publishers Inc., 51–62.

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

Ihde, D. (2002). *Bodies of Technology*. Minneapolis: University of Minnesota Press.

Introna, L. and Nissenbaum, H. (2000). The Public Good Vision of the Internet and The Politics of Search Engines. In: Rogers, R. (Ed.), *Preferred Placement. Knowledge Politics on the Web*. Maastricht, Jan van Eyck Akademie Editions, 25–47.

Mitcham, C. (1994). *Thinking Through Technology. The Path Between Engineering and Philosophy*. Chicago: The University of Chicago Press.

Moor, J.H. (1985). What is computer ethics? In: *Metaphilosophy*, 16, 4, 266–275.

Moore, G.E. (1965). Cramming More Components onto Integrated Circuits. In: *Electronics*, 38, 8, 1–4.

Rumsfeld, D. (2002). Transcript of Department of Defence News Briefing, 12 February 2002. URL: http://www.defenselink.mil/transcripts/transcript.aspx?transcriptid=2636. Last accessed on: February 9, 2008.

Vedder, A. and Lenstra, D. (2006) Reliability and security of information. In: *Journal for Information, Communication and Ethics in Society*, 4, 1, 3–6.

Verbeek, P.P. (2006). Materializing Morality. Design Ethics and Technological Mediation. In: *Science, Technology and Human Values*, 31, 3, 361–380.
Winner, L. (2004). Do Artifacts Have Politics? In: Kaplan, D.M. (Ed.), *Readings in the Philosophy of Technology*. Lanham, Maryland, Rowman & Littlefield Publishers, Inc., 289–302.

Part I
A Case Study: Ultrafast Communication

Chapter 2
Ethical Aspects of Research in Ultrafast Communication

Alfred Driessen

Abstract This chapter summarizes the reflections of a scientist active in optical communication about the need of ethical considerations in technological research. An optimistic definition of ethics, being the art to make good use of technology, is proposed that emphasizes the necessarily involvement of not only technologists but also experts in humanity. The paper then reviews briefly the research activities of a Dutch national consortium where the author had been involved. This mainly academic research dealt with advanced approaches for ultrafast communication. In the next section an assessment is given of the potential impact of the technological results on society. In order to emphasize the positive aspects and counteract the negative ones, three steps are proposed: (i) create conditions for a dialogue between experts in ethics and technology; (ii) work out scenarios for the introduction of new techniques in society; (iii) anticipate opportunities and threats. Finally the conclusions are presented.

Keywords Optical communication · Ethics in technology · Future scenario · Education · Technology assessment

2.1 Introduction

Over the last decade there has been an increasing interest in ethics in a number of fields not traditionally associated with ethics, such as business, medicine, publishing, science, and engineering. In the Netherlands an initiative started as part of a national research project administrated by the Dutch Foundation for Technology (STW).[1] In this project ethics-related activities were carried out parallel to the technological research to which it applied, and interaction between the technological

A. Driessen (✉)
MESA+ Research Institute, University of Twente, Enschede, The Netherlands
e-mail: Driessen@ewi.utwente.nl

[1] Towards Ultrafast Communication, Freeband Kennisimpuls (Dutch for "knowledge impulse"). Financed by the Dutch ministries of EZ and OCenW and coordinated by Prof. D. Lenstra, VUA.

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_2,

and ethical parts was stimulated.[2] At the end of the project two symposia were organized[3] where researchers from both parts presented their findings. The present paper is based on a presentation given at one of the symposia.[4] Being involved in applied science and—more specifically—being a project leader of the above mentioned national research project, I intend to summarize my reflections on the need of ethical components in technological research. The STW project is called Towards Ultrafast Communication, and it involved three Dutch universities: The Technical University of Eindhoven (TU/e), the University of Twente (UT), and the Free University Amsterdam (VUA).

In order to start one should explain what ethics could mean in the context of technology. Already in Greek philosophy ethics was studied and brought in connection with what is specific to human beings.[5] In the Aristotelian vision this is the capacity to guide oneself by using reason. In modern times the same connection is made, as this capacity is seen as the foundation of human dignity, and it has led to the conceptual foundation of human rights.[6] One therefore could say that a certain behavior or choice is ethical if it is in accordance with reason and thereby in accordance with the dignity of human nature. An optimistic definition is given by J.L. Lorda: "Ethics is the art to live well".[7] Focusing on our specific field of interest one could say that ethics in connection with technology is *the art to make good use of technology.*

In this definition the term "technology" can be taken as the concrete apparatus, a car, for example, or as a certain knowledge and infrastructure of apparatuses and equipment as expressed in terms like "nanotechnology". In the latter meaning, the design and development of new products and the choices made during development should also be considered as an important part of ethics of technology. The qualification "good use" should be taken in the widest range of meaning. The use of technology should objectively be good, that means all technical aspects should be optimized including related issues, e.g., the responsible use of resources and possible waste. But the goodness should also appear in the subject, that is, in the human being who makes use of the technology. Only if technology is used reasonably—respecting commonly accepted human values, i.e. according to human dignity—can one speak of ethical behavior.

---

[2] For more details see A.H. Vedder and D. Lenstra, "Reliability and Security of Information". J. Inform. Commun. Ethics Soc. 4(1), 3–6 (2006).

[3] The two symposia were held at NWO, The Hague on 14-2-2007 and 20-4-2007 respectively.

[4] A. Driessen, "*Is er een behoefte aan een ethische component bij technologie onderzoek?*" A lecture presented at the NWO symposium on Moral Technology Assessments, The Hague, 20-4-2007.

[5] See, e.g., Aristotle, Eudemian Ethics.

[6] Universal Declaration of Human Rights, UN. 10-12-1948, Article 1, "All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood." See http://www.un.org/Overview/rights.html.

[7] J.L. Lorda, *Moraal de kunst van het leven*, De Boog, Amsterdam 2006.

In the above definition ethics is defined as an art, meaning that two aspects should be considered: knowledge and skills. These two aspects are not necessarily found in a single person. Focusing, for example, on the art to play a violin one finds that excellent theoretical knowledge of a violin neither implies that you are a virtuoso nor a composer, and the same is true vice-versa. Applying this paradigm to ethics of technology one could state that the engineer or scientist is not able to deal on his own with all aspects of the art to make good use of technology. He misses the expertise to guide the subject, the human being, in his acting ethically. For this experts in humanity are needed who scientifically study human behavior and the values related to human dignity.

The paper is organized as follows: after the brief introduction with a tentative definition of ethics in the field of technology, a sketch of the field in which the technological research project and ethics apply is presented. In order to illustrate the work performed in the four years of the project, two examples of research highlights are given. Thereafter an assessment is made of the consequences of ultrafast communication on society. It is obvious that besides the positive aspects there are also quite a number of negatives that can be distinguished. In order to emphasize the positive and counteract the negative, three steps are proposed. Finally the conclusions are presented.

2.2 Technical Aspects of Ultrafast Communication

The national research project Towards Ultrafast Communication deals with transmission of data with speed beyond 100 Gbit/s or 0.1 Tbit/s. This speed is still slow when compared to the capacity of a single optical fiber, which exceeds 10 Tbit/s. It is high, however, if one realizes that in this way the data of 3 DVDs can be transmitted within a second. Comparing this speed to what is called today a broadband connection of 1–10 Mbit/s one may put question marks about the practical applications. Recent predictions, however, assume the introduction of 1 Gbit/s per user already within a decade (see Table 2.1).[8] In this case, at the nodes of the access networks data in the Tbit/s range would have to be routed.[9] Currently available electronic equipment is far from being able to handle such high bitrates. Moreover, even if they were able, power consumption would be a serious problem. Optical techniques would offer attractive solutions. At the TU/e a significant breakthrough was made in 2004 by realizing an all-optical switching element, a “flip-flop”, with a switching time below 18 ps (see Fig. 2.1).[10] After that, Dorren et al. constructed devices with

[8] R.E. Wagner, J.R. Igel, R. Whitman, M.D. Vaughn, A.B. Ruffin and S. Bickham, “Fiber-Based Broadband-Access Deployment in the United States”, J. Lightw. Techn. 24, 4256–4539 (2006).

[9] A.M.J. Koonen, “Fiber to the Home/ Fiber to the Premises: What, Where, and When?”, Proc. IEEE, 94, 911–934 (2006).

[10] M.T. Hill, H.J.S. Dorren, T. de Vries, X.J.M. Leijtens, J.H. den Besten, B. Smalbrugge, Y.S. Oei, H. Binsma, G.D. Khoe and M.K. Smit, “A Fast Low-Power Optical Memory Based on Coupled Micro-Ring Lasers”, Nature 432, 206–209 (2004).

Table 2.1 Projection of user demand for bandwidth, showing a Gbit/s target within a decade[8]

Technology	Per user	Start	50%	
Phone modems	<100 kb/s	1993	2001	History
Cable modems	1 Mb/s	1998	2006	
FTTx approaches	10 Mb/s	2004	2012	
Next gen fiber technology	100 Mb/s	2010	2018	Forecast
Big broadband technology	1 Gb/s	2016	2024	

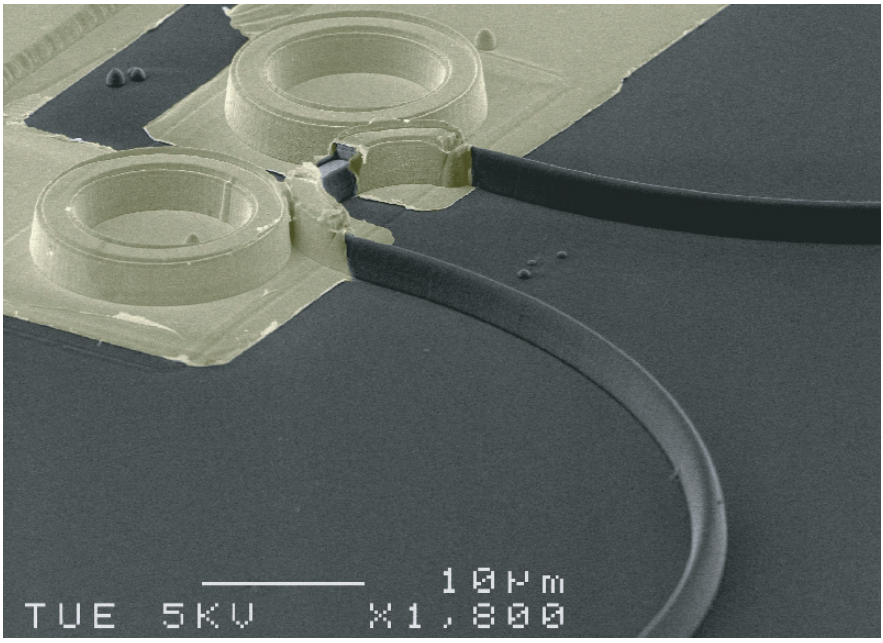
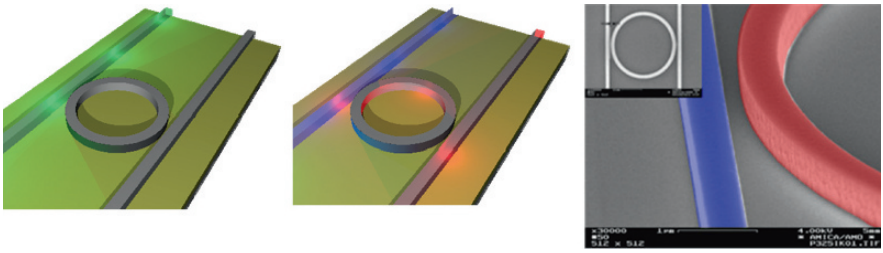


Fig. 2.1 Electron microscope picture of an optical flip-flop based on ring resonators with <18 ps switching time. The device has been fabricated at TU/e[10] and occupies approximately $60 \times 40\ \mu m^2$

even higher switching speed up to 640 Gbit/s. Dekker and al. from the UT used a different approach and were able to achieve sub-ps switching in a compact silicon waveguiding structure (see Fig. 2.2).[11]

Will there be an upper limit to this development? Already more than a decade ago Harmen R. van As mentioned during an international conference that there will be

[11] R. Dekker, A. Driessen, T. Wahlbrink, C. Moormann, J. Niehusmann and M. Först, "Ultrafast Kerr-induced All-Optical Wavelength Conversion in Silicon Waveguides Using 1.55 μm femtosecond Pulses", Opt. Express, 14, 8336–8346 (2006).



Pump pulse: off | Pump pulse: on | Electron microscope picture of device

Fig. 2.2 Optical switch based on ring resonators with <1 ps switching time.[11] The device has been designed and characterized by UT in cooperation with the RWTH Aachen and occupies approximately $8{,}000 \times 15\ \mu m^2$

no objective limit in communication with respect to duration and speed.[12] In the not so distant future—he states—technology will provide virtually unlimited bandwidth for everybody, anytime, and anywhere. The only limit will be the subject, the human being, who has not more than twenty-four hours a day available for communication.

How long does it take for new technologies to achieve widespread use and impact society? And how does this process proceed? In many cases roughly five stages can be distinguished. There is first the discovery of a scientific phenomenon, based mostly on years of intensive fundamental research. Many applications of the technology are immediately foreseen, but it takes time to demonstrate the potential in a device or apparatus that can be shown to the non-specialist who eventually will be able to bring it to market. Often the engineers and scientists directly involved believe in their approach and start their own spin-off company to work on development and perhaps eventually on small-scale, commercial production. When shifting to large-scale production the impact of the new technology on society becomes visible with all the beneficial and sometimes less attractive, unintended consequences. In the previous phases, certain societal aspects have already been explored, but only after large scale application the juridical and ethical issues will be analyzed and solved. There is no law established for the different stages and their duration. Yet based on empirical observation one could set-up the following normal schedule:

Year X:	discovery of scientific phenomenon
Year X + 10:	demonstration of application in laboratory
Year X + 20:	start of commercial production
Year X + 30:	impact on society becomes visible
Year X + 40:	ethical and juridical issues studied and (often) solved.

When applying this schedule to ultrafast communication one may expect that within the next two decades a number of far-reaching developments are made. First there is a seemingly unlimited “tele-presence” with audio and video that

[12] H. R. van As, Vienna University of Technology, private communication.

includes—in a later stage—perhaps even touch and smell. The application fields include entertainment, education, healthcare, the nursing and caring industry, as well as things related to safety and vigilance.

But not only what we call "the real world" will be connected. Virtual worlds are already part of our real worlds—created by human-computer interaction, Second Life is one example.[13] The access to these increasingly sophisticated worlds will become easier with fewer technological limits. What is called "intelligence" will be everywhere around us, at home, at the office, on the road. The technological basis responsible for this is the presence of personal networks around us which are built of all kinds of mobile devices and is connected by broadband to powerful servers and databases.

In an optimistic but nevertheless realistic view, truth and knowledge in all fields of human activity can be regarded as positive. This is valid even if one takes the considerable risk of potential misuse into account. This fundamental openness to progress applies also to technology of ultrafast communication. One should, however, also consider the inherent threats of the new technology. One finds, for example, possibility of new forms of intimidation and criminal behavior such as identity theft and voyeurism. The efficient control and manipulation of individuals and groups can become easier. Considering the single person in front of the new possibilities, one discovers new forms of addiction and, in consequence, new forms of exploitation of persons.

2.3 Measures to Be Taken

In order to emphasize the positive aspects of ultrafast communication and to avoid both the misuse and less desirable consequences, three steps are proposed:

1. *Create conditions for a dialogue between experts in ethics and technology.*

This apparently redundant point is nevertheless not so obvious. The reason is that beginning in secondary education our society creates a sharp separation between the studies in humanities ("alpha-studies" in the Dutch nomenclature) and science or technology ("beta-studies"). This separation commences in the first years of secondary school education where talented pupils are faced with courses exclusively in a single of the two mentioned subject areas. Why should a future engineer or scientist spend substantial time in studying history and classic languages instead of concentrating exclusively on physics, mathematics, and modern languages? And vice-versa? As a consequence of this unnecessary specialization, there is not only a fundamental ignorance of the other subject area, but sometimes even a certain mutual disdain. Apparently, in our enlightened age the old ideal of the *homo universalis* is not valid any more.

[13] http://www.secondlife.com.

There are several other issues that would make the dialog embarrassing. The expert in technology is mostly a specialist in a certain field of technology, often working in industry or commercial environment. His colleague in ethics, however, could be classified as a generalist, and is connected mostly to governmental or semi-governmental institutions (at least in Europe). Their language is different, as the jargon in the one field is unknown to the other.

2. *Work out scenarios for the introduction of new techniques in society.*

As explained above, the evolution of technology is often accurately predicted years or even decades in advance. Detailed roadmaps are established and business plans are made. These roadmaps are extremely important, as is illustrated by Moore's law predicting the increase of density of the electronic building-block—the transistor—on integrated circuitries.[14] This law has been valid for approximately four decades, starting from a few to currently one billion transistors per chip. This has only been possible through a common effort of an industry including millions of man-years of research and development (R&D) activities worldwide. There are a large number of roadmaps describing mostly a confined range of technological activities. All these start with past and today's state-of-the-art knowledge and then extrapolate with increasing uncertainty into the future. The uncertainty arises from nature, as technological breakthroughs are not directly related to quantitative effort in man-year and capital. But even more important to the timeline is the decision to put more resources into the development and the acceptance of the products made possible by new technologies. Here the unpredictable behavior of decision makers and end-users—free human beings—determines the success of technological innovation. In other words, what started as a technological issue is confronted with the world of humanities and the social sciences, which besides economics includes a broad spectrum of other disciplines. The response of the decision makers in any stage of the roadmap (and eventually the end-users) will determine the effort spent in implementing the roadmap.

The impact on society of a new technology depends on the acceptance of the technology by human beings. Therefore a joint effort has to be made by experts in the humanities as well as technology. In this way, scenarios can be made that alongside technical details include the contribution of disciplines dealing directly with man and his behavior. In order to illustrate this point one could recall the above-mentioned comment from the mid-nineteen-nineties, that the only limit in communication will be the limited hours a day the human being is able to communicate. Otherwise, there would come unlimited bandwidth anytime and anywhere. The uncertainty at the time of the prediction was not whether it would happen but only when. Similar far-reaching statements can be made today in other fields. It is important that the knowledge already now available in technology-related forums would be evaluated with regard to economical, juridical, pedagogical, psychological, sociological, medical, ethical, and other aspects. In this context ethics plays a

[14] see http://www.intel.com/technology/mooreslaw/index.htm

guiding role, as it provides the last word about the good use of a technology. For example, non-profitable technical developments—like walking on the moon—can be carried out because of a higher cause. An unethical application, where one could include the use of certain weapons of mass destruction, however, should be hindered even if it is connected with large economic profits.

A personal experience with scientists and engineers working on technology related scenarios could be mentioned. If one asks the people that are directly involved in technology the same question one does not always get the same answer. The reason for this is the different personal background and capacity to reflect on one's own scientific and technological effort. For a reliable view it is advantageous to be used to work on a meta-scientific level and—probably more important—to have the memory of several decades of active work in the field.

3. *Anticipate opportunities and threats.*

As mentioned before, the scenarios are increasingly uncertain when extrapolated further into time. In the major developments, the uncertainty is mostly related to the time of introduction and large-scale implementation. Concerning the roadmaps of specific issues, however, certain developments may be cut out completely. This demonstrates the high risk of investment in a particular technology. For example, the results obtained in our own TUC project will perhaps never directly be used in any future commercial production or product. Thus, when considering ethical aspects one probably should concentrate on the major developments (where the risk is reduced), as one can concentrate on certain device or system concepts which are largely independent of the specific technological implementation.

Once a scenario is worked out, it is possible to identify action points to arrive at a good use of the new technology. Currently, for example, we see in the Netherlands a nearly 100% penetration of broadband access and multimedia in families with school-age children. This situation could be foreseen at least a decade ago, see, for example the comment of Harmen van As made in Section 2.2. But only recently has society reacted to the new situation. Now children are consciously educated for a useful and enjoyable application of the new media, and there is consequently concern how to avoid any form of addiction. There is a demand for measures against undesired contacts (strangers) and undesired content (adult content). The juridical system, however, is not prepared for the new situation. In addition, research is needed on the long term consequences of intense use of multimedia with regard to health (RSI, obesity, changes in thought patterns), social behavior, creativity, and the ability to concentrate on intellectual work.

2.4 Conclusions

The main conclusion of the foregoing is that there is urgent need of an ethical component in technology research. Furthermore, it is not easy to organize the effective cooperation between experts in technology and ethics. In the case of environmental

issues, it is already common practice that for a new project with impact on the environment an Environmental Effect Report is produced in which the possible impact of the project is documented. In analogy to this one could suggest a Technology Effect Report (TER) every time a larger technological project is started, especially with public funding. This report should be a public document where a certain technological development and its possible alternatives are studied. On this basis then the expected consequences for man and society can be described in as systematic and objective as possible way. Once a TER has been worked out, the ethical evaluation proceeds to provide input for the decisions needed at critical points in the development and introduction of the new technology. These decisions, it should once more be emphasized, are made to enforce the positive aspects inherent in new technologies and to avoid the negative aspects as well as possible misuse.

The most important effect of the work on a TER could be that the two worlds of the scientist/engineer and the expert in humanities/social science would interact with each other. Often the first are not trained to reflect on the human and social aspects of their technology, whereas the second group does not realize the potential of currently-devised and often already demonstrated technological findings. By working together, the scenarios could be broadened to include all actors in the development of a new technology, reducing in this way the risk of poor decisions and improving the ethical character of technological progress.

The proposed activities are ambitious and much time from experts in technology and ethics is needed. Here the saying, “time is money”, is valid. On the long view, however, one may expect that investments in well-used technology, i.e., ethical technology will be even more economically profitable than the quick and often expedient rush to market with technology of unknown character and social effect.

Acknowledgment The author would like to thank Daan Lenstra, the coordinator of the national project on Ultrafast Communication, Anton Vedder for motivating discussions, and Paul Sollie, Jonathan Price and the referees for their stimulating comments. The financial support of STW is gratefully acknowledged.

Chapter 3
Whose Responsibility Is It Anyway? Dealing with the Consequences of New Technologies

Anton Vedder and Bart Custers

Abstract The infrastructure of our information and communication networks is quickly developing. All over the world, researchers are successfully working on higher capacity data transmission and on connectivity enhancement. Traditional limitations of time, space and quantity are gradually loosing their grip on the availability and accessibility of information and communication. These developments will change the world for the better in many ways. They can, however, have drawbacks as well. These are primarily concerned with the societal impact of the broader use of the technologies after they have been introduced into the market. In this chapter, we ask in which stage of the process of designing, developing, producing and introducing into the market of the technology these consequences should be identified, and by whom this should be done. We also focus on the responsibilities for addressing and solving these drawbacks. In this latter part of the essay, we detach ourselves a little from the practical setting of fast and ubiquitous networks and address a recently often heard claim, i.e., that reflection on the social, moral and legal aspects of technology should primarily take place in the phase of development so that solutions of possible problems can be quasi built into the device. We take a critical stance towards this claim and argue that concern and care for the social, moral and legal aspects should take place during the whole process, by different parties to the extent of their specific capacities and possibilities.

Keywords Ethics of technology · Responsibility · Privacy · Security · Reliability · Equal access to information

3.1 Introduction

The infrastructure of our information and communication networks is quickly developing. All over the world, researchers are successfully working on higher capacity data transmission and on connectivity enhancement. Traditional limitations

A. Vedder (✉)
Tilburg Institute of Law, Technology, and Society, Tilburg University, Tilburg, The Netherlands
e-mail: anton.vedder@uvt.nl

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_3,

of time, space and quantity are gradually loosing their grip on the availability and accessibility of information and communication. These developments will change the world for the better in many ways. They can, however, have drawbacks as well. These have to do with the quality of information, and the privacy, security, and accessibility of information and communication. They are primarily concerned with the societal impact of the broader use of the technologies after they have been introduced into the market. Subsequently, we will ask in which stage of the process of designing, developing, producing and introducing into the market of the technology these consequences should be identified, and by whom this should be done. Finally we will focus on the responsibilities for addressing and solving these drawbacks. In this latter part of the essay, we will detach ourselves a little from the practical setting of fast and ubiquitous networks. We will address a recently often heard claim, i.e., that reflection on the social, moral and legal aspects of technology should primarily take place in the phase of development so that solutions of possible problems can be quasi built into the device. We will take a critical stance towards this claim and argue that concern and care for the social, moral and legal aspects should take place during the whole process, by different parties to the extent of their specific capacities and possibilities.

3.2 Ultrafast Communication

Easy and fast network access is being realized in different ways. One way in which this is done, is by implementing high-capacity optical connections and flexible access to and in home networks. The use of a variety of wireless networks is rapidly growing. Examples are wireless local area networks, bluetooth and mobile telephony. The growth of both of these wireless networks and fibre-to-the-home connections will dramatically increase the need for more capacity in the wired part of the network. During the last decade, a vast amount of optical fibre cable has been installed in communication networks all over the world and even today new cables are laid at an astonishing rate. Especially the increasing number of fibre-to-the-home connections will put enormous pressure on the capacity of the upper hierarchy of long-distance networks. Most of the growth will be due to the expanding internet traffic.

In fact, these technological developments seem to be exponential. According to Moore's Law, the number of transistors on an integrated circuit (a "chip" or "microchip") for minimum component costs doubles every 24 months (Schaller, 1997). This more or less implies that storage capacity doubles every two years or that data storage costs are reduced by fifty percent every two years. This empirical observation by Gordon Moore was made in 1965; by now, this doubling speed is approximately 18 months. Moore's Law deals with storage capacities, but similar observations are made for communication speed and volume. According to Gilder's Law, the total bandwidth availability of US communication systems has tripled every twelve months since the 1980s and will expand at the same rate for the next 30 years to come (Raessens, 2001). Moore's Law is not only about making existing

technologies more efficient. It also takes into account the new ideas and inventions in the field of information technology. The latest developments to increase the speed and volume of information transfer on communication networks are focused on changing from electronic communication to optical communication. This is likely to result in a significant increase in the speed and volume of information transfer on communication networks. This new type of communication is referred to as *ultrafast communication* (Miller, 2004).

Ultrafast communication networks are mainly based on optical communication. Without doubt, light has become the dominant medium for transmitting information. In fact, photonics is considered to be the most important key technology of this century, to such an extent that one might refer to the present century as that of the photon, just like last century was that of the electron.

A crucial element in every network is the communication node, a facility in which information packages are received, inspected, buffered, labeled, redistributed and sent out again. They are present in every network, and the demand for higher capacity and throughput will manifest itself first at the higher levels of the network hierarchy. Presently, these nodes are fully electronically operating. This means that incoming optical signals are converted into electronic signals, then electronically processed, i.e. identified, buffered and labeled, and finally converted back into optical signals and transmitted to the user or to a next node. The processing speed of one conventional electronically operating node is generally 1 Gbit/s, i.e. 10^9 data bits per second. This may seem incredibly fast, but one should realize that the transmission capacity of a single ordinary optical fibre transmission cable is generally more than 100 Tbit/s, more than 100,000 times higher capacity than one electronic node. This means that if the fiber links in a telecommunication network are used to their full potential, the communication nodes will become bottlenecks for fast processing and rerouting of the data packages. Congestion of the whole network will unavoidably happen, not to speak of the danger of data packages getting lost forever. The solution to this problem is to develop a sufficiently fast alternative technology for data processing, preferably at teraherz speed, on the basis of which new types of nodes can be constructed. The potential maximum bit rate for a telecommunication link is set by the above-mentioned optical bit transmission capacity of the glass fiber. In order to deal with this enormous capacity in the node and to avoid the previously mentioned congestion problem, it would be very logical to stay in the optical domain all the way and hence make the network nodes optical as well. This means that a basic ultrafast optical device technology should be developed which will make the realization of all types of functionalities possible, such as, buffering of data, header recognition, switching of packages and regeneration of pulses. These devices must allow digital processing functions to be performed on data signals while “on the fly” and never leaving the optical domain (Cotter et al., 1999).

The engineers working on these devices find themselves in a situation comparable to that of the microelectronical challenge for electronic information processing in the1960s. Knowing the basic components needed for realizing the necessary functionalities, the challenge then was to realize microelectronical building blocks that could be integrated onto one single electronic chip device. We all realize now that

this development led to a revolution in electronic devices, ultimately bringing fast electronical equipment within reach of the general public; the personal computer being the most remarkable example. The ambition of Photonics is to make all of this much faster, not only in transmission speeds but also in bit manipulations per second. At the same time, the photonic circuits should become less power consuming in order to create opportunities for personal applications in portable versatile communication devices or for personal electronic health care. Thus, light will influence our way of living to an extent we never could have imagined just a few decades ago. Photonics will play a crucial role—often the central role—in our daily life, notably in the ways we communicate and in the tools we use to explore the frontiers of science.

To realize the photonic ambition, one needs to find optical alternatives for each electronic building block, such as flips flops, gates, buffers, memories, shift registers, transistors etc. The information in electronics is normally present in the form of binary units or bits, simply on or off. This is less restrictive in the photonics domain, since here the possibility of different parallel wavelengths in adition to the binary information handling introduces per wavelength channel an enormous flexibility in the way the information is digitized, which introduces a lot more design possibilities.

In short, new technology is developed for all-optical ultrafast signal processing and handling. This will lead to all-optical ultrafast telecommunication nodes that can handle the full potential of the existing optical fiber transmission capacity. All-optical building blocks have been realized in concept and the first integrated device versions will soon be fabricated. This development will make telecommunication networks in general and the Internet in particular orders of magnitude faster.

3.3 Consequences

In this section we will focus on a number of possible drawbacks. We focus on quality and security of information and communication, privacy, public security, accessibility, and exclusivity. We will provide only an overview of these drawbacks here, since detailed discussions are beyond the scope of this contribution.[1]

3.3.1 Quality and Security of Information and Communication

The introduction of the Internet has brought about considerable changes in the ways in which people communicate and disseminate and gather information. Remarkably,

[1] For more detailed discussions we refer to earlier publications, e.g., for more on quality and reliability, see Vedder and Wachbroit (2003) and Vedder and Lenstra (2006). For more on the drawbacks regarding privacy and public security, see Custers (2008) and for more on the drawbacks on exclusivity and the digital divide, see Compaine (2001).

people's ways of assessing reliability of information and safeguarding the security of communication are still, to a high degree, geared to traditional media. (Vedder, 2002) They relate to the—often institutionally embedded—signs of authority of the sources and intermediaries and to the recognisability of the details of the process of the transactions involved. With the growing speed of the information and communication networks two characteristics of the Internet are further enlarged. First, as the number of content providers and the ease of uploading information further increases assessing the true nature of sources and intermediaries of information becomes more difficult. Second, as the technologies involved become more sophisticated and complicated, the processes of interaction become less transparent. All of this diminishes the possibilities of assessing the trustworthiness of partners in communication and of information content providers, and of assessing the validity and reliability of information and of ensuring the security of transactions (Vedder and Lenstra, 2006).

Sometimes, the drawbacks of limited quality or reliability of information may not be obvious. However, some examples may illustrate the consequences of lacking data quality or flawed security. For instance, many people tend to increasingly rely on the use of medical information on the Internet for diagnosing their own medical situation. Since not everyone is a medical expert, this may lead to errors in such diagnosis. As a result, people may start to use the wrong medication or treatment. Obviously, not all medication can be obtained without seeing a health care professional, but the Internet also provides plenty of options of ordering medication abroad. Medication produced in countries with less strict quality assurance may result in even worse consequences.

Another example of how limited reliability and security of information on the Internet may have serious drawbacks concerns financial data. *Phishing* is the process of attempting to acquire personal information, such as passwords and credit card details, from people by pretending to be a trustworthy person or organization. A phishing attempt usually involves an email that asks users to fill out their personal data. Such emails may look reliable to the users, they may even be exact copies of messages from their own bank, in order to convince users to hand over their personal data. Once the data are sent, they can be used by the criminals to make financial transactions for their own benefit.

Reliability in the epistemologically normative sense that is relevant here is a matter of proper justification in terms of "content criteria" and "pedigree criteria" of reliability (Vedder and Wachbroit, 2003; Vedder, 2005). "Content criteria" refer to the conditions or criteria of reliability that are a function of the content of the information itself. Among these are the criteria of evidence that mostly belong to the domain of experts—people familiar with the subject or with a specific educational background or experience. Other examples of content criteria are logical criteria and, arguably, subject-matter criteria. "Pedigree criteria" are the conditions or criteria of reliability that relate to the authority and the established legitimacy and credibility of the source or intermediary of the information. Pedigree criteria are not merely used by non-experts. Experts use them as well. Pedigree criteria are established by credibility-conferring institutions. Here one can think of institutions

in a very broad way, ranging from well-organized institutes to broader—sometimes intricate and tangled—networks of cultural and societal arrangements. Earlier research has shown that many problems regarding reliability of online information on the Internet are not problems of information *lacking* reliability, but of receivers misperceiving or not perceiving (un-) reliability.

The very possibility of adequately recognizing pedigree criteria will increasingly diminish where fast networks with enhanced connectivity are concerned. Increasingly often, a content provider will be anonymous or will have merely a virtual identity. Also, the lack of traditional intermediaries (such as libraries, librarians, specialized publishers) has a negative influence on the capabilities of information seekers to assess the reliability of information. These kinds of factors leave the users often without clues or any indication whatsoever about the character, background, and institutional setting of the content provider. As the accessibility to a broader public of information originally intended for experts increases, the absence of intermediaries becomes gradually more problematic. Finally, as the connectivity and the possibility of providing content through networks increase, the opportunities for content providers to present themselves as others than they are—resulting in misinformation and, for instance, phishing—will multiply accordingly. Consequently, there will be a growing need for—not only—the development of new credibility conferring systems, but even for possibilities of identification and authentication of content providers.

3.3.2 Privacy and Public Security

Another issue is the role of ultrafast networks in public security and its implications for the privacy of individuals. Currently, lots of information on communication is collected and processed by judicial authorities and intelligence services themselves and by third parties such as telecom corporations and internet providers to support governments in their fight against crime and terrorism (Vedder et al., 2007) This surveillance can be distinguished in two main types, i.e. tapping and data retention. Tapping, or wire tapping, has been used for a long time and aims at monitoring the contents of any specific communication, such as phone call or an e-mail. In most modern countries there are very strict regulations to comply with before tapping is allowed. Data retention is a more recent form of surveillance. It is not aimed at the contents of any communication, but on the traffic data, i.e., surveillance of the call detail records of telephony and Internet traffic and transaction data. In March 2006 the European Union adopted a directive that requires telecom operators and Internet providers in all member-states to implement data retention systems for both telephone and Internet traffic.[2] Data retention focuses on the storage of call detail

[2] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.

records of telephony and Internet traffic and transaction data. Basically this concerns phone calls made and received, emails sent and received and web sites visited. These data provide an idea of who stays in contact with whom, when and how frequent. When possible, further identifying information may be added, as well as location data.

The new generation of ultrafast communication networks is likely to be a combination of optical and wireless. The former is relatively hard to tap; the latter is relatively easy to tap. Particularly for the wireless parts of ultrafast communication it is therefore recommended that cryptography is used to prevent unauthorized tapping. However, the cryptography used should not be too strong to be deciphered in cases in which tapping is allowed. The use of trapdoors and technologies such as key recovery systems, key escrow systems and trusted third-party encryption may be helpful to achieve this. These are systems and technologies in which exceptional access is possible. This enables users with additional information to circumvent the regular access and security procedures. For more detailed information, see Abelson et al. (1998) and Akdeniz (1998).

Whereas tapping concerns the contents of the communication, data retention focuses on storing and analyzing communication data, particularly call detail records regarding phone calls and Internet traffic. Ultrafast networks will require larger capacities for storing and analyzing data. The former is relatively easy, since storage capacity is ever increasing (though the costs involved are subject of a major discussion); the latter is a significant problem. Analyzing vast amounts of data needs to be done in automated ways, such as with the use of data mining. However, most data mining technologies are not yet very sophisticated for large-scale use. Furthermore, a major disadvantage is that the risk profiles resulting from the automated analyses may not be accurate, see Custers (2003). False-positives may result in investigating and even arresting innocent people. False-negatives may result in criminals and terrorists being out of scope. When risk profiles have limited accuracy, they should only be used with the utmost care, in order to prevent investigating and arresting innocent people. It is recommended to keep performing double checks on existing risk profiles and not to merely rely on data in databases, but also perform significant field work. In order to prevent the worst forms of unjustified discrimination and social polarization, it is recommended not to include sensitive personal data, such as religion and ethnic background, in the risk profiles, see Custers (2004).

In sum, tapping and data retention in the age of ultrafast communication networks may be very useful to reveal criminal and terrorist networks and to find first offenders. Both aspects are increasingly needed in the fight against crime and terrorism. However, because of the increasing amounts of data that are communicated over ultrafast networks, it is absolutely necessary to make, from the outset, selections on which data should be collected. Even though all data can be stored, it is not recommendable to do so, because the overview will be lost. A better idea is to make a selection of the data that may be useful. This will be a more targeted and effective approach than storing and analyzing all available data. This will be a lesser infringement of the privacy of those data subjects whose data is involved,

particularly of those who are innocent. Obviously, privacy infringements may be allowed in some cases, both from a moral and a legal perspective, particularly when dealing with organized crime and terrorism, but such infringements should be limited to a minimum. According to Etzioni (1999), limitations of privacy should only occur when there is a well-documented and macroscopic threat to the common good at stake. Even then, it should be considered whether the intended measures limiting the privacy are effective and whether the goal cannot be achieved with other measures that are less privacy-invasive. When this is the case and the privacy of individuals or groups of people is actually violated, measures should be considered to treat undesirable side effects.

3.3.3 Accessibility and Exclusivity

In the last decades, it has become clear that people are not only increasingly using information and communication technologies, but are also becoming increasingly dependent on them. As a result, many actions that people used to do in person or on paper are now performed digitally. For instance, many people are no longer booking their flight tickets through a travel agency in town, but use the Internet. In many countries, people request their tax returns via their home computers, no longer using paper files. Instead of going to a shop to buy a CD, many people nowadays download their music from the Internet. The dependency of people on information and communication networks raises questions on the accessibility and exclusivity of these networks. These questions are closely related to the debate on the so-called *digital divide*. For more on this debate, see, for instance, Compaine (2001), Van Dijk (2005), Mossberger et al. (2003).

Ultrafast communication networks are likely to introduce two barriers of access to users: costs and knowledge. These access barriers may result in excluding groups of people from access to these networks. In the early stages of introduction of new technological developments, it is likely that the costs will be kept low in order to get a critical number of users that communication networks usually require. This is a different approach the normally used in non-networked technologies, such as apartments, cars, or books, in which cases profit has to be made from selling the product itself. After a sufficient number of users is connected to the network, profit has to be made to compensate investments that were made, and costs for consumers to buy the products and services offered on the network will increase. These costs may decrease again after the introduction of more competition. Generally spoken, the number of users may depend on the costs involved and even though costs may be kept low, it is likely that there will always be groups (small or big) of people who are excluded.

The same goes for the knowledge that is required from users of sophisticated networks. In general, older people seem to have more trouble adapting to the latest technological developments. New technologies may expect more of users regarding education levels and may require users to be able to adapt to new concepts, such as using a mouse or a touch screen or talking to a computer on the phone. This

may also result in excluding some groups of people from access to communication networks.

When larger groups of people are excluded from the networks, this may cause social polarization between those who are included and those who are excluded. Apart from social polarization, another effect of exclusive networks may be the limited number of providers of structure and content, which may lead to manipulation of the information provided.

Because digital services may address most customers and may involve fewer costs, it may ultimately no longer be profitable for companies to have offices in town where people can go to for their products or services. As a result of Internet trade, many music stores, travel agencies and bank offices have already downsized or closed. It is expected that many more will follow in the years to come. This is likely to increase the exclusivity of networks, particularly when, in the next stage of technological developments, ultrafast networks will replace the existing networks. Here we will offer three suggestions that may help to deal with the above-mentioned effects of the exclusivity of ultrafast communication networks. The first and second aim to decrease the exclusivity of the networks by addressing the access barriers, the third addresses alternatives to exclusive networks:

- *Remove the costs barrier*
  The first reason for exclusivity are the costs involved. There may be several ways to remove this barrier. For instance, the costs may be compensated, or free access points, such as in libraries, may be provided.
- *Remove the knowledge barrier*
  The second reason for exclusivity is the knowledge required. There may be several ways to remove this barrier. For instance, by educating these groups and by providing more user-friendly access points.
- *Ensure off-line alternatives for basic needs*
  Some networks will be exclusive, when the barriers above cannot be removed. For most commercial networks this is not necessarily a problem. It may become a problem when there are basic needs involved for the users. Booking a flight ticket is generally not considered a basic need, but buying clothing or completing tax return forms may be considered so. For these applications, it seems reasonable to provide off-line alternatives, even if they may no longer be profitable after some time.

Current communication networks, such as the Internet, do not show large-scale social polarization and manipulation of information. Most basic needs can still be purchased off-line. Although ultrafast communication systems are likely to show at least some of the effects described on a larger scale, the suggestions above may help to minimize or avoid these effects. Exclusivity of networks is not necessarily a negative thing, as long as there remain some choices and alternatives for both those who are included and those who are excluded. However, exclusivity may become a drawback when it causes social polarization and prohibits people from access to basic needs.

3.4 Responsibilities Involved

In the previous sections we expounded the possible benefits and drawbacks of the development and introduction of technologies that can further increase the speed and accessibility of information and communication networks. We mainly concentrated on the broad social impact that these technologies may have in the future. We have not and we will not say much about possible benefits and drawbacks of earlier phases of the development and introduction into the market of these technologies. Of course, it is possible that at some stage of these earlier phases morally questionnable situations arise. It could be possible, for instance, that poisonous materials were used exposing researchers to health risks. Or that materials were used that are rare and extracted under very bad circumstances in developing countries. We did not refer to possible problems such as these (and to our current knowledge would not need to do so) because the expertise needed to identify these problematic issues belongs to the domain of other specialists, e.g. the engineers that perform the research themselves. This brings us to questions regarding the responsibilities for identifying and actually addressing possible positive and negative consequences of technologies.

For this question to be answered, it is important to consider the different phases in developing and introducing new technologies. Here we will distinguish the following phases:

1. Research and development
2. Production
3. Introduction into the market/society

When thinking through the division of responsibilities with regard to the diagnosis of and the response to possible opportunities and risks, it must be taken into account that many of the actors and stakeholders involved in the phases mentioned only have a very restricted insight into the opportunities and risks involved. Moreover, many of them have restricted means to respond. For instance, engineers are involved in the first phases, but have limited influence on the introduction of new technologies into the market/society. End users may have effect on how the new technologies are introduced into society and how the new technologies are actually used. However, end users have restricted means to influence research, development and production of new technologies.

It seems, therefore, appropriate to first distinguish between responsibilities to identify benefits and drawbacks, on the one hand, and responsibilities to act in order to respond to them, on the other. Recognizing advantages and drawbacks requires insight and expertise in the technological developments and also a kind of ability to see what would be socially beneficial or detrimental. This insight and ability need not necessarily be accompanied by the ability and capacity to respond, once advantages or drawbacks have been identified. For example: it may be an engineer's responsibility to identify risks to the environment involved in the production of an artifact, but that does not mean that the solution for this problem is this person's responsibility as well.

Second, a distinction should be made between the responsibilities of the different parties and individuals who are somehow involved, for instance: the researchers, supervisors, organizations funding the research, enterprises, governmental authorities, consumer organizations and other NGOs. The responsibilities with regard to the identification of drawbacks and advantages may vary according to the different expertise in technology and acquaintance with the various needs and preferences in society. The responsibilities for acting can vary with the different capacities and powers of the parties involved. An NGO may have a responsibility to do something about environmental risks caused by the production of an artifact, which an engineer involved in that process may not have, although he or she identified the risk and communicated it to the NGO involved—simply because the NGO is in a much better position and is apt to respond to phenomena like these.

It is by all means undesirable that all responsibilities are assigned to just one group of stakeholders, such as the researchers or engineers themselves. Nonetheless, the currently popular value-sensitive design approach in ethics of technology (Friedman et al., 2006) has a natural focus on the stage of design. This focus may be appropriate in light of the still very dominant view that technology and its design are in themselves morally neutral. Simultaneously, it is apt to draw away the attention from the other stages and from stakeholders other than the directly involved engineers. Coincidence or not, there is a growing tendency to restrict the window of possible interference in the case of flaws in technology to the pre-market phase. This tendency, pervades the plans of the Bush administration to restrict liabilities to risks involved in technology to the risks that were foreseeable at the time of design (Pear, 2004).

The responsibilities should be assigned to the various parties involved because of their different expertise, abilities and powers. They should not be restricted to the stage of research and development, simply because not all possible advantages and drawbacks can be known at that stage. The process of appraisal and critical evaluation should start in that first stage, but it should not also be finished in that stage. This will prevent that those involved in the different stages will push off responsibilities to others. It is still often the case that engineers and technicians suggest that they only build a particular technology that others can use for better or for worse. The end users, however, often suggest that they only use technologies for the purposes for which they were intended or designed. In the case of weapon technology, for instance, manufacturers usually claim end user responsibility, whereas victims of this technology suggest that their harm may have never occurred when the weapons were not manufactured in the first place.

In order both to facilitate the identification and to find creative responses, communication by all the parties involved throughout all stages would be desirable. With regard to possible drawbacks, one might be tempted to think only of suppression by for instance prohibiting the development or the exploitation of certain techniques, such as was initially the case regarding stem cell research in the United States. By initiating the communication and debate of all the stakeholders, however, one may try to find technical solutions in an earlier stage of the development. Or, conversely, it may become clear that certain technical solutions may not work, so

that accompanying regulatory measures are called for, once an artifact is introduced in the market. Ensuring higher degrees of user-friendliness may be a typical example of the former approach, educating users may be a typical example of the latter approach.

In this section, we have emphasized the plurality of parties involved and the variety of responsibilities of those parties. Each of these parties may sometimes feel tempted to shun away from specific responsibilities. For instance: an engineer specialized in a very specific part of the development of a new artifact may think that he or she is not responsible for thinking about its broader social impact since other parties have responsibilities as well. The fact that consumer organizations or a governmental authority may have responsibilities with regard to the design, production and introduction of an artifact, however, does not exempt the engineer. All of the parties have responsibilities based on their expertise and capabilities. It is hard to see, how anyone of them, especially the engineer, could deny that his or her expertise and ability are relevant. This may again be illustrated with the example of weapon technology: it may not be realistic for weapon manufacturers to push off responsibility to end users that may use the weapons for better or for worse. When weapons are manufactured on a large scale, it may be assumed that they will be used sooner or later to some extent. Instead of pushing off responsibilities, it is preferable to have joint responsibilities. Instead of creating gaps in the responsibilities, i.e., parts of the research and development process where nobody is responsible, this may create joint responsibilities. We consider overlapping responsibilities an advantage rather than a drawback in these cases.

3.5 Conclusion

New technologies are changing the world we live in. Many benefits come with the development and introduction of these new technologies. Using the example of ultrafast communication technologies, we investigated typical consequences regarding the quality of information, privacy, security, and accessibility of information and communication. As these effects cover the different phases in developing and introducing new technologies, the question was raised who is responsible for these effects, particularly the drawbacks of new technologies. Hence, it was suggested to distinguish responsibilities to identify benefits and drawbacks versus responsibilities to respond to them. A second distinction was made regarding the parties involved in the different stages of the development and introduction of new technologies.

Sometimes the parties involved are tempted to shun away from specific responsibilities. However, the fact that there are more than one party bearing certain responsibilities does not exempt the parties involved to take up their specific responsibilities. Joint responsibility during the whole process of development and introduction is recommended. This may be achieved by communication by all the parties involved throughout all stages. In this way, they will have more overview

over the whole process, benefits may be maximized and drawbacks and risks may be minimized. Minimizing drawbacks and risks may involve taking accompanying regulatory measures in cases where the drawbacks cannot be avoided, but this is not necessarily so.

References

Abelson, H., Anderson, R., Bellovin, S.M., Benaloh, J., Blaze, M., Diffie, W., Gilmore, J., Neumann, P.G., Rivest, R.L., Schiller, J.I., and Schneier, B. (1998) *The risks of key recovery, key escrow and trusted third party encryption*; A report by an ad hoc group of cryptographers and computer scientists. www.cdt.org/crypto/risks.

Akdeniz, Y. (1998) No Chance for Key Recovery: Encryption and International Principles of Human and Political Rights. In: *4th International Conference on Ethical Issues of Information Technologies,* Ethicomp 98, Rotterdam.

Compaine, B.M. (2001) *The Digital Divide: Facing a Crisis or Creating a Myth*, Cambridge MA: The MIT Press.

Cotter, D., Manning, R.J., Blow, K.J., Ellis, A.D., Kelly, A.E., Nesset, D., Phillips, I.D., Poustie, A.J., and Rogers, D.C. (1999), Nonlinear optics for high-speed digital information processing. *Science,* 286, pp. 1523–1528.

Custers, B.H.M. (2003) *Effects of Unreliable Group Profiling by Means of Data Mining*. In: G. Grieser, Y. Tanaka and A. Yamamoto (eds.) *Lecture Notes in Artificial Intelligence, Proceedings of the 6th International Conference on Discovery Science (DS 2003)* Sapporo, Japan. Berlin, Heidelberg, New York: Springer-Verlag, Vol. 2843, pp. 290–295.

Custers, B.H.M. (2004) *The Power of Knowledge*, Tilburg: Wolf Legal Publishers.

Custers, B.H.M. (2008) Tapping and Data Retention in Ultrafast Communication Networks, *Journal of International Commercial Law and Technology*, 3(2), pp. 94–100.

Etzioni, A. (1999) *The Limits of Privacy*, New York: Basic Books.

Friedman, B., Kahn, P.H., Jr., and Borning, A. (2006). Value Sensitive Design and information systems. In: P. Zhang and D. Galletta (eds.) *Human-Computer Interaction in Management Information Systems: Foundations*, Armonk, New York; London, England: M.E. Sharpe, pp. 348–372.

Miller, D.A.B. (2004) Ultrafast Digital Processing. In A. Miller, D.M. Finlayson, and D.T. Reid (eds.) *Ultrafast Photonics*, Bristol, Philadelphia: Institute of Physics Publishing.

Mossberger, K., Tolbert, C.J., and Stansbury, M. (2003) *Virtual Inequality: Beyond the Digital Divide*, Washington DC: Georgetown University Press.

Pear, R, (2004) In a Shift, Bush Moves to Block Medical Suits. http://query.nytimes.com/gst/fullpage.html?sec=health&res=9B05EED6173DF936A15754C0A9629C8B63 (accessed on 14th August 2007).

Raessens, B. (2001) *E-Business, Your Business*. Utrecht: Lemma.

Schaller, R.R. (1997) Moore's Law: Past, Present and Future, *Spectrum*, IEEE, Volume 34, June 1997, pp. 52–59.

Van Dijk J.A.G.M. (2005) *The Deepening Divide: Inequality in the Information Society*, Thousand Oaks CA: Sage Publications.

Vedder, A. (2002) What people think about the reliability of medical information on the Internet. In: I. Alvarez, T. Ward Bynum, J. Alvaro de Assis Lopes, and S. Rogerson, (eds.) *The Transformation of Organisations in the Information Age: Social and Ethical Implications.* Lisboa: Universidade Lusiada, pp. 281–292.

Vedder, A. (2005) Expert knowledge for non-experts: Inherent and contextual risks of misinformation. *ICES, Journal of Information, Communication and Ethics in Society*, 3, 113–119.

Vedder, A. and Robert W. (2003) Reliability of information on the Internet: Some distinctions. *Ethics and Information Technology,* 5, pp. 211–215.

Vedder, A. and Daan L. (2006) Reliability and security of information. *Journal for Information, Communication and Ethics in Society*, 4(1), pp. 3–6.

Vedder, A.H. et al. (2007) *Van Privacyparadijs tot Controlestaat? Misdaad- en Terreurbestrijding in Nederland aan het Begin van de 21ste eeuw.* Studies Rathenau Instituut (Ext. rep. 49). Den Haag: Rathenau Instituut.

Chapter 4
Ethics in and During Technological Research; An Addition to IT Ethics and Science Ethics

Anke van Gorp

Abstract Although IT ethics deals with interesting ethical issues that are of importance in and during technological research and development, these ethical issues are not always relevant for the individual researcher developing new technologies. I identify four reasons why ethical issues raised by IT ethics might be relevant for the goal of a project as a whole but not for the individual researchers working in the project. There are other ethical issues that are relevant for individual technological researchers. Based on the results of a case study a list of ethically relevant questions is developed that should help individual researchers to address the ethical issues he or she can address during research. This list should be seen as an addition to, not a substitution of, IT ethics and science ethics. Researchers can use the internet to find information about the ethical questions. However, they should keep in mind the issues concerning the reliability of the information on the internet. Using some examples I will show that researchers can obtain information about the ethical questions in the list relatively easily and they should be able to judge the reliability of the gathered information.

Keywords Ethics and technology · Technological research

4.1 Introduction

In this paper a list of questions will be proposed that helps researchers doing research in technological fields to identify ethical aspects of their research. This list can be seen as an addition to science ethics that focuses on informed consent in research with human subjects, the use of animals, ideas about authorship etc, see (Barnbaum and Byron, 2001). It is also an addition to IT ethics because IT ethics has not really addressed the individual researcher working in the early stages of developing new technologies. Both science ethics and IT ethics deal with important questions and this proposal is not meant as a substitute but as an addition to these

A. van Gorp (✉)
Innovation Policy group, TNO Quality of Life, Leiden, The Netherlands
e-mail: anke.vangorp@tno.nl

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_4,

fields. In the following I will first give an outline of the reasons why IT ethics is of little relevance to researchers working in early developments of what can become new IT. Following this I will introduce the ethical aspects that could be relevant in the early development of IT; these ethical aspects are summarized in a table at the end of Section 4.3. In the subsequent section sources are given where researchers can find relevant information to answer the questions in Table 4.1. To show that the questions and the information sources can give interesting and sometimes surprising results I will elaborate on two examples in Section 4.5. Further thoughts are presented in the last section.

4.2 IT Ethics and Technological Research

Most of the ethical aspects identified in IT ethics deal with the use of IT and problems associated with that use. The following ethical issues are often mentioned: reliability of information on the internet, security of IT, digital divide, cybercrime, privacy, see (Johnson, 1985; Introna, 1997; Nissenbaum, 1998; Tavani, 1999;

Table 4.1 First draft of a checklist with ethical issues for technological researchers

Safety:
– Are the compounds and processes used toxic?
– Are the compounds and processes hazardous for the health and safety of the researchers?
Sustainability:
– Are the compounds and processes eco-toxic?
– How much energy will the technology that is being developed use (including production and discarding)?Is it possible to decrease the energy use?
– Are scarce material used in the development of the technology?
– Is recycling being taken into account?
International justice:
– Where and in what circumstances are the raw materials mined?
– What are the consequences of the technological development for the position of poor people and marginalized groups?
Security:
– Is it possible to use the technological development for weapons or terrorists' attacks?
– What are the possible forms of misuse? Can this misuse be prevented? Is regulation necessary?
Methods:
– (science ethics)
– (issues relating to human subjects and animal testing, for example informed consent of human subjects and the decrease of the use of animal tests)
– In what way are the assumptions used in models and load scenarios validated? What data is necessary to have a better idea about the validity of the model?

Floridi, 2001; Rooksby and Wecktert, 2007). These ethical issues are important and should be taken into account in the development of technology. There are ideas about how to take these issues and human values into account during design processes, see the chapter about value sensitive design in this book and (Friedman, 1997). The mentioned ethical issues are also relevant to the goal of the Towards Ultrafast Communication project as a whole. This goal was described by the project leader as "2010: access for each individual to all layers of the communication network, everywhere, anytime" (Lenstra, 2006). However, there are four reasons why these ethical issues are not that relevant for the *individual* researchers working within the TUC project.

The first reason is that the ethical issues that are mentioned above all deal with the use of IT not the technological research that might lead to the design of new technology. For the researchers the connection between their research and the use and users is very remote. The researchers do not know whether their research will be used by users and if it is going to be used when it will be used. This is not only the case in TUC. Even though more emphasis is placed on the utilization of scientific research within the EU, a lot of technological and scientific research at universities is remote from users and the use-context. Researchers first have to provide a "proof-of-principle" meaning that the proposed component can indeed perform the desired function. Providing a proof-of-principle requires a lot of research and tests. After a proof-of-principle is provided the technology can be developed further into a prototype. If that works then the technology can be used to make products. In Vincenti's conceptualization the TUC project would be really radical design. Vincenti has introduced the concept of design type. The design type ranges from normal to radical design. In normal design the operational principle and normal configuration are the same as in previous designs (Vincenti, 1990). "Operational principle" is a term introduced by Polanyi (Polanyi, 1962). It refers to how a device works. The normal configuration is described by Vincenti as: "... the general shape and arrangement that are commonly agreed to best embody the operational principle." (Vincenti, 1990: 209). In the case of TUC the research should lead to a new operational principle. It may well be that a researcher has to decide after a few years of research that he or she is not able to get a component based on the new operational principle working properly. In such radial design there is usually an idea of the use of the product but the day-to-day research practice is focused on providing a proof-of-principle or a working prototype.

The second reason that the ethical issues mentioned above might not be very relevant for technological research such as TUC is that the research deals with components. The researchers discussed components and tests of components. They talked about algorithms, quantum dots, band gaps and photonics. The researchers and especially the PhD students did not know what was meant by ethical issues. They thought that ethical issues such as privacy, the digital divide and security were irrelevant for their research practice. The researchers admitted that these issues are important but they thought that these issues were not important at this moment of the technology development. They also doubted whether they should be the persons to deal with these issues. When looking at the daily research practice the relevance

of the ethical issues most often mentioned in IT ethics is hard to see. These ethical issues might be very relevant for the total research. However, when testing a component it is questionable whether the researcher or anyone else is able to make any meaningful claims about for example the relation between the testing of the component and the infringement of privacy caused by the use of a product that will perhaps be made containing the component.[1] There are, however, opportunities to have meaningful ethical reflections by the researchers working on early development of technologies if the ethical issues are more related to their daily research practice. The daily research practice for the PhD students and post-docs refers in TUC to the experiments that the researchers do and the way they perform and evaluate these. Senior staff will have management tasks, fund-raising and teaching tasks besides their research.

The third reason why ethical reflection by the researchers on the issues as mentioned in IT ethics is not relevant is that the researchers have very different backgrounds and work in a specific context. Their different backgrounds make it very difficult for a large group of researchers to engage in a meaningful ethical reflection and debate about the technology they are developing. Technological research is international at least in the Netherlands. At some institutes over 50% of the employed researchers are foreigners. For example at the Dutch Technological Top Institute NIMR (Netherlands Institute for Metals Research) over 60% of the employed researchers were foreigners, the same holds for the TU Delft University Institute DIMES (Technologisch Weekblad, 2007). In the Netherlands technological research is done by Chinese, East-European and former Sovjet-state citizens and some Dutch researchers. These researchers are usually employed for only two or three years. For their career development and for obtaining another contract after the current one, researchers need to publish and obtain results in their technological research. Researchers have therefore only limited time to reflect on ethical issues and some might not have a clue about what ethical reflection should entail. Some researchers might have had ethics and technology courses during their BSc or MSc studies but not all. Some PhD programs could include a course on ethics but then what about the postdoc researchers that did not do their PhD program in the Netherlands. This does not have to mean that no ethical reflection should be asked from technological researchers. Limited time, pressure to publish, different backgrounds and an insufficient background to analyse the ethical issues related to technological research are not a justification for researcher to simply ignore ethical issues. However, they pose constraints on what can be expected from technological researchers at this moment. Researchers can be asked to reflect on their research practice, but the most fruitful way is at this moment to ask researchers concrete questions that they can answers within a limited amount of time. If answering these concrete questions is seen as a too limited interpretation of the moral responsibility of technology researchers

[1] Ethical issues such as the ones mentioned in the first section can become very relevant later on in the development of a technology and at these stages researchers and engineers make decisions that could influence for example privacy. In these cases engineers and researchers should reflect on these issues. I would like to refer to the chapter about value sensitive design in this book.

and more ethical reflection is expected then ethics courses for PhD students and post-docs should be started.

A fourth reason why ethical reflection on the ethical issues most often mentioned in IT ethics by researchers might not be relevant is that the division of responsibility should be taken into account. With regard to large research projects several actors play a part, for example project leaders, professors, funding institutions, users, the government etc. Von Schomberg claims that the idea of role responsibility cannot be used any longer in the complex society in which we live (Von Schomberg, 2007). No one person has the overview of all consequences of a technological development and therefore he pleads for an ethics of knowledge policy and knowledge assessment. Citizens should be involved in the assessment and policy making. I agree with Von Schomberg that no individual researcher has the oversight over the consequences of a technological development. PhD students and postdoctoral researchers know their own research and perhaps that of some colleagues but they do not even have an oversight of the whole project. However, I think that role responsibility is still a useful concept in this complex society although it should be complemented with collective responsibility (van Gorp and Grunwald, 2009). See the chapter of Anton Vedder and Bart Custers in this book for more about the division of responsibility with regard to large technological research projects.

So my claim is that there are ethical issues that are relevant for technology development as a whole that are at the same time not that relevant for individual researchers that work in the first stages of technology development. This does not mean that these ethical issues should not be addressed in the first stages of technology development, but that one should not expect the individual researchers to address these issues. For example the reliability of information on the internet is an important ethical issue but not an issue that is relevant for the individual researchers in the TUC project. Of course the project team, together with funding agencies, politicians, ethicists and citizens could, and perhaps should, think about what all optical and faster switches would mean for the reliability of information on the internet. This is however a collective responsibility. There are, however, ethical issues that individual researchers in the first stages of technology development should address.

4.3 Technological Researchers and the Ethical Issues They Should Address

It is difficult if not impossible to make a complete checklist of ethical issues that is valid for researchers in all technological research. New research might bring forth new ethical issues that are not foreseeable. A checklist can therefore never guarantee that all ethical issues will be identified. The checklist can, however, make sure that ethical issues that are foreseeable are indeed identified. Taking the division of responsibility seriously, the checklist for the researchers should only include those topics that the researchers can really address in their research. So this checklist is only one tool to identify ethical issues and it is meant for researchers. Bearing

these limitations in mind a checklist might prove to be fruitful because it helps researchers to identify ethical issues. The researcher within TUC thought that the concept of ethical issues was vague and they did not know what it meant and how they could identify them. This list helps researchers by asking concrete questions. Below I will present a first draft of such a checklist for technological research, e.g. research including chemicals, biological material, technological equipment and computer modelling. This checklist is relevant for a lot of technological research, but specific research projects might require additional questions. For example, in medical technology development, questions related to informed consent in trials are important and should be added.

The checklist is only a tool to quickly identify ethical issues. If ethical issues are identified then a thorough ethical analysis should be made. Some researchers will be able to make a thorough ethical analysis themselves, for example because they have been taught how to do this in an ethics and technology course. Other researchers will need the help of an ethicist to make an ethical analysis.

4.3.1 Safety

The most important issues with regard to safety are toxicity of chemicals and processes and the health and safety of the researchers in their research practice. Many researchers do not know the hazards attached to the processes and substances they use during their research. Researchers usually know the catastrophic scenarios and probably know when to evacuate a lab. To illustrate this point, when I was doing my MSc thesis work in Materials science in 1999, I knew that if you smelled something that reminded you of almonds during some experiments, you should evacuate immediately because this could indicate that hydrogen cyanide was forming. If you were using hydrofluoric acid in etching then you had to wear a complete rubber suit. Both chemicals are very dangerous and every student was well aware of this. Besides these very dangerous or even lethal chemicals, there were standard lab safety instructions.

Usually researchers only know standard lab safety instructions and the hazards of a few potentially lethal processes and chemicals. Researchers, however, do not know exactly the risk and hazards of their research practice. Most researchers do not check whether the substances and processes they work with pose a risk to themselves and trust that their supervisors will warn them if necessary. This is problematic because these supervisors sometimes also lack knowledge of working conditions regulation. Moreover, some chemicals are reprotoxic which means that they influence, sperm, egg cells and or embryos negatively. Reprotoxicity is important for a group of people, namely those thinking about starting a family. Many researchers especially PhD students and postdoc researchers are thinking about starting a family. Wishing to start a family is not something you discuss regularly at work, at least not if you have a temporary contract. So researchers need to know themselves whether they work with chemicals or processes that are reprotoxic. For female researchers this might mean that they have to announce that they are pregnant much earlier than

they would have liked because they should not perform certain experiments during pregnancy.

One could say that the issues related to health and safety during research are addressed in working conditions regulation. However, in order for these regulations to be effective and protect researchers, the researchers need to know these regulations and the relevance of it with regard to their specific research. Moreover, knowledge of health and safety problems with the substances and processes researchers work with is not only relevant for the researchers themselves. If researchers use a toxic chemical in a component then this could lead to problems during production and use of this component. This is more important when the component is part of a product that is going to be mass produced than if the component is only used in a few very complicated and technological advanced devices such as parts for a particle accelerator. However, researchers can try to find a non-toxic replacement in the early stages of their research this would prevent problems later on.

4.3.2 Sustainability

The following subjects might be relevant in technological research: the ecotoxicity of chemicals and processes, the use of scarce raw materials, the amount of energy used and the possibilities for recycling. It is easier and cheaper if toxic or scarce materials are replaced at the beginning of a research project. It is more expensive if the replacement needs to be made after the proof-of-principle has been delivered because it would require getting to a new proof-of-principle based on more sustainable processes and materials.

The ecotoxicity of the chemicals and materials used in a research is relevant for the way the chemicals are handled and discarded during the research. The ecotoxicity is also important if the production of the component is taken into account. If a component is developed that should be produced in huge quantities for example in computers then the use of ecotoxic material is problematic.

The amount of energy that is consumed by products and processes is important for sustainability. Energy use was an important topic within TUC not because they wanted sustainable components and switches but because if the components and switches would produce too much heat then the combinations of thousands of these would cause overheating problems. This same problem could, however, be interpreted as a sustainability problem. Just recently some attention is given to the energy consumption of computers and internet and the total energy consumption of computers and the internet infrastructure seems to be significant on a global scale.

Some resources and raw materials are scarce, using these in mass products is not a wise idea. Researchers could try at the beginning of their research to substitute the scarce material for a not so scarce one. This point is related to recycling, some decisions made at the beginning of the development of a technology can influence the possibility to recycle products once they have been used. The points about recycling and scarce materials are only relevant when material components or products

are developed, not if computer models are made to simulate a process. The cradle to cradle ideas developed by Braungart and McDonough require that all material should be recycled, either by being reused or as nutrients for the natural environment (see www.mbdc.com).

4.3.3 International Justice

Non-governmental organizations such as Human Rights Watch have shown that the mining and harvesting of raw materials are in certain countries related to conflicts and oppression of indigenous people. Famous examples are the so-called blood-diamonds but also the illegal logging in the nineties in Cambodia, the profit of which was used by the Khmer Rouge to finance the violent conflict. Another example is the relation between the mining of Coltan in Congo and the conflicts there, I will elaborate on this example in Section 4.5.

Besides problems with the mining of raw materials, the development of new technology can change the position of poor countries and marginalized groups. New technology might benefit poor countries but it might also disadvantage the position of poor countries even further. Some technologies are developed with the benefits of poor countries in mind, for example golden rice in the nineties and the production of artemisinin in synthetic biology.[2] It is claimed that these technologies can help to solve problems like malnutrition (golden rice) and disease (malaria) in poor countries. Although these technologies can indeed help; they make people in poor countries even more dependent on western companies to provide them with seeds or medicine. Problems with IPR (Intellectual Property Rights) and the difficulty to find companies that will produce cheap generic medicines can add to this problem. So it remains a question whether people in poor countries would indeed benefit form technological developments such as the synthetic production of artemisinin.

4.3.4 Security

Intuitively there is a difference between intended use and misuse of an artefact or technology.[3] Technological developments can give rise to new forms of crime, an example is cybercrime. Besides use and misuse there is also dual-use. Dual-use means that the knowledge or technology can be used for the promotion of the good but also for attacks. According to Miller dual-use is an ethical dilemma for the researcher because the researcher him or herself might do the research in order to

---

[2] Golden rice is rice with more betacarotene, it should reduce vitamin A deficiencies. Artemisinin is a part of a combination medicine against malaria.

[3] Whether the use is indeed proper or not is a difficult question that I cannot address in this paper. I refer to a special issue dedicated to the dual nature of artefacts for elaborate ideas about function ascription and proper functions (Kroes and Meijers, 2006).

promote the good but other researchers might use the same knowledge to create terrorist weapons (Miller, 2007). Miller and Selgelid have introduced three axes in relation to the goals of the research: good/ evil; military and non-military and military for offensive and defensive purposes. Research can be categorised on these axes. The problem with dual use technology is that although research results were obtained in research with a good purpose these results can be used in research with a bad purpose (Miller and Selgelid, 2006). Some research is intended to be used for both military, either offensive or defensive, and non- military purposes, this is sometimes called dual-purpose. The researcher is usually aware of this dual-purpose. If there is no dual-purpose there might still be possibilities to use the technology in an attack (dual-use). The researcher might be able to foresee unintended offensive possibilities that the technology he or she is developing could allow for.

4.3.5 Methodology

There are also ethical issues with regard to the research methods that are used and the conduct of researchers. Science ethics addresses the professional norms that scientist should adhere to, for example norms about authorship, plagiarism and the independence of researchers. Science ethics is relevant for all scientific research not only or specifically for technological research. I will not address science ethics here but refer to (Barnbaum and Byron, 2001; Whitbecks 1995) and resources available on the online ethics centre for engineering and science (onlineethics.org). There are also extensive ethical guidelines and committees for research involving human subjects or involving animal tests. Ethics committees have guidelines and procedures to assess whether the use of human subjects or the use of animals meet all moral and methodological requirements (the online ethics centre provides an extensive list of links to ethical guidelines for human subject and animal research under the heading responsible research). Because issues concerning human subjects and animal tests are not often relevant in technological research and if they are relevant they are dealt with by ethical committees, I will only mention them here.

Making computer models of processes and products is nowadays an essential part of technological research. The models are used to predict properties of the final product, process or component. In order to simulate the properties, the process or product first needs to be modelled, e.g. a mathematical description needs to be made. After the process or product is modelled scenarios have to be made for the loading. The loading might be stress or heat or radiation etc. It is important that all assumptions that have been made can be justified. It is impossible in the ill-structured problem of developing and designing a completely new product to make a model and loading scenarios completely independent from each other. The ill-structured nature makes this impossible. According to Simon ill-structured problems are problems where:

> There is initially no definite criterion to test a proposed solution, much less a mechanizeable process to apply the criterion. The problem space is not defined in any meaningful way. (Simon, 1973: 311)

This implies that modelling a product or process and formulating the load scenarios are ways to understand the problem at hand better. During simulation the model and the load scenarios are adapted to each other in an iterative process (Van Gorp, 2007). The models combined with the scenarios produce beautiful and convincing pictures but whether or not these pictures give a good indication of the properties the product or process will eventually have depends on the validity, stability and sensitivity of the model and the scenarios.

In the following table the ethical issues are summarized in a checklist.

The ethical issues that require a broader discussion than only researchers are not included in this list. Questions related to the desirability of imaginable or probable changes in society and culture as a result of a certain technological development are very important and should be addressed. However, these questions should not be addressed by the researchers alone. Politicians, funding agencies, non-governmental organisations, philosophers of technology etc should participate in these discussions. The researchers doing the technological research should also be invited to participate in these discussions. Perhaps the researchers have relevant information about the possibilities or impossibilities of the technological development, but it is not their responsibility to make final decisions about the desirability of technological developments for the whole of society.

4.4 Information About the Questions of the Checklist

By answering the questions in Table 4.1 researchers can identify some ethical issues in their research project. There are different ways of obtaining information to answer these questions.

An internet search usually already provides some information about for example the toxicity of compounds. An internet search is the least researchers can do. A problem of information on the internet is the amount of information and that it is difficult to assess the reliability of the information.

Following Vedder and Wachbroit I will interpret reliable information as information that is properly justified (Vedder and Wachbroit, 2004). Reliable information is not necessarily true but people are justified in believing that it is true. Judging whether information is reliable then is judging whether the justification meets certain standards. Vedder and Wachbroit have introduced the distinction between content and pedigree criteria. (Vedder and Wachbroit, 2004).Content criteria are “a function of the content of the information itself” (Vedder and Wachbroit, 2004). One could think about the evidence that is provided for a claim, its consistency etc. The researchers should have the knowledge about their own subject to make a judgement about the reliability of the obtained information. The questions in Table 4.1, however, require knowledge outside of the researcher’s background and knowledge.

So, for the researchers it is probably impossible to judge whether the information they have found on the internet is reliable only using content criteria. Vedder and Wachbroit have introduced a second type of criteria called pedigree criteria. These pedigree criteria refer to "the conditions or criteria of reliability that are a function of the source of the information." (Vedder and Wachbroit, 2004). They refer especially to established credibility-conferring institutions for example academic institutions, professional societies and peer review systems. This means that the place where information was found on the internet is an important pedigree criterion. If information is found in peer-reviewed scientific journals or on the website of well-known universities, the researchers can usually rely on that information. Organisations such as ILO (International Labour Organization) and the WHO (World Health Organisation) provide information regarding the toxicity of compounds on their websites. The fact that the information is on the WHO and ILO websites and the procedures these organizations use to gather information indicate that the information is reliable. The researchers know these credibility-conferring institutions because they also publish articles in peer-reviewed journals and they know the reputation of some universities. Hence, researchers have the knowledge and skills to use both content and pedigree criteria to judge the reliability of information gathered in an internet search.

In addition to an internet search researchers can get relevant information from existing regulative frameworks. This information is available on the internet but also at health and safety departments of universities and research organisations. A regulative framework consists of

> . . .all relevant regulation, national and international legislation, technical codes and standards and rules for controlling and certifying products. A regulative framework is socially sanctioned, for example by a national or the European parliament or by organizations that approve technical codes. (Van Gorp, 2005)

EU directives are usually goal-based and not very detailed. For example machinery should be safe. The European technical standards, the so-called, EU codes, provide elaboration and operationalisation of what safe machinery is. Besides EU regulation there are national regulation and standards (in the Netherlands for example NEN and Kema standards). Professional societies define good practices that might provide relevant information. The regulative framework for working conditions consists in the Netherlands of EU directives and national regulation. Both EU directives and the national regulation include a list with dangerous compounds (89/391/EC and Arbeidsomstandighedenbesluit) and can therefore provide valuable information.

It can be difficult to imagine misuse and dual-use. Some dual-use is not too hard to imagine, for example if a researcher is working on aerosols some knowledge about these aerosols could be used in an attack with toxic chemicals as aerosols. Other use is more difficult to imagine. Literature and movies might help researchers to broaden their imagination. I do not want to claim that researchers should take all Hollywood scenarios seriously but they could think whether the writer indeed points to a possibility. A movie such as Minority Report can help people imagine

what it would mean if we could read each others minds. The question of the desirability of technologies leading to the possibility of "reading someone's mind" should not be answered by technological researchers alone but these researchers could reflect whether the technology they are developing could be used for these purposes.

4.5 Examples

To show that the information sources mentioned in the previous section can provide relevant and interesting information within a limited amount of time, I give two examples. One of the examples comes from the TUC project the other one was chosen because the material is used a lot in electronic equipment.

In a meeting of the researchers and people from participating companies and the funding organization (27th of January 2006 in Amsterdam) a researcher presented her research on quantum dots. In an internet search using the Google search engine and the search terms "quantum dots" and "health" a review article about the toxicity of quantum dots was found (Hardman, 2006). Searching using the search engine Scholar Google and the terms "quantum dots" and "toxicity" gave 1630 hits on 18th of April 2007 and 2500 two months later.[4] These articles all have different subjects and some are relevant others not. It is too time consuming to read all these articles but based on the title it is already possible to look for the most interesting articles with regard to questions about toxicity and eco toxicity. One of the articles addressed the health and safety issues with regard to semi-conductor production using a certain production process (Shenai-Khatkhate et al., 2004). In this article the workplace exposure limits (WEL) were given for a number of compounds that were also used in the TUC project. One of the compounds was aluminumoxide.

Another researcher presented during the same meeting his research on the properties of components of Erbium doped aluminumoxide. The WEL of aluminiumoxide is relevant for his research. I have also done a search using the terms "erbium doped aluminumoxide" using Scholar Google. This search led to articles about toxicity and eco toxicity of nanoparticles. The English version of Wikipedia gave the following information about Erbium. Erbium is mined from the minerals monazite, xenotime and euxenite in Norway, Sweden, USA, Australia, India Canada and Brasil (Erbium Wikipedia, 2007).[5] These minerals are usually radioactive. There is no information

[4] I use Scholar Google to show that in using a publicly available search engine, researchers can already obtain some relevant information in a very short time. The results of the search have different "pedigrees" and the researchers have to take this into account. Alternatively, the researchers can search their institutions library for information but it then depends on the collection of the library what information is collected.

[5] Again I deliberately used a publicly accessible source, in this case wikipedia, to show that every researchers should be able to find this information easily. Of course, there is a debate whether wikipedia is a reliable source of information but researchers should be able to assess the pedigree

about the scarcity or issues concerning international justice. Erbium is a rare earth element and there is very limited knowledge of the toxicity of rare earth elements. The wikipedia entry states that there are reasons to regard Erbium as moderately toxic. An internet search for a material safety data sheet for Erbium provided several data sheets, amongst other the sheets from the US Department of labor. The data sheets also indicate that a lot is unknown. At this moment there are no reasons to try to substitute Erbium doped aluminumoxide but it is probably wise to handle Erbium carefully. If incomplete data are found then it is a good idea to repeat the internet search again a few years later perhaps there is at that moment more information regarding the toxicity.

A second example is Tantalum (also Tantalium) that is used in most communication devices. Capacitors in mobile phones and computers usually contain Tantalum. I have tried to answer the questions in Table 4.1 for Tantalum. The English tantalum wikipedia entry refers to problems surrounding the mining of Tantalum in Congo (Tantalum Wikipedia, 2007). According to the wikipedia entry there is a relation between drugs use, rapes, criminality and corruption and the mining of Tantalum. Tantalum is refined from the mineral Coltan. A second search with the terms "Coltan" and "Congo" in Google provides a lot of information. Amongst others a newspaper interview with Klaus Werner and Hans Weiss about brands, war and child labour (Sioen, 2002). In this article Coltan mining is one of the examples. The money from Coltan was used for financing the civil war in Congo. Websites from a Belgian member of the EU parliament (Bart Staas), several NGO websites and the environmental website of Leonardo DiCaprio all claim that the mining of Coltan destructs the habitat of Gorilla's, moreover people mining the Coltan eat "bushmeat" (gorillameat). It can be concluded that there are problems concerning the mining of Coltan in Congo. Coltan is not scarce and it is also mined in for example Australia and Brasil. So in the case of Tantalum it is important to know where it was mined.

If a company is interested in the technology that a researcher has developed using Tantalum, then this researcher should warn the company about the problems concerning Coltan mining in Congo. Moreover, researchers should try to avoid using Tantalum from Coltan mined in Congo. It is difficult to establish where a product or material comes from. In the Werner and Weiss interview they say that although companies claimed that they did not use Tantalum from Coltan mined in Congo, they used it at least in 2002.

A search using the terms "tantalum" and "safety" leads to several material safety data sheets amongst others the safety sheet from the International Program on Chemical Safety from the WHO, European Commission, ILO, and United Nations Environment Program (UNEP). These sheets include the toxicological properties, symptoms of exposure, safe practices and the environmental effects of compounds.

---

criteria of the references used in the wikipedia entry and check these references. This gives researchers in a limited amount of time an idea of the ethical issues. If they find serious problems, the researchers should of course do other more "scientific" literature searches. Used in this way a wikipedia entry can be a good starting point for answering the questions of the checklist.

According to this data sheet Tantalum is not very toxic but it should not be inhaled and contact with eyes should be prevented. There is a risk of dust explosions and Tantalum should not be brought in contact with water.

4.6 Further Thoughts

The list mentioned in Table 4.1 is just a first draft but with the examples I have shown that these questions can help researchers in identifying ethical issues in their research projects. The list is an addition to the ethically questions raised in IT ethics and science ethics. An internet search combined with documents form relevant regulative frameworks give enough information to answer the question provisionally. This does not have to cost a lot of money or time, only internet access and perhaps an appointment with a health and safety official from the research institute.

An important question after gathering information is what researchers should do if they identify ethical issues. I do not want to claim that all research where toxic compounds are used or where international justice issues are important should be stopped immediately. It is, however, important that a researcher knows the risks of the compounds she is working with, she can change her research practice accordingly. Researchers can also inform funding agencies or companies about the identified ethical issues. Together they can decide whether it is better to change the research project or not. There are regulations, institutions and processes within society to deal with risks, this means that in instances where technological research imposes risks on people living near the research centre, researchers should inform the appropriate people and institutions. One could think of informing politicians, local government, certifying organizations, the ministry of Environment or other institutions depending on the specific risks.

References

Arbeidsomstandighedenbesluit version 25 Feb 2004, Sdu Uitgevers Den Haag

Barnbaum, D.R. and Byron M. (2001). *Research Ethics: Text and Readings*. Upper Saddle River, NJ: Prentice Hall

Erbium, Wikipedia, accessed 7 June 2007, http://en.wikipedia.org/wiki/Erbium

EU Directive 89/391/EC health and safety at work

Floridi, L. (2001) Information ethics; an environmental approach to the digital divide, *Philosophy in the Contemporary World*, 9(1), 39–45

Friedman, B. (ed) (1997) *Human Values and the Design of Computer Technology*. Cambridge:Cambridge university press

Hardman, R. (2006) A toxicological review of Quantum Dots: toxicity depends on physicochemical and environmental factors, *Environmental Health Perspectives*, 114(2), 165–172

http://cordis.europa.eu/

http://onlineethics.org, accessed 28 June 2007

Introna, L.D. (1997) Privacy and the computer: why we need privacy in the information society, *Metaphilosophy*, 28 (3), 259–275

Johnson, D. (1985) *Computer Ethics*. NJ: Prentice Hall

Kroes, P. and Meijers, A. (eds) (2006), The dual nature of artefacts (special issue), *Studies in History and Philosophy of Science Part A*, 37(1), 1–158

Lenstra, D. (2006) Lecture at workshop Methodologies for the moral evaluation of technology development 24 en 25 Maart 2006 Utrecht

Miller, S. and Selgelid, M. (2006) *Ethical and Philosophical Consideration of the Dual-use Dilemma in the Biological Sciences*. Centre for Applied Philosophy and Public Ethics, Australian National University and Charles Sturt University, Canberra, Australia

Miller, S. (2007) Dual use dilemma in the biological sciences: ethical and philosophical aspects, presentation at Expert meeting "Bioterrorism & dual use", 3TUcentre for ethics and technology, 6 March Den Haag

Nissenbaum, H. (1998) Protecting privacy in an information age: the problem of privacy in public, *Law and Philosophy*, 17, 559–596

Polanyi, M. (1962) *Personal Knowledge*. Chicago: The University of Chicago Press

Rooksby, E. and Wecktert, J. (2007) Information technology and social justice, Idea Group Inc

Shenai-Khatkhate, D.V. Goyette, R.J., Dicarlo, R.L., Jr., Dripps, G. (2004) Environment, health and safety issues for sources used in MOVE growth of compound semiconductors, *Journal of Crystal Growth,* 272

Sioen, L. (2002) Interview Klaus Werner en Hans Weiss over wereldmerken, oorlog en kinderarbeid, *De Standaard 31 mei* 2002

Simon, H. (1973) The structure of ill-structured problems, *Artificial Intelligence,* 4, 181–201

Tantalum, Wikipedia, accessed 18 April 2007, http://en.wikipedia.org/wiki/Tantalum

Tavani, H.T. (1999) Computer ethics text books: a thirty year retrospective, *ACM SIGCAS Computers and Society*, 29 (3), 26–31

Technisch weekblad, Top 30 kennis- en onderzoeksinstituten, 31 maart 2007

US Department of Labor, Material safety data sheet Erbium, accessed 7 juni 2007, www.mpc.ameslab.gov/downloads/MSDS/MSDS_Erbium.pdf

Van Gorp, A.C. (2005), *Ethical Issues in Engineering Design; Safety and Sustainability,* Delft: Simon Stevin Series in the Philosophy of Technology

Van Gorp, A. (2007) Ethical issues in engineering design; regulative frameworks for safety and sustainability, *Design Studies*, 28 (2), 117–131

Van Gorp, A. and Grunwald, A. (2009) Ethical responsibilities of engineers in design processes; Risks, regulative frameworks and societal division of labour, in *The Ethics of Risk* eds. S. Roeser and L. Asveld, 252–268

Vedder, A. and Wachbroit, R. (2004) Reliability of information on the internet: some distinctions. *Ethics and Information Technology*, 5(4), 211–215

Von Schomberg, R. (2007) From ethics of technology towards an ethics of knowledge policy & knowledge assessment, *Working document from the European Commission Services*, Jan 2007

Whitbeck, C. (1995), Trustworthy research; editorial introduction, *Science and Engineering Ethics*, 1(4), 322–328

www.goldenrice.org

www.mbdc.com

Chapter 5
The Need for a Value-Sensitive Design of Communication Infrastructures

Noëmi Manders-Huits and Jeroen van den Hoven

Abstract In this chapter, we will discuss Value-Sensitive Design as a general approach to thinking about design in the TUC – *Towards Ultrafast Communication* – project from an ethical perspective. We begin with the historical context by describing developments both in information technology and ethics that preceded Value-Sensitive Design. Both ethics and computer science demonstrate a converging development towards interest in the relation between values and technical design, respectively coined by us as "The Value Turn" in (information) technology and "The Design Turn" in ethics. After a sketch of the development and history of Value-Sensitive Design we will discuss the methodology of both Value-Sensitive Design and a related approach called "Values at Play". Their methodology consists of three modes of inquiry, respectively the conceptual, empirical, and technical mode and the discovery, translation, and verification phase. These three parts of the methodology are used to deal with concerns arising from moral theory with respect to the way technology affects moral practices in direct and indirect ways. Towards the end of the chapter we will discuss in general how Value-Sensitive Design applies to the TUC project.

Keywords Design · Values · Information technology · Ethics · Value-sensitive design · Communication infrastructures

In this chapter, we will discuss Value-Sensitive Design (VSD) as a general approach to thinking about design in the TUC – *Towards Ultrafast Communication* – project from an ethical perspective. First we give a short introduction to developments both in information technology and ethics that preceded Value-Sensitive Design. As we will see, both ethics and computer science demonstrate a converging development towards interest in the relation between values and technical design. After a sketch of the development and history of Value-Sensitive Design we will discuss the methodology of both Value-Sensitive Design and a related approach called "Values

N. Manders-Huits (✉)
Faculty of Technology, Policy, and Management, Delft University of Technology, Delft, The Netherlands
e-mail: n.l.j.l.manders-huits@tudelft.nl

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_5,

at Play". Their methodology consists of three modes of inquiry, respectively the conceptual, empirical, and technical mode and the discovery, translation, and verification phase. These three parts of the methodology are used to deal with concerns arising from moral theory with respect to the way technology affects moral practices in direct and indirect ways. Finally, we will in general outline how the VSD approach relates to the TUC project. Other contributions in this book provide detailed analyses of the ethical problems.

5.1 Developments in Information Technology: "The Value Turn"

When the computer was introduced around the middle of the twentieth century, scholarly attention was focused on the technology simpliciter. The computer was developed without too much thought about (1) the use and application in real life (2) the social, organisational and political changes it would require or it would bring about. Computers were a new and fascinating technology: solutions looking for problems.[1] Analogous to Neil Postman's "tool-using culture" (1993), where technologies are perceived only as ways to solve physical problems, the technology initially appeared to be "context-free", "context-independent" and neutral.[2]

In the seventies and eighties attention was increasingly drawn to the context of the technology, i.e. real organizations, (human) user needs and requirements, work conditions, etc. The social and behavioural sciences became involved with Information Technology (IT) in the form of Human-Computer Interaction,[3] Participatory Design[4] and Social Informatics. However, the focus of these efforts and commitments was initially mainly on functional values, such as user-friendliness and worker-safety. Social and organizational context was often taken into account merely as a way to identify potential barriers to the successful implementation of systems and in order to save money. Attention for human and social context of computing was thus determined by fear of failures in implementation and acceptance and the economic damage that can be associated with it.

In the last decade of the 20th century the successful application of information technology was increasingly seen as *dependent* on its capacity to accommodate human values. Human beings, whether in their role as employers, consumers, citizens, or patients, have moral values, moral preferences and moral ideals. In every society there are ongoing moral and public debates about liability, equality, property,

---

[1] Consider the characteristic illustration of this rather instrumentalist view of technology e.g. by the expression "guns don't kill people, people kill people".

[2] For an assessment of the embeddedness of technology in society see for instance Tiles, M., and Oberdiek, H., *Living in a technological culture: human tools and human values*. Routledge, New York, 1995.

[3] For an extensive literature on Human Computer Interaction see http://www.hcibib.org/readings.html.

[4] See for instance Kensing, F., *Methods and Practices in Participatory Design*. ITU Press, Copenhagen, 2003.

privacy, autonomy and accountability. Successful implementation is more and more construed in terms of how and to what extent values are taken into account in the design and architecture of systems. Values may even become driving factors in the development of IT instead of being an impediment in the design of information technology. We seem to have entered a third phase in the development of IT that we'd like to refer to as "The Value Turn in IT", where the needs and values of human users, as citizens, patients, are considered in their own right.

5.2 Development in Ethics: "The Design Turn"

Simultaneous to the development of the views on technology and society, a development in ethics occurred during the course of the last century. From a predominantly meta-ethical enterprise in the beginning of the 20th century, where the focus was on questions concerning the meaning of ethical terms such as "good" and "ought" and on the cognitive content and truth of moral propositions containing them, the philosophical climate changed in the sixties and ethics witnessed an "Applied Turn".[5] Moral philosophers started studying problems and practices in professions, in public policy issues and public debate. Especially in the USA, a notable development took place as philosophers gradually started to realize that philosophy could contribute to social and political debates by clarifying terms and structuring arguments, e.g. concerning the Vietnam War and civil rights, abortion, environmental issues, animal rights, and euthanasia. The focus at this point was on the application of normative ethical theory, for instance utilitarianism or Kantianism. However, in the last decade, applied ethics is taken a step further[6] as emphasis is now being placed by some authors on the *design* of institutions, infrastructure and technology, as the shaping factors in our lives and in society.

With respect to institutional design authors such as Thomas Pogge, Russell Hardin, Cass Sunstein, Robert Goodin and Dennis Thompson have taken applied ethics a step further down this path. Besides offering applied analyses, they also address the economic conditions, institutional, legal frameworks and incentive structures that constitute the moral situation. Moreover, they realize that if ethicists want to bring about real and desirable moral changes in the world and if the applied

---

[5] For a discussion on the development to applied ethics see for instance Macklin, R. "Theoretical and Applied Ethics: A Reply to the Skeptics" in David M. Rosenthal, Fadlou A. Shehadi (eds.), *Applied Ethics and Ethical Theory*, University of Utah Press, Salt Lake City, 1988, Noble, Cheryl, "Ethics and Experts," *The Hastings Center Report* (June), 7–9, 1982, and on the practical implications the proceedings of the 1992 conference of the European Society for Research In Ethics, *The Turn to Applied Ethics: Practical Consequences for Research, Education, and the Role of Ethicists in Public Debate*, David Brown, 1993.

[6] It certainly remains questionable whether meta-ethics, applied ethics, and the focus on design is of the same theoretical and moreover, justificatory level. This discussion was already initiated with respect to the applied turn (see the discussion following Noble's piece on 'ethics and experts' for example, to be found in the previous note). Our aim for this article is merely to point to the developments in and the extension of the ethics field.

analyses are to be adequately implemented, then systems, institutions, infrastructures need to be designed in accordance with moral values. This notable shift in perspective might be termed "The Design Turn in Applied Ethics" (Van den Hoven, 2008).

Although design in the work of the authors mentioned above is primarily focused on institutional design, the Design Turn clearly brings into view the design of socio-technical systems and technological artefacts. These two separate but simultaneous developments of the past decades in IT (technology – social and organizational context – human and moral value) and the other in ethics (theory – application – design) come together in the idea of Value-Sensitive Design (VSD).[7] VSD was first proposed in connection with information and communication technology and that is still its main area of application. There were several important ideas and proponents of those ideas that led up to it.

5.3 Value-Sensitive Design

As a strong proponent of private transport, famous architect and urban planner Robert Moses designed the construction of overpasses on New York parkways low, so that cars could easily access e.g. Jones Beach, while at the same preventing buses to pass under. This turned out to have severe social and political implications, following Langdon Winner (1980), as the poor and (mainly) colored population – who are largely dependent on public transport – was prevented from accessing Jones Beach. Indirectly, the overpass functioned as a border-mechanism separating the wealthy from the poor with respect to the area that lies behind. Although it is still contested whether Moses' design was consciously *intended* to have the implication of "natural" or even racial selection as it did, according to Winner it is nevertheless a clear-cut illustration of the political dimensions that artifacts may have. With his account of "The Politics of Artifacts", he was one of the first to point to the political and social ideologies, values and biases our technologies have embedded in them.

Other studies into the philosophy and sociology of technology have also revealed numerous illustrations of the fact that social and political biases and values are incorporated in technical artifacts, systems and infrastructures (see, for example, Cowan, 1985; Lansing, 1991; Latour, 1992; Mumford, 1964). The examples in these studies illustrate how technologies tend to promote certain ideologies, while obscuring others. Batya Friedman, Helen Nissenbaum, and other scholars in ethics of information technology have extended this research into questions of how information technologies specifically can carry values and contain biases (see, for example, Friedman, 1997; Moor, 1985; Nissenbaum, 2001; Tavani, 2004). The presumption here is that technology is not neutral with respect to values. Value-Sensitive Design

[7] There is a rapidly expanding literature on Values and Design or VSD; see websites of Batya Friedman (http://projects.ischool.washington.edu/vsd/), Helen Nissenbaum (http://www.nyu.edu/projects/valuesindesign/) and Jean Camp (www.designforvalues.org).

(VSD) recognizes that the design of technologies bears "directly and systematically on the realization, or suppression, of particular configurations of social, ethical, and political values" (Flanagan et al., 2008).

VSD is an approach to systems development and software engineering which was developed in the last decade of the 20th century. It was developed by Friedman and others building on insights of the human-computer interaction community (HCI) to draw attention to the social and moral dimensions of design. In VSD the focus is on incorporating human and moral values into design of (information) technology, by looking at the endeavors of design from an ethical perspective concerned with the way moral values such as freedom from bias, trust, autonomy, privacy, and justice, are facilitated or constrained. Where other research and technical communities have also been working on the value implications of computer technology, such as computer ethics, computer-supported cooperative work (CSCW) and participatory design (PD), VSD focuses *primarily* and *specifically* on addressing values of *moral* import. Other (mainly older) frameworks tend to focus more on functional and instrumental values, such as user-friendliness and worker-safety. Albeit building a user-friendly technology might have the side-effect of increasing a user's trust or sense of autonomy, in VSD incorporating moral values into the design is the primary goal instead of a by-product. Even though VSD does not commit to a specific normative framework, according to Friedman, "Value-Sensitive Design is primarily concerned with values that center [sic] on human well being, human dignity, justice, welfare, and human rights. Value-Sensitive Design connects the people who design systems and interfaces with the people who think about and understand the values of the stakeholders who are affected by the systems. Ultimately, Value-Sensitive Design requires that we broaden the goals and criteria for judging the quality of technological systems to include those that advance human values." (Friedman, 1997) It is at the same time, as pointed out by Van den Hoven (Van den Hoven, 2005: 4), "a way of doing ethics that aims at making moral values part of technological design, research and development". Notably, the aim of this methodology is to broaden the goals and criteria for evaluating technology from a moral perspective, where the moral framework that serves as starting point can vary depending on the decision made by the performers of the assessment. Self-evidently, this includes the specific definition of the scope and contents of (certain) values.

There are several Value-Sensitive Design communities that have developed in the last decade focusing attention on the way moral values are facilitated or constrained as an integral part of the conception, design, and development of technological artifacts and systems. These include Design for Values (Camp), Values at Play (Flanagan, Howe, and Nissenbaum) Value-Sensitive Design (Friedman, Kahn, and Borning) and the related methodology "Disclosive Computer Ethics" developed by Philip Brey[8] (2000, 2001). They underscore that technology has moral and political implications for those affected by the technology. Technology is not merely

[8] For a description of the research centre working on values in design see www.ethicsandtechnology.eu.

enabling, but *constitutive*; it shapes our practices and institutions in important ways, such as those of health care and transportation, Van den Hoven argues (2005: 4).

The main methodological structure used by VSD initiatives is an integrative and iterative tripartite methodology consisting of conceptual, empirical, and technical investigations (See Friedman et al., 2005 on VSD, or Flanagan et al., 2005 on VAP). Each of the conceptual, empirical, and technical investigations and analyses are carried out iteratively, mutually informing and being informed by the other investigation. These interdependencies are metaphorically described by Nissenbaum as "balls in play" (Flanagan et al., 2005), where attention to three different modes (balls) of investigation must be maintained and balanced for successful implementation. "Conscientious designers must juggle and keep in the play the results of at least three modes", i.e. the results of empirical, conceptual, and technical research (Flanagan et al., 2005).

The first "ball", the conceptual analysis, is informed by ethics and moral philosophy regarding particular value constructs relevant to the design in question. "Value-Sensitive Design provides us with the opportunity to deal with these ethical issues in a new and fresh way: by "frontloading ethics" and by means of the proactive integration of ethical reflection in the stage of design." (Van den Hoven, 2005). The (design of a) technology is evaluated in light of certain concerns informed by moral theory with respect to the way (the use of a) technology impacts our life. VSD does not promote a specific moral theory, although the available case studies refer to values that are central in modern western philosophical thought, e.g. equality and justice. Furthermore it is always the case that a particular conception of these values is used to inform design, e.g. a particular conception of the general concept of privacy could be taken as a vantage point for the design of a particular Privacy Enhancing Technology. There are different ways of proceeding in cases of value conflicts, to be decided by the responsible for the assessment of the technology at stake. Justification of value hierarchies may be found in classical ethical theories, such as utilitarianism or Kantianism; others may prefer to weigh different value interpretations and preferences as cases each treated on its own or subject to the prevailing academic philosophical discussion on the conceptualization of these values.

The second ball in play is the empirical mode of investigation. In this mode empirical data is collected to investigate whether and how the values of the conceptual mode are supported by different stakeholders of the technology. Empirical data can also be used to provide feedback in support of certain design choices made in the technical mode of investigation with respect to a particular design. The empirical data is collected using any of the available methods from the social sciences, e.g. interviews, polls, questionnaires. The inclusion of an empirical mode makes VSD a multidisciplinary approach bringing together humanities, social sciences and technology development. How to proceed in case of value conflicts or conflicts due to different aims is dependent on the moral theory on the one hand, and methods and goals of the social sciences on the other and how these two relate to each other. Moreover, it needs to be clear what the overall aim of the VSD investigation is, so that different stakes can be weighed and balanced accordingly.

Finally, the third ball, the technical analysis investigates particular technical design specifications and variables that might promote or obscure given values within the context of the technology being designed. Decisions during the design process knowingly or unknowingly determine to a large extent the moral and political implications a technology may have in practice. Any particular design enables features, opportunities and possibilities, while playing off others. In the technical analysis the focus is primarily on how technologies can support or compromise human values. Subsequently, it tries to incorporate the results of the conceptual and empirical phases into design in a proactive manner (Friedman, 1997).

Another main approach within VSD is the Values at Play (VAP) approach, which offers a similar tripartite methodological framework consisting of discovery, translation, and verification phases (Flanagan et al., 2005). Whereas Friedman's account of VSD offers a separate empirical mode and a technical mode, these parts of the process are subjected to the three phases offered by the VAP account. VAP focuses mainly on the integration of values as its main aim, as opposed to the more interdisciplinary outset offered by VSD. It starts with the discovery phase, where the goal is to identify the values that might be relevant to the design of a particular technology, including those explicit in the aspirations of the technology's designers, as well as those that only emerge when the technological design process is underway. The translation phase is the activity in which designers translate the value considerations identified in the discovery phase into the architecture and features of the technology. The final phase is verification, ensuring that the designers have successfully implemented the values identified throughout the discovery process. In both the VSD and VAP versions, these three modes of investigation are intended to form an integrative and iterative methodological framework for embodying human and moral values into the design of technology.

5.4 Towards Ultrafast Communication

Towards Ultrafast Communication (TUC) is a Dutch research project that aims at using photonics (light) succeeding the electron as basic material for conducting information in information and communication technologies (ICT). It will increase the speed of major technologies such as the Internet tremendously. Furthermore, it will improve (the speed of) access and the effective use of networks of ICT notably. Arguably, this will also lead to expected – as well as unexpected – transformations of human practices, as ICT has already shown to have a pervasive impact on society. By applying a Value-Sensitive Design perspective, awareness can be raised regarding considerations with respect to possible future consequences of using this technology and the affordances and constraints this may have in a moral respect. Accordingly, relevant human and moral values can be explicitly and proactively incorporated in design.

As a potential future building block for large-scale information infrastructure(s), which will be responsible for the highly intensive trafficking of data, the photonic cell is an interesting node to study from an ethics of technology perspective.

Typically for infrastructures is that they are not often topic for discussion when everything is functioning well. According to Bowker and Star (1999: 34), infrastructures have a tendency to disappear, what they call "infrastructural inversion": The infrastructure fades as it is embedded in structures, social arrangements and [other] technologies (Star and Rohleder, 1996 in Bowker and Star, 1999: 35). Only when the system breaks down, it becomes visible again. To study its moral and political implications, the whole system would need to be dissected. As pointed out by Bowker and Star, large-scale infrastructures systems are often "invisible, erased by their naturalization into the routines of life. Conflict and multiplicity are often buried beneath layers of obscure representation." (Bowker and Star, 1999: 47). Value-Sensitive Design however offers an account to study possible implications and effects *during* design, *before* the standardization, even before the effectuation of a technology. Software and hardware together regulate cyberspace. They are analogous to the way law regulates society in the offline world. As Lessig puts it (drawing on William Mitchell), "computer code is law" (Lessig, 1999: 6). We need to think about the way in which we want this code to regulate our communication infrastructure. How do we want our institutions to be designed? Lessig therefore argues that we need to identify the values that a space should guarantee (Lessig, 1999: 6, 109). An infrastructure should support ways to guarantee, protect, uphold and develop those values. Besides formulating explicit rules and regulations, we can also contribute to the design of basic institutions in society and therefore to society itself, by taking part in the social shaping of (information) technology. It is along these lines that we are co-responsible for institutional and infrastructural design.

Nevertheless, when designing a technology that has the potential of having transformative implications for our daily life, as TUC has, future outcomes are fundamentally complex and uncertain. It is almost impossible to predict and anticipate the future consequences and implications of a technology. Moreover, as it is taken into use, the technology is adjusted and changed in a dynamical development process; it may be applied and used for purposes other than those intended in design. The loss of control in this sense is referred to as the *Collingridge Dilemma* (Collingridge, 1980). It should be recognized that decisions made in design determine future opportunities and possibilities which new technologies will provide, while obscuring others. In a complex socio-technical system such as an information and communication infrastructure, numerous different components are interdependent and influence each other with respect to the degree of possibilities that they are able to provide. Different layers e.g. physical (optical fibers switches etc) pass on or inherit properties or shape other layers: physical – routing protocols – middle ware – application layer. Degrees of freedom at one level are constrained by design decisions on previous levels. This illustrates the complexity of creating a Value-Sensitive Design. VSD entails not only thinking through different stages of design, but also through the different layers of physical entities within a system, let alone the different operational layers within the infrastructure. Finally, it calls for the analysis of the technology in question within its socio-political context, before, during and as a result of its implementation.

For the performers of assessing the technology using VSD, it is important to identify what and whose values should be taken into account with respect to the new technology: Who are the relevant stakeholders, e.g. the engineers, the contractors, managers, companies, potential customers, etc.? A related and significant choice for the assessment of values is the issue whether one wishes to work with a predetermined normative framework, in order to justify value choices and (the outcomes of) value considerations to be encountered during design. If so, this framework should be made explicit so that it is clear what is to be expected with regard to the value commitments of the technology before, during, and after the design process. Ultimately, decisions in design and their consequences may promote or undermine specific human values. As Friedman pointed out however, values are neither solely designed into technology, nor solely conveyed by social drivers and forces; influence is exerted bi-directionally. This is what is called an *interactional* position; design and social context and the interaction between both matters (Friedman and Kahn, 2003).

The Value-Sensitive Design of (a new information technology as) TUC leads to several problems that deserve attention. Firstly, designing an entirely new communication infrastructure has transformative consequences; it may transform institutions, communication, relations, etc. in many, often unforeseen ways. Therefore it is important to consider the desirable principles and values one wishes to support and play out in design. Secondly, infrastructures are complex socio-technical systems where values are expressed and opportunities are enabled through numerous different physical and operational layers. Design decisions determine consequences for values in design at more than one spot, reaching all the way to the context of use. Thirdly, it is our moral responsibility to be aware of the values in the design of communication infrastructures because they shape –and indeed regulate – our communication and therefore an influence an important part of our society.

5.5 Conclusion

In this chapter we have explored Value-Sensitive Design as an approach for thinking about design in the TUC – *Towards Ultrafast Communication* – project from an ethics of technology perspective.

Value-Sensitive Design is a relatively new approach for becoming aware of and assessing moral values in design. By applying a three partite methodology, including a conceptual, empirical, and technical mode, or mainly focusing on the integration of conceptual considerations (VAP), moral concerns arising from moral theory are taken into account and the design of a technology is evaluated in light of the constraints and affordances it brings about in moral practices.

As we have seen there are several important aspects to keep in mind. First of all the shaping of a communication infrastructure has pervasive consequences for everyday life; it transforms institutions, communication, relations, etc. in many, often unforeseen ways. For this reason it is crucial to be aware of the principles and values one wishes to support. Second, one should be aware of how design decisions have

far reaching consequences for values in design at more than one spot, reaching all the way to the (moral) practices in context of use. Third, communication infrastructures regulate our communication and therefore an important aspect of our society. Therefore it is our moral responsibility to create a value-sensitive design that is compliant with our moral framework.

References

Bowker, G., Star, S.L. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press, 1999.

Brey, P. "Disclosive Computer Ethics." In *Readings in Cyberethics*. R.A. Spinello and H.T. Tavani (eds.), Jones and Bartlett Publishers Inc., 2001, 51–62.

Brey, P. "Method in Computer Ethics: towards a multi-level interdisciplinary approach." *Ethics and Information Technology* 2000, 125–129.

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

Cowan, R.S. "How the Refrigerator Got Its Hum." In *The Social Shaping Of Technology: How The Refrigerator Got Its Hum* D. MacKenzie and J. Wajcman (eds.), Philadelphia: Open University Press, 1985, 202–218.

Flanagan, M., Howe, D., and Nissenbaum, H. "Values at play: Design tradeoffs in socially-oriented game design." In *Conference on Human Factors in Computing Systems* 2005, 751–760.

Flanagan, M., Howe, D., and Nissenbaum, H. "Values in design: Theory and practice." In *Information Technology and Moral Philosophy* J. van den Hoven and J. Weckert (eds.), Cambridge: Cambridge University Press, 2008.

Friedman, B. *Human Values and the Design of Computer Technology* New York: Cambridge University Press, 1997.

Friedman, B. and Kahn, P. "Human values, ethics, and design." In *The Human-Computer Interaction Handbook* J.A. Jacko and A. Sears (eds.), Lawrence Erlbaum Associates, 2003, 1177–1201.

Friedman, B., Kahn, P., and Borning, A. "Value Sensitive Design And Information systems" In *Human-Computer Interaction in Management Information Systems: Foundations* P. Zhang and D. Galletta (eds.), M.E. Sharpe, Inc. 2005.

Lansing, J.S. *Priests and Programmers: Technologies of Power in the Engineered Landscape of Bali*. Princeton, NJ: Princeton University Press, 1991.

Latour, B. "What are the Missing Masses? The Sociology of a Few Mundane Artifacts." In *Shaping Technology/Building Society: Studies in Sociotechnical Change,* W.E. Bijker and J. Law (eds.), Cambridge, MA: MIT Press, 1992, 225–258.

Latour, B. *Aramis, or, the Love of Technology*. Cambridge, MA: Harvard University Press, 1996.

Lessig, L. *Code and Other Laws of Cyberspace*. New York: Basic Books, 1999.

Moor, J. "What is computer ethics?" *Metaphilosophy* 1985, 16(4), 266–275.

Mumford, L. "Authoritarian and democratic technics." *Technology and Culture*, 1964, 5(1), 1–8.

Nissenbaum, H. "How computer systems embody values." *IEEE Computer* 2001, 34(3), 118–120.

Postman, N. *Technopoly: the Surrender of Culture to Technology*. New York: Vintage Books, 1993.

Tavani, H. *Ethics and Technology: Ethical Issues in an Age of Information and Communication Technology*. Hoboken, NJ: Wiley, 2004.

Van den Hoven, J. "Design for values and values for design" *Information Age* 2005, 4, 4–7.

Van den Hoven, J. "Moral methodology and information technology." In *Handbook of Computer Ethics*. H. Tavani and K. Himma (eds.), Hoboken, NJ: Wiley, 2008.

Winner, L. "Do artifacts have politics?" *Daedalus* 1980, 109(1), 121–136.

Part II
Evaluating New Technologies: Methodological Issues

Chapter 6
The Moral Relevance of Technological Artifacts

Peter-Paul Verbeek

Abstract This chapter explores the ethics of technology in a double sense: it lays bare points of application for ethical reflection about technology development, and it analyzes the ethical dimensions of technology itself. First, the chapter addresses the question of how to conceptualize and assess the moral significance of technological artifacts. If technologies help to shape moral actions and decisions, how to understand and evaluate this moral role of material artifacts? Second, the chapter analyzes the implications of the moral significance of technology for the question of responsibility. If moral actions and decisions are the result of complex interactions between humans and technologies, how to attribute responsibility in such situations? And third, the chapter analyzes how designers can take responsibility for the moral dimensions of their designs. By integrating the methods of stakeholder analysis and mediation analysis, the chapter proposes an expanded framework for the ethics of design.

Keywords Ethics of design · Technological artifacts · Technological meditatiën · Responsibility · Stakeholder analysis

6.1 Introduction

How to do ethics of technology? Over the past decades, several answers to this question have developed, which differ radically from each other. In its early days, the ethics of technology took the form of *critique*. Rather than addressing specific ethical problems related to actual technological developments, ethical reflection on technology consisted in criticizing the phenomenon of 'Technology' itself. Classical approaches in the philosophy and ethics of technology feared the ongoing fusion of technology and culture, and aimed to protect humanity from the alienating powers of technology. They saw the technologization of society as a threat to

P.-P. Verbeek (✉)
Department of Philosophy, University of Twente, Enschede, The Netherlands
e-mail: p.p.c.c.verbeek@gw.utwente.nl

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_6,

human authenticity and to the meaningfulness of reality. People would only exist as cogs in the machine of a technologized society, reduced to the function they have in the apparatus of mass production (cf. Jaspers, 1951), while reality would only have meaning as a heap of raw material, available to the human will to power (cf. Heidegger, 1977).

Gradually, however, the field of 'ethics of technology' developed, which sought a more detailed understanding of and contact with actual technological practices and developments. Rather than placing itself outside or even over against the realm of technology, ethics now came to address actual ethical problems related to technology. Applied subfields emerged, like biomedical ethics, ethics of information technology, and ethics of nanotechnology. Moreover, ethics became more interested in the *process* of technology development. Fields like engineering ethics and ethics of design came into being, which are explicitly directed at the practice of technology development.

There are good arguments, though, to claim that the current connection between ethics and technology does not go far enough yet. Paradoxically as it may seem, many ethical approaches to technology still have too little contact with technology itself and its social and cultural role. For quite often the ethics of technology takes an equally externalist position toward technology as did the early critique of technology. At the basis of both approaches is a radical separation between technology and society. On the one hand, ethical reflection directs itself at the individual responsibility of engineers, who have to blow the whistle when they discover the existence of immoral practices, or immoral consequences of specific innovations. On the other hand, ethics focuses on the risks connected to the introduction of new technologies. Often-used case studies concern the roles of engineers in the development of the exploding Space Shuttle 'Challenger' and the Ford Pinto with a gas tank that ruptured in collisions at 25 m/h. Technologies are approached here in a merely instrumentalist way: they fulfill a function, and if they fail to do this in a morally acceptable way, the whistle should be blown. Humans are placed in one domain, technologies in the other, and the question is how to make sure that technology does not have detrimental effects in the human realm and that humans can control the technological realm.

What remains out of sight here, is the fundamental intertwinement of both domains. Technologies play an important role in human practices and experiences, in moral actions and decisions, and in the quality of our lives. When technologies are used, they inevitably help to shape the context in which they fulfill their function. They contribute to the coming about of specific relations between human beings and reality, and co-shape new practices and new ways of living. Cell phones, for instance, explicitly help new ways of communicating and interacting to come about: they create new ways of dealing with appointments (long-term planning becomes less necessary if everybody can be reached everywhere anytime), new styles of communication (including SMS), and new definitions of the separation between public and private (by making it easy to have private conversations in public). Likewise, medical technologies play an important role in defining illness and health and in making ethical decisions regarding the lives of unborn children.

Technologies like ultrasound makes it possible to detect serious congenital defects, and in doing so they translate the fetus into a possible patient and charge parents with the responsibility to make a decision regarding the life of their unborn child.

This intertwinement of humans and technology has been analyzed in terms of 'technological mediation': technologies mediate the experiences and practices of human beings. The examples of the mobile phone and ultrasound imaging are not unique – all technologies, once they are used, help to organize the relation between human beings end their environment. This phenomenon of technological mediation can be studied without reverting to classical determinism, which saw technology as an all-determining and threatening power, but also without marginalizing the role of technology to pure instrumentality. Technologies-in-use enable human beings to be involved in practices and experiences, which get shape through the specific mediating role of technologies.

Such technological mediations have at least as much ethical relevance as the prevention of disasters or attempts to find responsible ways to deal with risks. By mediating our actions and experiences, as I will explain, technologies implicitly help to answer the ethical questions of 'how to act' and 'how to live'. Technologies are morally significant; they help human beings to do ethics, by informing our moral decisions and by giving shape to our actions. In order to deal adequately with the moral relevance of technology, therefore, the ethics of technology should broaden its scope. Rather than approaching ethics and technology as belonging to two radically separated domains, the interwoven character of both should be central (cf. Latour, 1993). It is a mistake to locate ethics exclusively in the 'social' realm of the human, and technology exclusively in the 'material' realm of the nonhuman. Technologies are social too, if only because they help human beings to do ethics – and human beings belong to the material realm too, since their lives get shape in close interactions with the technologies they are using. Only by crossing the divide between both spheres, the ethical dimensions and relevance of technology can be understood. 'Ethics of technology', therefore, can be read both as an objective genitive and a subjective genitive. It can indicate the ethical activity that has technology as its object or reflection, as a form of applied ethics. But it can also indicate the ethical dimension of technology itself.

This chapter aims to bring together these two dimensions of the ethics of technology. First, I will address the question of how to conceptualize and assess the moral significance of technological artifacts. If technologies have such a large impact on our actions and decisions, and on the quality of our lives, to what extent does that imply that they can be analyzed and assessed in moral terms? Do technologies 'have' morality, and is it possible to 'read' and to evaluate this morality? Second, I will address the implications of my analysis for the ethics of design. If technologies help to shape the moral actions of human beings and the character of the decisions that inform our actions, technology designers are actually doing 'ethics by other means'. This insight implies that designers bear responsibility for the moral dimension of their designs. How can they deal with this responsibility, and how can it be integrated in existing frameworks for the ethics of design?

6.2 Capturing the Moral Dimension of Technology

6.2.1 Technological Mediation

As stated above, the moral dimension of technology can be analyzed with the help of the notion of technological mediation, as elaborated by authors like Don Ihde and Bruno Latour (cf. Verbeek, 2005). This concept indicates the ways in which technologies inevitably and often implicitly help to shape human actions and perceptions, by establishing relations between users and their environment. The central idea in the theory of mediation is that technologies should not be understood as functional *instruments*, but as active *mediators* in relations between humans and reality. This mediation can occur because most technologies-in-use are not the *terminus* of human perception and action, but rather withdraw from our attention in order to make possible a specific involvement with the world. When driving a car, for instance, our attention is not with the car itself, but with the road, other road users, the surroundings, and traffic signs. Only when someone does not yet know how to drive a car, or when the car breaks down, our attention is drawn to the technological object itself. But as soon as someone knows how to drive and if the car functions properly, he or she will be able to *embody* the car, to use a concept of Don Ihde (1990). Rather than interacting with the car, we then interact with the world *through* the car.

Technologies-in-use mediate between humans and reality, and establish specific relations between both. This phenomenon of technological mediation has two dimensions, each of them pertaining to one aspect of the relations between humans and reality. First, technologies help to shape how reality can be present to human beings, by mediating human *perception* and *interpretation*; second, technologies help to shape how humans are present in reality, by mediating human *action* and *practices*. The first dimension has a *hermeneutic* character, since it concerns meaning and interpretation; the second one a *pragmatic* one, since it concerns human actions and activities.

In Science and Technology Studies, the concept of 'script' was developed to indicate and analyze the mediating role of technologies in human action (Akrich, 1992; Latour, 1992). Just like the script of a theater play or a movie tells the actors what to do at what moment, technologies can prescribe their users how to act when they are used. Typical examples are speed bumps, that present car drivers with the choice to either slow down or to damage their shock absorbers (Latour, 1992), and the turnstiles at metro stations that require travelers to buy a ticket in order to enter the train (Achterhuis, 1995). Mediation of action usually has the shape of encouraging or inviting specific forms of action, while discouraging or inhibiting other actions. The mediating role of technologies in human experience is often indicated in terms of *technological intentionality*: technologies are 'directed' at specific aspects of reality and help to shape human perceptions and interpretations accordingly. Such intentionalities usually take shape in amplifying specific perceptions while reducing others, thus providing a specific basis for interpreting what is perceived – like ultrasound, which makes visible the fetus in terms of specific bodily dimensions

which are compared to medical norms and help us to interpret the fetus in terms of health and illness, and even in terms of a moral dilemma about abortion.

6.2.2 Material Morality

The approach of technological mediation lays bare an inherently moral dimension in technology. It shows that technologies always help to shape human actions and interpretations on the basis of which (moral) decisions are made. And if ethics is about the question of how to act, this implies that technologies help human beings to do ethics – and that they, therefore, embody some form of morality themselves. After all, when we take away the technologies, our moral questions and the answers we give to them would be entirely different.

A good example here is genetic diagnostics for hereditary forms of cancer. Genetic diagnostic testing for breast cancer, e.g., focuses on mutations in the breast cancer genes BRCA1 and BRCA2 which can predict the probability that somebody will develop this form of cancer. Carriers of such mutations (mostly women, but also men can develop breast cancer) are presented with the choice to do nothing and run a high risk to develop breast cancer; to undergo regular testing in order to discover cancer in an early stage; or to have both breasts preventively amputated (cf. Boenink, 2007)

Discovering such mutations, therefore, transforms healthy people into potential patients. Moreover, this form of genetic testing translates a congenital defect into a preventable form of suffering: by choosing to have your breasts amputated, you can prevent that you will develop breast cancer. When this technology is used, it organizes a situation of choice. This choice is complicated, because it involves a new category which is introduced by this new technology: between health and illness, genetic testing places the area of being 'not-yet-ill'. The very fact that this technology makes it possible to *know* that it is very likely that a person will become ill, added to the possibility to preventively remove organs, makes this person responsible for his or her own disease. In so doing, the technology of genetic testing creates a moral dilemma and also suggests ways to deal with this dilemma.

This example shows that medical technologies can mediate the moral choices that both medical doctors and patients make, by organizing situations of choice and by suggesting answers to this choice. The moral agency involved in making a decision about breast amputation is not exclusively with the patient and the medical doctor who advises her, but should also be located in the technologies that raise the moral dilemma in the first place, and that also implicitly suggest ways to deal with this dilemma.

From this point of view, it is reasonable to attribute to this technology a specific form of morality – without reverting to some form of animism that approaches things as spirited entities. Crucial here, though, is that the morality of technology can only exist in the context of human-technology relations. In themselves, technologies cannot be moral agents; only within the context of use practices they can become meaningful and play their mediating roles. Technologies are no fully fledged moral

agents in the sense that they can make moral decisions themselves. Moral agency requires intentionality and some form of autonomy – properties which artifacts do not possess. But when technologies are used, they do take part in the moral agency that eventually comes about, because they profoundly help to shape moral decisions and actions. Human intentionality is co-shaped by these technologies then, and their autonomy is technologically mediated. The relevant question, therefore, is not if artifacts can be moral agents, but how to understand moral agency when human actions are technologically mediated.

If ethics holds on to an approach in which moral agency requires a high degree of autonomy and intentionality, the ethics of technology will ultimately become an impossible enterprise. After all, this would imply that technologies can only be neutral means in the hands of people, because artifacts cannot have the ability to act morally, and because technologically mediated action does not stem from the autonomy that is required for moral agency. And this would in fact imply that we would need to ignore the phenomenon of technological mediation and its moral implications altogether. Conversely, an ethical theory that wants to take technological mediation seriously cannot discard the notions of autonomy and intentionality, since specific forms of autonomy and intentionality are needed to understand ethics, and to maintain the concept of moral responsibility.

In order to further address this moral significance of technological artifacts, I will first elaborate some implications of the moral relevance of technology for the notion of *responsibility*, which is so central in ethics. After that, I will explore what this 'material dimension' of ethics can imply for the practice of doing ethics, by bringing in the concept of technological mediation in the ethics of technology design.

6.3 Responsibility of Humans and Nonhumans

The distributed character of moral agency over humans and nonhuman entities has important implications for our understanding of questions of responsibility. Because of the important contribution of technologies to our practices and experiences, the question rises to what extent humans can be held responsible for actions that were induced by technologies. Does somebody act responsibly if he or she keeps to the speed limit because a speed limiter enforces this? And who is to be held responsible when an automatic face recognition system in security camera's abusively identifies a person as suspect – which actually appears to happen more often for people with a dark skin and for older people, because the software in these systems is tuned to light contrast on a young, white skin? (cf. Introna, 2005)

In order to deal with the complexity in attributing responsibility that arises when taking into account the moral relevance of technological artifacts, it is necessary to first make an elementary distinction between two kinds of responsibility involved here: causal responsibility and moral responsibility. In a causal sense, someone is responsible when he or she is the *cause* of some event or state of affairs. But this can also be the case without a possibility to keep this person responsible in a *moral*

sense. The event or state of affairs can, for instance, be caused accidentally, or under pressure. Only when somebody acts purposively and in freedom, he or she can be held morally responsible for his or her actions. And this freedom and intentionality are precisely the two elements of human agency that appeared to be so complicated in the case of technologically mediated action.

By their influence on human action – or: by their contribution to causal responsibility – technologies also contribute to the moral responsibility of human being for the actions that come about in interaction with this technology. But this does not imply that technologies should be held morally accountable for their mediating roles in human behavior – just like it does not make sense to consider technologies fully fledged moral agents in the way human beings are moral agents.[1] Yet, this does not take away the fact that technologies play a more-than-causal role here: by mediating human interpretations and actions, they actively co-shape moral responsibility. Moral decisions regarding the preventive amputation of breasts, for instance, are not simply 'causally influenced' by technologies of genetic testing. Rather, the moral questions themselves and the various possibilities to answer them are co-shaped by these technologies; it is the amalgam of humans and technologies that acts morally here and that bears moral responsibility.

The fact that responsibility needs to be distributed over humans and technologies, however, does not mean that there are no points of application available to deal adequately with the question of responsibility in technology design and use. In order to think adequately about responsibility in relation to technologies, we need to pay separate attention to the human and the technological share in the eventually resulting decisions and practices without losing out of sight the intricate connections between both. Both technologies themselves and their designers play an important role in the actions and decisions of technology users, and therefore we need to address them both.

6.3.1 Responsibility of Designers

Implicitly, the theory of technological mediation reveals an inherently moral dimension in technology design. Showing that technologies always help to shape human actions and interpretations on the basis of which (moral) decisions are made, has important implications for our understanding of the ethical roles of both technological artifacts and their designers. If ethics is about the question of how to act, and designers help to shape how technologies mediate action, designing should be considered 'ethics by other means'. Every technological artifact that is used will mediate human actions, and every act of design therefore helps to constitute specific moral practices.

Usually, this 'material ethics' of technology design happens in an implicit way. Engineers design new technologies with specific functionalities in mind, without

[1] Contrary to the approach Aaron Smith takes in his humorous paper 'Do You Believe in Ethics? Latour and Ihde in the Trenches of the Sciences Wars' (Smith, 2003).

explicitly aiming to influence the actions and decisions of users. The question, therefore, is how considerations regarding the mediating role that the technology-in-design will eventually play in society could be explicitly integrated in the design process. Two forms of designer responsibility can be distinguished here. First, designers can *anticipate* the impact, side-effects and mediating roles of the technology they are designing. On the basis of such anticipations, they could adapt the original design, or refrain from the design at all. Second, designers can also take a more radical step and deliberately design technologies in terms of their mediating roles. In that case, they explicitly design behavior-influencing or 'moralizing' technologies: designers then inscribe desirable mediating effects in technologies.

6.3.2 Moralizing Technology

The latter direction was taken by the Dutch philosopher Hans Achterhuis (1995, 1998), who translated Latour's analysis of scripts into a plea for an explicit 'moralization of technology.' Instead of only moralizing other *people* ('do not shower too long'; 'buy a ticket before you enter the subway'), humans should also moralize their *material environment*. To a water-saving showerhead the task could be delegated to see to it that not too much water is used when showering, and to a tourniquet the task to make sure that only people with a ticket can enter the train.

Achterhuis' plea for a moralization of technology received severe criticism (cf. Achterhuis, 1998, 28–31). In the debate that arose around this issue in The Netherlands, two types of arguments were brought in against his ideas. Firstly, human freedom was thought to be attacked when human actions are explicitly and consciously steered with the help of technology. This reduction of human freedom was even perceived as a threat to human dignity; if human actions are not a result from deliberate decisions but from steering technologies, people were thought to be deprived from what makes them human. Moreover, if they are not acting in freedom, their actions cannot be called 'moral.' Human beings then simply show a type of behavior that was desired by the designers of the technology, instead of explicitly choosing to act this way. Secondly, Achterhuis was accused to attack the democratic principles of our society, because his plea for developing behavior-steering technology was considered an implicit propagation of technocracy. When moral issues are solved by the technological activities of designers instead of democratic activities of politicians, these critics hold, not humans but technology will be in control.

These arguments can be countered, though. Anticipating the mediating role of technologies during the design process – either to assess possible undesired forms of mediation or to explicitly 'moralize' technologies – does not need to be as immoral as it might seem to be. First of all, human dignity is not necessarily attacked when limitations of freedom occur. Our legal constitution implies a major limitation of freedom, after all, but this does not make it a threat to our dignity. Human behavior is determined in many ways, and human freedom is limited in many ways too. Few people will protest against the legal prohibition of murder, so why protest to

the material inhibition imposed by a speed bump to drive too fast at places where children are often playing on the pavement?

Secondly, the analysis of technological mediation above shows that technologies *always* help to shape human actions. Seen from this perspective, paying explicit attention to the mediating role of technologies should be seen as simply taking the responsibility that the analysis of technological mediation implies. When technologies are always influencing human actions, we had better try and give this influence a desirable and morally justifiable form. Moreover, the 'moralizing' role of technologies does not necessarily have the form of exerting *force* on human beings to act in specific ways. Technologies can also *persuade* or *seduce* people to do certain things; they can invite specific actions without forcefully exacting them.

These counterarguments, however, do not take away the anxiety that a technocracy would come about when technologies are explicitly moralized. It might be true that technologies do not differ from laws in limiting human freedom, but laws come about in a democratic way, and the moralization of technology does not. Yet, this does not justify the conclusion that it is better to refrain from paying explicit attention to technological mediation during the design process. If technologies are not moralized explicitly, after all, the responsibility for technological mediation is left to the designers only. Precisely this would amount to form of technocracy. A better conclusion would be that it is important to find democratic ways to develop desirable forms of 'moralizing technology'. In the following, I will elaborate a way to do this – as an important route to take seriously the moral charge of technological artifacts in the ethics of technology.

6.4 Designing Mediations

6.4.1 The Complexity of Anticipating Mediations

Before discussing the *moral* aspects of the deliberate 'moralization' of technology, it is important to first discuss the *practical* difficulties involved here. For designing mediation is not as easy as it might seem to be. 'Building in' a specific form of mediation in a technological artifact is a complex task, since there is no univocal relationship between the activities of designers and the eventual mediating role of the technologies they are designing. Technological mediations comes about in a complex interplay between technologies themselves and the activities of users and designers. Technologies are 'multistable,' as Don Ihde calls it; they need to be interpreted and appropriated by users in order to be used, and different forms of appropriation result in different technologies. The typewriter, for instance, was originally designed, not as a piece of office equipment but as a device for helping visually impaired people to write (cf. Ihde, 1993).

Technological mediations, therefore, come into being only within specific contexts of use, rather than having a fixed existence which stems directly from the activities of their designers. This makes it difficult to fully predict the ways in which

technologies will influence human actions. First of all, technologies can be used in unforeseen ways, and therefore have an unforeseen influence on human actions. The energy-saving light bulb is a good example here, having actually resulted in an increased energy consumption since such bulbs often appear to be used in places previously left unlit, such as in the garden or on the façade, thereby canceling out their economizing effect (Steg, 1999; Weegink, 1996). Second, unintentional and unexpected forms of mediation can also arise when technologies do get used in the way their designers intended, but meet unforeseen use practices. A good example here is the revolving door, which keeps out not only cold air but also wheelchair users.

In short, designers play a seminal role in the coming about of technological mediations, but not the only role. The mediating role of technologies also depends on the users, who interpret and appropriate technologies, and on emergent characteristics of the technologies themselves, which can evoke unanticipated mediations. Designers, therefore, cannot simply 'inscribe' a desired form of morality into an artifact, but need to align their agency with the agency of users, and of the technologies themselves. The Fig. 6.1 draws together these factors in the coming about of mediation.

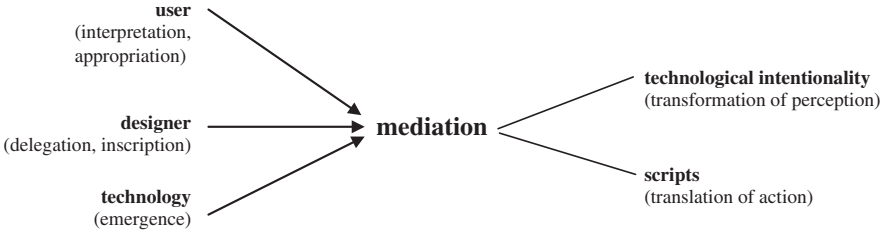


Fig. 6.1 Sources of mediation

6.4.2 Mediation Analysis and Stakeholder Analysis

This complexity, however, does not imply that it is impossible to 'inscribe' morality into a technological artifact. Even though it will never be possible to fully predict the impact of any designed technology, designers can still formulate product specifications not only on the basis of the desired functionality of the product, but also on the basis of an *informed prediction* of its future mediating role and a moral assessment of this role. Such a prediction could be informed by a 'mediation analysis:' an analysis of the future role of the technology-in-design in human actions and experiences, which moves beyond merely addressing its functionality. The most important 'instrument' for performing such analyses, however trivial it may sound, is the designer's imagination. Imagination can establish a link between the context of design and the future context of use. When designers attempt to imagine how

the technology they are designing might mediate the actions and perceptions of its users, they could feed back these anticipations into the design process.

In order to design 'moralizing technologies' in a morally responsible way, designers would consequently need to integrate such a mediation analysis in *moral* reflection about the quality of their designs. A common method for such an applied form of moral reflection, also in the ethics of technology, is the method of *stakeholder analysis*. Applying this method to technological mediation is one important way to take seriously the moral charge of artifacts in the ethics of technology. The aim of stakeholder analysis is to lay bare all moral arguments relevant for a specific ethical problem, by making an inventory of all stakeholders involved and of all arguments that are relevant from their point of view. Some stakeholders, for instance, might suffer from specific negative consequences of using the designed technology. And toward other stakeholders we might have the moral duty to introduce specific forms of mediation, e.g. when they help to save lives, like lane changing assistants in cars, which produce an alarm on unsafe attempts to overtake a vehicle. In order to carry out a stakeholder analysis regarding the morality of technology, three points of application for moral reflection should be distinguished: (a) the intended mediations that were deliberately inscribed in the technology; (b) the forms of mediation involved; and (c) the eventual outcomes of the technological mediations (cf. Berdichewsky and Neuenschwander 1999, p. 54).

Intended mediations (a), of course, are only involved when designers explicitly aim to influence user behavior with their design, rather than merely delivering a functional commodity. The *forms* of mediation (b) that are used can differ radically. In some cases, technologies actually *force* their users to act in specific ways, like a speed bump on a road or a tourniquet at a metro station. But in other cases, technologies rather *persuade* their users, like an econometer in a car which gives feedback on the fuel efficiency of one's driving style. And other technologies *seduce* their users, in non-cognitive ways, to perform or refrain from specific actions – like the designs of *Eternally Yours*, an ecological Dutch industrial designers association which aimed to stimulate product longevity by designing products that do not get discarded prematurely, but are cherished by their owners. The eventual *outcomes* (c) of the technological mediations are the actions and decisions that eventually get shape – and these can, again, differ radically from the originally intended mediations.

At all of these three levels, stakeholder analysis can be a fruitful method to lay bare the specific moral relevance of the technology in design. The intended influence of the technology on the behavior of users needs to be morally defendable; the specific forms of mediation that are used to evoke this influence should be proportional and acceptable; and the eventual results of all these efforts should be justifiable. Together, these elements determine the moral quality of the activities of the designer and of the technology in design. Expanding the method of stakeholder analysis along these lines makes it possible to move beyond the predominant focus on risk analysis and whistle blowing in engineering ethics and the ethics of technology. Also, it opens the way to give the impact of technologies on the qualities of our lives and their implicit morality a more explicit role in the moral reflection of designers.

6.4.3 Democratizing Technology Design

This integration of mediation analysis and stakeholder analysis, however, leaves one important form of criticism to the moralization of technology out of scope, and that is the alleged threat to the democratic quality of society. After all, such technologies influence human behavior and install visions of the good life while users are not always aware of this. And in our liberal democracy, the freedom of the individual is a very important value, while finding answers to the question of the good life explicitly belong to the private sphere rather than the public space. Too much interference of technology in our daily lives, can be a direct threat to democracy.

This threat is no science fiction. The growing attention of companies like Philips for what is called 'Persuasive Technology', for instance, might well result in an increasing number of devices that aim to persuade people to change their behavior. The recently developed 'persuasive mirror', e.g., aims to persuade its users to adopt a healthier lifestyle by presenting them with an image of how they will look in the future if they would stick to their current pattern of living (Knight, 2005). If the state would enforce such a healthier lifestyle by law, this would cause a lot of consternation, while technologies like the persuasive mirror can introduce similar effects 'through the backdoor'.

This explicit moralization of technology, again, is not wrong or undesirable per se. But it needs a more democratic structure. For this reason, it is important to develop democratic procedures for both the evaluation and the design of 'moralized technology'. A highly interesting point of application for such a democratization of design processes is formed by the method of Constructive Technology Assessment (CTA; cf. Schot, 1992; Rip et al., 1995). CTA establishes a link between the contexts of design and use by engaging all relevant stakeholders in the very process of technology design. CTA is based on an evolutionary view of technology development. It approaches the process of technology development as a generation of 'variations' that are subsequently exposed to a 'selection environment,' like the market and government regulations. In such a selection environment, only the 'fittest' variations will survive. There is an important difference between the generation of technologies and the generation of biological species, though. Contrary to biological evolution, in technology development there is a connection or 'nexus' between variation and selection. After all, designers can anticipate the selection environment when they are designing technologies, in order to prevent that much effort is put in developing technologies which will not be accepted by consumers or by government regulations.

CTA is a method to employ this nexus in a systematical way, by feeding assessments of the technology-in-design by all relevant actors – like users, pressure groups, designers, companies et cetera – back into the design process. It does so by organizing meetings of all relevant actors in which the aim is to reach consensus about the design of the technology that is 'constructively assessed.' This form of technology assessment is called 'constructive' because it does not assess technologies *after* they have been developed, but *during* their development, so that these assessments can be used to modify the original design.

CTA can be seen as a democratization of the designing process. When a CTA design methodology is followed, not only designers determine what a technology will look like, but all relevant social actors. Following this method, therefore, could take away the fear for technocracy that was discussed above; it opens a space for deliberative democracy in processes of technology design. And this is especially relevant when the design has explicit 'moralizing' or 'behavior-influencing' aspects. Seen from the perspective of technological mediation, however, CTA also has limitations that need to be overcome. CTA primarily focuses on *human* actors, and pays too little attention to the actively mediating role of the *nonhuman* actor that is at the center of all activity: the technology-in-design. CTA claims to open the black box of technology by analyzing the complex dynamics of technology development. It is based on the constructivist notion that technologies are not 'given,' but the outcome of a process in which many actors are involved. Other interactions between the actors might have resulted in a different technology. But analyzing the dynamics of *technology development* only opens the black box of technology half way. It reveals how technologies emerge from their *design context*, but their role in *use contexts* remains blackboxed.

Therefore, organizing a democratic, domination-free discussion between all relevant actors is not enough to lay bare all relevant aspects of the technology in question. The mediating role of the technology-in-design is likely to remain implicit during the entire CTA process, if it is not put explicitly and systematically on the agenda. Whether technologies be designed in merely functional terms or in terms of explicit 'moralization', the ways in which they will perform their functionality and moralization will always involve a process of mediation which needs the designer's attention. For this reason, participants in the CTA process should not only be invited to integrate assessments of users and pressure groups in product specifications, but also to anticipate possible mediating roles of the technology-in-design.

Creating space for all relevant stakeholders to anticipate the possible mediating role of the technology-in-design enhances the chance that as many mediating roles as possible are taken into account and explicitly assessed. To be sure, this augmentation of the CTA methodology does not guarantee that all mediating roles of the technology in design will be predicted. It does create a connection between the 'inscriptions' within the context of design and the 'interpretations' or 'appropriations' within the context of use, but this connection cannot possibly cover 'emergent' mediating roles of the technology. Yet, the integration of CTA and mediation theory might be a fruitful way to give shape to the responsibility of designers implied in the phenomenon of technical mediation.

6.5 Conclusion

The ethics of technology needs to take seriously the moral relevance of technological artifacts. In order to do justice to the profound role of technology in society and in people's everyday lives, technologies need to be approached as morally relevant

entities, rather than mere instruments in the hands of moral human beings. Technologies help to shape human actions and decisions by mediating our interpretations of the world and the practices we are involved in. In this way, they contribute importantly to moral actions. For that reason, the phenomenon of technological mediation deserves a central place in the ethics of technology, in addition to the predominant focus on the risks caused by technologies and the whistle blowing that is needed on the part of designers and engineers.

Approaching technologies as morally relevant entities has important implications for our understanding of the notions of moral agency and responsibility. By approaching agency and responsibility as phenomena that are distributed over human and nonhuman elements, ethical theory can do justice to the hybrid character that many actions and practices have taken on. Moreover, this expansion of ethics to the realm of materiality also broadens the locus of ethical activity: it moves from the realm text to that of materiality and design. If ethics is about the question of how to act, and technologies help to answer this question, technology design is 'ethics by other means'. Designers cannot but help to shape human actions and experiences via the technologies they are designing. Therefore, design processes should be equipped with the means to do this in a desirable, morally justifiable and democratic way. The methods of stakeholder analysis and Constructive Technology Assessment, augmented with insights from the theory of technological mediation, offer a fruitful basis for doing this – and to 'make things better' in a truly moral sense.

Acknowledgment This article incorporates some thoroughly reworked fragments from my earlier papers 'Materializing Morality – design ethics and technological mediation' (*Science, Technology and Human Values* 31 – 3, May 2006) and 'Morality in Design' (in: Pieter E. Vermaas e.a (eds.), *Philosophy and Design: from Engineering to Architecture*. Dordrecht: Springer, 2008).

References

Achterhuis, H. (1995), 'De moralisering van de apparaten'. In: *Socialisme en Democratie* 52 nr. 1, 3–12.

Achterhuis, H. (1998), *De erfenis van de utopie*. Amsterdam: Ambo.

Akrich, M. (1992), 'The de-scription of technical objects'. In: W.E. Bijker and J. Law (eds.), *Shaping Technology / Building Society*. Cambridge: MIT Press, 205–224.

Berdichewsky, D. and E. Neuenschwander (1999), 'Toward an ethics of persuasive technology'. Communications of the ACM 42: 5, pp. 51–58.

Boenink, M. (2007). 'Genetic diagnostics for hereditary breast cancer. Displacement of uncertainty and responsibility.' In: G. de Vries and K. Horstman (eds.), *Genetics form the Laboratory to Society*. Palgrave/Macmillan.

Heidegger, M. (1977), 'The question concerning technology'. In: *The Question Concerning Technology and Other Essays* (trans. W. Lovitt). New York: Harper & Row.

Ihde, D. (1990), *Technology and the Lifeworld.* Bloomington/Minneapolis: Indiana University Press (The Indiana Series in the Philosophy of Technology).

Ihde, D. (1993),*Postphenomenology*. Evanston, Il.: Northwestern University Press.

Introna, L. (2005), *Disclosive ethics and information technology: disclosing facial recognition systems*. Ethics and Information Technology 7: 75–86.

Jaspers, K. (1951), *Man in the Modern Age (trans.* E. Paul and C. Paul). London: Routledge & Kegan Paul.

Knight, W. (2005), 'Mirror that reflects your future self'. *New* Scientist, issue 2485, 05 February 2005, p. 23.

Latour, B. (1992), 'Where are the missing masses? – The Sociology of a Few Mundane Artifacts'. In: W.E. Bijker and J. Law (eds.), *Shaping Technology / Building Society.* Cambridge: MIT Press.

Latour, B. (1993), *We have never been modern* (trans. C. Porter). Cambridge: Harvard University Press. (translation of: *Nous n'avons jamais été modernes*, Parijs: La Découverte, 1991).

Rip, A., T. Misa and J. Schot (eds.) (1995), *Managing Technology in Society – the Approach of Constructive Technology Assessment*. London: Pinter.

Schot, J. (1992), Constructive Technology Assessment and Technology Dynamics: The Case of Clean Technologies. *Science, Technology and Human Values* 17-1: 36–56.

Smith, A. (2003), 'Do You Believe in Ethics? Latour and Ihde in the Trenches of the Sciences Wars'. In: D. Ihde and E. Selinger (eds.), *Chasing Technoscience: Matrix for Materiality*. Bloomington & Indianapolis: Indiana University Press.

Steg, L. (1999). *Verspilde energie? Wat doen en laten Nederlanders voor het milieu.* The Hague: Sociaal en Cultureel Planbureau (SCP Cahier no. 156).

Verbeek, P.P. (2005). *What Things Do – Philosophical Reflections on Technology, Agency, and Design*. Penn State University Press.

Weegink, R.J. (1996). *Basisonderzoek elektriciteitsverbruik kleinverbruikers BEK'95*. Arnhem: EnergieNed.

Chapter 7
Interdisciplinarity, Applied Ethics and Social Science

Niels Nijsingh and Marcus Düwell

Abstract The bonds between applied ethics and social science seem to become increasingly tight. This does not only manifest itself by social scientists and ethicists working together, but also by an increasing attention, from both sides, to the very nature of their cooperation. The debate on this topic has been ongoing for more than twenty years, but in the last years the discussion became more intense. However, one can doubt whether up until now the relevant questions have been sufficiently distinguished. Particularly the notion "empirical ethics" creates more confusion than clarification with regard to the different conceptual and methodological dimensions. The methodological debates show a wide spectrum of topics. This paper aims to make an inventory of the questions raised in the debate on the role of empirical research in (applied) ethics. We argue that there are at least five distinct groups of questions to be found in this debate, concerning: (1) the fact-value gap, (2) the notions of 'discipline' and 'interdisciplinary', (3) the nature of both ethics and social sciences, (4) the type of empirical research relevant for ethics, (5) the embeddedness of the individual researcher. The paper concludes with some suggestions for further debate.

Keywords Social science · Ethics · Fact-value gap · Interdisciplinarity · Empirical ethics

The bonds between applied ethics and social science seem to become increasingly tight. This does not only manifest itself by social scientists and ethicists working together, but also by an increasing attention, from both sides, to the very nature of their cooperation.

Many ethicists are currently involved one way or another in social science research (public forums, opinion polls, focus group studies, in-depth interviews, narrative analysis etc.). This prompts in a very straightforward way the question how

N. Nijsingh (✉)
Ethics Institute, Utrecht University, Utrecht, The Netherlands
e-mail: n.m.nijsingh@uu.nl

This paper is written in the context of the EU Project 'Challenge of Biomedicine', Contract No. SAS6-CT-2003-510238, funded in the 6th Framework-programme of the European Commission.

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_7,

ethics and social science should relate to each other. The debate on this topic has been ongoing for more than twenty years, but in the last years the discussion became more intense.[1] Moreover, recent discussions on the ethics of technology and the intractability of public debate on complex technological developments surrounded by uncertainty[2] have increased the necessity of discussing the relation between the different disciplines in a clear manner. However, one can doubt whether in this debate up until now the relevant questions have been sufficiently distinguished. Particularly the notion 'empirical ethics' creates more confusion than clarification of the different conceptual and methodological dimensions. The methodological debates show a wide spectrum of topics. This paper aims to make an inventory of the questions raised in the debate on the role of empirical research in (applied) ethics. We argue that there are at least five distinct groups of questions to be found in this debate, concerning: (1) the fact-value gap, (2) the notions of 'discipline' and 'interdisciplinary', (3) the nature of both ethics and social sciences, (4) the type of empirical research relevant for ethics, (5) the embeddedness of the individual researcher. The paper will conclude with some suggestions for further debate. It should be noted that our ambitions here are modest. We mean to do no more, but also no less, than to raise and distinguish some issues that we think need to be addressed if the debate is to progress any further. However, we also need to add that, although our ambitions are modest, we are by no means neutral in the different discussions, nor will we pretend to be. In fact, we believe that precisely what is lacking in some problematic accounts in this debate is sufficient clarity on which questions are being answered.

7.1 Fact-Value; Descriptive-Prescriptive

The first cluster of questions that we find in the debate on empirical ethics contains questions on the 'fact-value-distinction', the naturalistic fallacy, the distinction between the prescriptive and the descriptive sciences and the 'Is-Ought-gap'.[3] A fundamental distinction is made between on the one hand *normative ethics* as a prescriptive discipline and *descriptive and empirical research* on moral opinions, attitudes and beliefs on the other. While *descriptive disciplines* want to generate valid knowledge on whether or not people have specific opinions, attitudes or beliefs (including those on moral questions), *normative ethics* want to prove whether those

[1] Early predecessors being for example Fox and Swazey (1984) and Weisz (1990). The debate received an impulse by the theme issue of *Medicine Health Care and Philosophy* in 2004.

[2] See Bovenkerk (2006) and Sollie (Chapter 10, this volume).

[3] Although these terms are generally mentioned in the same breath, they strictly speaking do not mean the same. The fact/ value distinction is a distinction between different types of statements. The naturalistic fallacy is a particular type of fallacy, which tries to base the truth of moral statements on non-moral ones, thereby reducing moral properties to non-moral properties. The is-ought gap is concerned with the idea that it is not possible to derive a value statement from a factual statement. The distinction between prescriptive and descriptive generally refers to a difference in method. It does seem to be true that the last three presuppose the first (i.e. the fact/ value distinction).

opinions, attitudes or beliefs are valid, justified or appropriate. This distinction taps in directly on the meta-ethical assumption that we should and can distinguish between facts on the one hand and values on the other. If it is true that values form a separate category, and that they cannot be derived – as Hume famously claimed – from facts, then we have reason to be wary of empirical data in the ethical enterprise.

However, things are not that simple. It does not follow from a gap between 'is' and 'ought' that we can form moral judgements without reference to facts at all. On the contrary, in order to make a value judgment we first need to know some factual aspects of what it is we want to value. This includes e.g. biological or technical knowledge, but for most research in applied ethics some knowledge about 'social facts' is necessary as well. So claiming that you cannot derive an 'Ought' from an 'Is' doesn't make it superfluous to know something about the facts.

Furthermore, the fact-value dichotomy is not uncontroversial. There are several ways of framing the distinction. Some philosophers argue that facts and values have a distinct ontological status, others hold that prescriptive claims have to be justified in different way than descriptive ones. Yet others doubt whether we can meaningfully apply the distinction at all.[4] Therefore the distinction between the prescriptive and the descriptive is not based on such a solid ground as it may seem at first sight. The battle on the dichotomy between fact and value and the relationship between descriptive and prescriptive statements and theories has not been decided yet, and it seems that this debate gives rise to a number of highly controversial meta-ethical issues on the status and meaning of moral claims.[5] This debate can, however, not be answered in the domain of applied ethics, but has to be decided on the basis of valid meta-ethical argumentation. In this context it has been claimed that since the distinction between fact and value is blurred, the distinction between the prescriptive and the descriptive sciences falters also.[6] This, however, is not obvious. Even if we could work on the assumption that the fact-value gap is blurred, it is not clear what conclusions we can derive from this. It *could* be argued that it follows that moral claims cannot be ethically justified and that therefore the relation between ethics and social sciences is irrelevant. But this requires several controversial steps and assumptions on, for example, the methodological status of 'social facts' and the relation between the different disciplines and their objects.

7.2 'Discipline' and 'Interdisciplinarity'

The notion of interdisciplinarity plays a central role in these debates as well. A need for clarification of the relationship between different disciplines simply occurs because in applied ethics social scientists and ethicists are working together. But

[4] See Putnam (2002).
[5] See Miller (2003).
[6] Molewijk et al. (2003).

how is a 'discipline' defined? How we understand 'interdisciplinarity' will depend on the answer to this question.[7] Disciplines may be defined by reference to a variety of features, such as research questions, aims, methods, subjects, or any combination of these. It is at least clear that these features will have to relate to each other in an appropriate way. If our research question is 'How does the Dutch public judge developments in genomics?' there will be little point in trying to derive the conclusion by way of utilitarian calculus. Similarly, it will not do to interview lay people when we want to know something about the risk of accidents occurring in a nuclear power plant. In other words, the first question we have to ask ourselves when we think of the relation between different disciplines, is how the disciplines are defined, and how the abovementioned features interrelate. This might turn out to be not such a simple task.

Furthermore, we may ask what the notion of *inter*disciplinarity could consist in. One possibility is that of the (more or less complete) integration of two or more disciplines.[8] In this context the term 'transdisciplinarity' has been coined.[9] Transdisciplinarity, in contrast to interdisciplinarity, requires the disciplines to merge and thus to form a new discipline. For ethics and social science this results in a model of 'integrated empirical ethics' with 'its own premises, theories, topics and methods'.[10] This however raises the question how such a 'transdiscipline' can be conceptualized. 'Integrated empirical ethics' apparently transcends the perspective of philosophical ethics and social sciences in a discipline on a higher level. But this makes it unclear how we should describe the resulting discipline, if not as an amalgam of two. To our opinion this is problematic, because it raises the question from what perspective we can evaluate such a 'super-discipline'. What are the criteria for success? Of course, the defender of such an approach may take a modest stance, and argue that only specific elements from the different disciplines are merged, but then it becomes questionable whether what we are talking about is the merger of *disciplines*, rather than a sensitiveness from *within* different fields to new or deviant perspectives. More generally, it raises the question what the added value is of describing a certain research activity as a merger, rather than as something that makes sense to do in its own right.

A radical opposite of the transdisciplinarity approach is a very minimal conception of interdisciplinarity: the exchange of results, or 'raw data'. On this account, social scientists deliver a result (for example 'the Dutch public tends to think negatively about the role of genomics in society'), which the ethicists can use afterwards. The ethicists on the other hand would provide the social scientists with a normative argument about the political and moral relevance of their findings.

---

[7] See Van der Steen (1995).

[8] Here the claim is not that 'a discipline' as such is integrated with another, but usually that given a certain problem or research field, the different features of the disciplines involved are merged. Admittedly, the concept 'discipline' reaches the limits of comprehensibility here.

[9] This use of the term differs from the way some other authors, such as Mittelstrass (2005) have used it.

[10] Van der Scheer and Widdershoven (2004).

This conceptualization of the relationship between the different disciplines does to some extent accept the different research aims of the disciplines, but does not seem to do justice to the nature of the interdisciplinary effort. Firstly, it raises the question of what a 'result' is. It seems that results of social science research can only be significant if we know at least something about the method used, and about the context in which the questions were framed. Secondly, we may further wonder whether a complex and abstract datum such as 'the opinion of the Dutch public' can be characterized as 'raw datum' at all. And thirdly, the normative assumptions within social science and the social scientific assumptions within applied ethics are deeply embedded in the research activities themselves. Such assumptions will have to be explicated if the interdisciplinary effort is to be fruitful.

7.3 The Nature of Ethics

Apart from general questions on the concepts of 'disciplinarity' and 'interdisciplinarity', there are also questions concerning these specific disciplines themselves. What is applied ethics and what should it be?

As the term suggests, applied ethics is one branch of ethics, which in turn is generally considered to be a discipline of philosophy. Often a distinction is made between applied ethics, normative ethics, descriptive ethics and meta-ethics. Applied ethics deals with moral questions concerning specific contested practices (as a general term for bioethics, environmental ethics, technological ethics etc.); normative ethics with more general normative considerations and principles; descriptive ethics is a form of social science that studies opinions, attitudes and convictions concerning morality; and meta-ethics deals with questions related to moral concepts themselves. Each of these approaches connects to empirical research in different ways.

Focussing on applied ethics, we may for example wonder whether applied ethics should mainly be concerned with conceptual problems or with concrete cases. Also, one can ask whether applied ethics should address such fundamental 'philosophical' questions implicit in the debate – such as 'What is pain?' and 'What is sickness?' – or whether it should stick to 'merely' normative reasoning of specific cases. Also, should applied ethics attempt to 'solve' moral issues, merely to 'elucidate' or perhaps be concerned with something else entirely? These questions have been answered in many different ways, which of course impacts significantly on the way we think about applied ethics in an interdisciplinary context.

Furthermore, there is a specific question on the justification of normative claims made in applied ethics. That is, which normative theory is the right one? Naturally, this is one of the large questions of practical philosophy, and therefore raises many complex issues. It goes without saying, therefore, that little can be taken for granted when discussing the role of 'applied ethics' in an interdisciplinary context. One specific debate has to be mentioned here, however, since it directly relates to the notion

of empirical evidence in ethics: the debate on what has been dubbed 'empirical ethics'.[11]

An influential view within the empirical ethics enterprise arose out of discontentment with what is often denoted as 'principlism'.[12] The thought is that principlism is a top-down ('foundationalist') strategy that presupposes universally held values. Instead, it is argued, we need to acknowledge that values are relative to their context, and that – thus – there is a pluralism of values. According to this critique the principlist approach is too rigid and rationalistic in its application to be able to answer to the subtleties of the practice. We do not claim that this is the only or even the most important critique of either principlism or top-down approaches, only that it is an influential one. We mention it here just to illustrate some of the confusions often found in the debate.

Inspired by the critique of top-down approaches, there have been a great number of attempts to be more sensitive to context and empirical reality. This has led to hermeneutic, reflective equilibrium and anti-theory approaches, which each in their own way try to give empirical data a place in ethics. The result is a 'contextualized' or 'empirical' ethics, which gives ethics a way of dealing with empirical data without claiming that the only proper form of thinking about morality is to succumb to relativism and to limit oneself to descriptive ethics.[13]

A confusing feature of this debate, however, is that it is not always clear whether ethics *itself* is considered interdisciplinary enterprise (consisting for example of contributions from philosophy, theology, social science, medicine, and so on), or whether the context-sensitiveness makes ethics fit for participation in an interdisciplinary undertaking. It is sometimes claimed that (bio-)ethics[14] is already interdisciplinary, but others seem to think that context-sensitiveness does not eliminate the own nature of the methodology of ethics.[15]

---

[11] It is perhaps interesting to note that when Ritchie and Kaplan (1940) speak of 'empirical ethics' they mean what we would now call 'descriptive ethics'. Here, however, we use the term to cover a wide variety of positions within applied ethics that assume that empirical evidence is directly relevant to the methodology of ethics, in contrast to approaches that consider empirical data to be no more than 'input' for ethical reflection, and thus external to the methodology of ethics itself.

[12] The term 'principlism' was first coined by Clouser and Gert (1990). The critiques on principlism are not in the first place focussed on Beauchamp and Childress (1989), traditionally associated with the 'Georgetown mantra', but instead primarily on their followers. Still, even for the most rigid principlist it seems odd to say that they are 'foundationalist'. If anything, the principlist approach seems pluralistic in character. It seems that the justification for calling it foundationalist is not so much that there is a solid foundation underlying the theory, but it seems to be meant that principlism is a top-down, instead of bottom-up-approach. See Borry et al. (2005), who in contrast to a foundationalist, principlist ethics discuss a model of 'ethics in action'.

[13] See for example Krones and Richter (2003), Den Hartogh (1999), Birnbacher (1999), p.322, Van der Steen (1995), p. 58 and Borry et al. (2005).

[14] For some reason bioethics seems to be associated with interdisciplinarity more often than applied ethics is.

[15] See Bennett and Cribb (2003).

But setting aside these questions on the nature of 'empirical ethics', does the argument against principlism justify the move towards empirical ethics? We believe it does not.

Note firstly that this critique brings in a meta-ethical claim on the nature of morality (i.e. the claim of pluralism) into the debate about empirical ethics. This is of course to some extent unavoidable, since any ethical discourse is in some way or the other related to the more fundamental questions. Yet, in this specific context the role of the meta-ethical assumption of pluralism is dubious, at best. It is one thing to hold that it is important to be conscious of different moral perspectives on a particular issue, but quite another to say that in ethical discourse the justification of a moral claim derives from public opinion. The dispute between top-down and bottom-up is on the issue of justification, rather than on the question whether there is a variety of perspectives.

Also, it is unclear how the claim of pluralism relates to the desire to be sensitive to the complexities of particular cases. It is of course true that mathematical deduction is an unattractive model of practising ethics; any practical discipline by definition has to be 'context sensitive'. However, it follows in no way from the observation that ethics needs to be sensitive to context, that it therefore needs to presume the truth of value pluralism.

And, lastly, the argument against principlism needs to seek out its target. If the target is a pluralistic theory, the reproach may for example be that it is unclear how the different principles relate. If on the other hand the target is a monistic theory no such question arises. Conversely, pluralistic top-down approaches are immune to some standard critiques aimed at monistic theories – for example those referring to the incommensurability of values.

7.4 The Type of Empirical Data

The questions concerning empirical data in ethics do not just arise depending on our conception of ethics. Naturally it makes a big difference what type of empirical data we are talking about. Empirical data stemming from physical, biological, historical, economic, legal and social science research all play a role in applied ethics in a wide variety of ways. This makes the question on the role of empirical data in ethics a complex one. It all depends whether we are talking about neuroscience insights in consciousness, or the implications of cross-cultural differences, or the social acceptance of stem cell research.

One of the reasons that so many authors have become fascinated with the relationship between social science and ethics is that the role of social scientific data in ethical discourse is complicated in ways in which the relation between ethics and, for example, biology is not.[16] The image of social science research as a fact-provider is perhaps distorting when it implies the assumption that the subject of this

[16] Hedgecoe (2001), p. 307.

discipline (opinions, narratives, arguments) is just 'out there'. The forming of an opinion is something that may only come about in the research setting: people may just begin articulating their opinion the moment they are asked. The question this raises is whether the data are intended to unveil what people actually think and do, or whether the research question was to discover what people would think or do.

Furthermore, social science data frequently emphasize the *reasons* for a particular opinion. This is particularly so in *qualitative* social science, which will in general have the structure of a reasoned opinion, not 'merely' statistical facts. It may be argued that this implies that qualitative social sciences cannot and should not live up to the standards of unbiased research prevalent in other branches of social science. It follows that concerning the relation between qualitative social science and ethics we have to take in account that the aforementioned cannot simply be classified as a provider of the facts on the way things are.

So, what kind of facts are we talking about here? When discussing public opinion, it is often more or less assumed that the morally relevant data we are looking for are moral opinions. But this is not necessarily so. Qualitative social science is a way to investigate a wide variety of different attitudes, preferences, arguments and experiences. So what may be interesting for the ethicist in a qualitative social science study may be intuitions, or it may be arguments, or it may be experiences from a certain perspective ('What does this mean for. . . a patient with disease X, a mother, an elderly person, etc.?').

From the perspective of the ethical discourse we can wonder what kind of empirical research could be interesting. There is a number of possibilities: We could be interested in the public *opinion as such*, which means looking for the opinion of the majority, opinions of specific ethnic, gender, religious etc. groups concerning a contested topic, let's say stem cell research. Another option could be to investigate *preferences* of specific societal groups: do they prefer research on embryonic stem cells or adult stem cells? Both approaches are only ethically relevant under the assumption that opinions and preferences have a certain authority. But if we really want to give moral authority merely to the *opinion or preferences* of the people, why not let the 'voice of the people' be decisive in all cases? Why hold on to the notion that ethics has any added value? Although this may sound as a rhetoric exaggeration, questions on the authority of the 'voice of the public' raise serious issues on the added value of ethics in public debates.[17]

In public debate and democratic process, social sciences and ethics meet in an interesting way. We think here specifically of the possibility for lay people or non-experts – 'the public' – to have a voice in decisions on particular topics, by way of qualitative participatory processes, such as focus groups. This phenomenon raises questions both on the authority of ethics, as well as on the authority of social scientific results.

Another topic of investigation one could be interested in is the *way people perceive* new technologies. If people talk about stem cell research, prenatal diagnosis

[17] Bovenkerk (2006), Bovenkerk and Poort (2008)

or organ transplantation, what kind of evaluative aspects are relevant to them? All kinds of ethical discourses make assumptions concerning the ways of perceiving of and dealing with technologies. If for the debate around prenatal diagnosis the autonomous decision of the pregnant woman is a relevant factor, it is important how the circumstances of the decision are perceived by women. The perception of technologies is e.g. also relevant in cross-cultural perspectives. Since nearly all modern technologies are applied in a global context it's relevant to know how the different cultural frameworks, the concepts of individuality, family, body and the like influence the way people perceive new technologies. This would be important for right-based approaches, for concepts of human dignity, for utilitarian concepts, and for liberal approaches as well. If it is morally important to protect the dignity of people, to protect their room of decision or to avoid harming them, one should be aware of the perspective of the people concerned.

A different possibility would be to focus on *arguments,* instead of on opinions and preferences. Since ethicists are interested in the validity of moral arguments, it should be the first interest of the discipline to get access to all available forms of argumentation. But it's very likely that the arguments that trained philosophers come up with are different from those of other people. It could be important to get access to the whole range of possible argumentations. That doesn't mean that the arguments used in non-academic or public contexts are 'better' arguments. It doesn't say anything about the validity of those arguments. The only relevant point is that the ethicist should have access to the whole range of possible argumentation.

These topics of research in social sciences – opinions, preferences, perceptions and arguments – are probably not the only interesting possibilities from an ethical point of view. But the points mentioned already show that there are different types of social science research that could be important from an ethical perspective in various ways.

7.5 The Embeddedness of the Researcher

Distinct from these considerations on the different disciplines and their relations, are the questions on the position of the individual researcher – in this case the ethicist. Many of the complaints on a lack of integration between social science and ethics have been phrased in terms such as 'leaving the ivory tower',[18] 'practical experience'[19] and critique of 'armchair philosophy'.[20] All this may be important, but it is not clear that this issue should be discussed as a methodological worry.[21] Whether

[18] Vorstenbosch (1993).

[19] Van der Scheer and Widdershoven (2004).

[20] Birnbacher (1999).

[21] The confusion becomes manifest when authors speak of the 'viewpoint' of the ethicist: is this the ethicist, qua person or is this the viewpoint of an entire discipline? The second interpretation has an odd ring to it. See Birnbacher (1999), p. 326.

it should be depends on the question whether the practical and empirical efforts of an ethicist are an integral part of the activity of practising ethics, or whether on the other hand empirical input functions as a mere vehicle for ethical reflection.

The question here is whether an ethicist should attain practical experience in order to have a better understanding of what he is talking about, and whether social science provides a good tool for doing so. Social science is by no means the *only* way for an ethicist to acquaint with society. An ethicist writing on ethnicity could live in a neighbourhood with people of many different ethnic backgrounds. An ethicist doing research on euthanasia could have his office in a nursing home. However, social science does seem to be a very efficient way of finding out what happens in society. In what way would this be relevant to his research?

These types of questions are not purely of a theoretical nature. If one ethicist is 'involved in society', and the other isn't, there is no a priori reason to think that the results of the second will be of a lower standard. As long as he gets his facts straight, there seems no reason to confront him with not 'being out there' enough.

But there may be more to 'understanding' than 'getting your facts straight'. This is a particularly attractive thought when we think that ethics is not only a theoretical enterprise, but also the exercising of a capacity by an individual researcher. If this is the case then this capacity might have to be trained properly. Then the demand of coming out of the ivory tower would be motivated by the desire to become more sensible to societal demands, or 'what it is like' to be in a certain position. Furthermore, by knowing what sort of arguments are generally considered important, the ethicist may develop a sense of what type of research is currently important, or socially relevant. For example, if the doctrine of double effect is very often considered to be an important argument in the public debate on euthanasia, this may give ethicists a reason to reflect on the doctrine of double effect, independently of theoretical reasons for finding this particular argument important. Also, an ethicist may like to consider the public as a 'sparring partner'. He may be curious how certain arguments are received and on what grounds they are accepted or rejected. This may help him to reconsider his own arguments, and to try to make them more convincing. We do not think that any of this is very controversial. However, the controversy starts when we ask what the methodological status is of this input. Some ethicists believe that these encounters with practice *should not* come back in the ethical texts themselves, other than as a mere illustration. Therefore, even if it could be useful for the individual ethicist to go out and explore the world, inspiration or motives to write on a certain topic have no place in ethical research, properly speaking, let alone as a justification for a normative claim. Others, however, think that practical experience can be an important *heuristic tool*. From this perspective, the embeddedness of the researcher is integral to the practice of ethics itself. This is certainly a possible position on the matter, but it is certainly not uncontroversial.

Since there is a case for the claim that these type of considerations are independent of the methodological considerations, they could play a role, irrespective of the methodological convictions of the ethicist. One may be utilitarian, Kantian or principlist, and still use social science in this way. This would mean, however, that many of the reproaches that have been addressed to ethicists (that they should

leave their ivory tower etc.) were actually missing their target, if the aim was the methodology of ethics.

7.6 What's Next?

In the past twenty years a complex field of discussion has emerged around the topic of empirical data in applied ethics. Our attempt to categorize these debates has opened different areas of possible discussion. In conclusion, we will give some suggestions for further debate.

First, an interesting question that has thus far received far too little attention is on *how the different ethical theories deal with research in social sciences,* and how proponents of the different theories participate in interdisciplinary research. Many of the questions that arise in the debate on interdisciplinarity depend in their answering on the background assumptions of the ethicist involved. However, we should not be tricked into thinking that the relation between these background assumptions and their application is always straightforward. It would be a mistake to think that only a utilitarian takes opinion polls into account, or that a Kantian is indifferent to empirical data.[22] But to some extent it's very likely that different normative ethical theories will ask different questions to the social scientist. This normative ethical horizon for the discourse with social scientists should be explicitly reflected. Furthermore, different metaethical assumptions may turn out to correlate to their application in unexpected ways. It is e.g. conceivable that contrary to what is often supposed emotivistic theories of the good are less open to empirical data than so-called 'rationalistic' ones are. An emotivist could only be interested in attitudes and opinions people have concerning new technologies, but for him the underlying argumentations would not necessarily be very interesting. A more rationalistic type of ethics could then be a much more appropriate partner for qualitative research in social sciences. In any case, a large field is there to be explored on these topics.

Second, it seems imminent that more research will be done on the relation between the *desirability of public participation and the public role of the ethicist.*[23] This is not so much a question on truth finding, but rather on the social implications of scientific and philosophical results. A distinction can (and should) be made between the ethicist, qua ethicist, and the ethicist that makes policy recommendations. In a policy recommendation social scientists and ethicists meet again. In a democratic society the opinion of the public is, rightly, deemed important. But this is not yet to say that it should always have priority over other inputs. On the other hand, the ethicist may have reason to be less uncompromising in his policy role than in his purely scientific work. How should these roles relate? In that context, however,

[22] Winkler (1993) treats armchair philosopher and Kantian as synonyms!

[23] Of course, we do not want to claim that such research would be entirely new. A lot of the literature we have referred to addresses the issue of the political role of ethics. We believe however that some elements in this discussion have been underemphasized.

the public role of the social sciences has to be discussed as well. In the process of policy making not only ethicists are involved but social scientists as well. While the role of ethics in policy making is often discussed, the role of social sciences has been a topic of scientific research on a much lesser degree. But for policy advice it is important to know precisely for which reason which kind of expertise is needed. That process should be discussed much more in detail.

A third important field of future research is the more theoretical question of the relation between the status of arguments and the status of opinions. If one holds that ethical argumentation is no more than the explication of a point of view, in other words the forming of a more sophisticated opinion, what else is ethics then, but the factual statement that one person has a certain opinion on a certain issue? This forces the ethicist to be explicit about why reasons count as *good* reasons. This is not just a question on the nature of the good, and ethics as such, but also a question about the relation between common sense and ethical reasoning. How does a good argument distinguish itself from a bad one? Is an ethical argument refuted when it turns out to be contrary to common sense? What is the status of 'intuitions' in ethics?

It turns out that the question on the interdisciplinary relation of ethics and social science falls apart in a large variety of questions. The different topics are interdependent: answers to the questions asked in one area are closely related to answers given in another. However, in dealing with these questions one should be aware of their distinctness, that is, one should be aware that answering one type of question does not automatically give you an answer on the other ones.

References

Alvarez, A.A.A. (2001), "How Rational should Bioethics be? The Value of Empirical Approaches", *Bioethics* 15 (5/6), 501–19.

Arnason, V. (2005), "Sensible Discussion in Bioethics: Reflections on Interdisciplinary Research", *Cambridge Quarterly of Healthcare Ethics* 14, 322–28.

Arnold, R.M. and Forrow, L. (1993), "Empirical Research in Medical Ethics: An Introduction", *Theoretical Medicine* 14, 195–96.

Ashcroft, R.E. (2003), "Constructing Empirical Bioethics: Foucauldian Reflections on the Empirical Turn in Bioethics Research", *Health Care Analysis*, 11 (1), 3–13.

Barbour, R.S. (2001), "Checklists for Improving Rigour in Qualitative Research: A Case of the Tail Wagging the Dog?", *BMJ* 322, 1115–7.

Beauchamp, T.L. and Childress, J.F. (1989), *Principles of Biomedical Ethics*. Beauchamp, T.L. and Childress J.F. (2001), *Principles of Biomedical Ethics, Fifth Edition*. Oxford: Oxford UP.

Bennett, R. and Cribb, A. (2003), "The Relevance of Empirical Research to Bioethics: Reviewing the Debate", in: Häyry & Takala (2003a).

Birnbacher, D. (1999), "Ethics and Social Science: Which Kind of Co-Operation?, *Ethical Theory and Moral Practice* 2, 319–36.

Borry, P., Schotsmans, P., and Dierickx, K. (2004a), "What is the Role of Empirical Research in Bioethical Reflection and Decision-Making? An Ethical Analysis", *Medicine Health Care and Philosophy* 7, 43–53.

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

Borry, P., Schotsmans, P., and Dierickx, K. (2005), "The Birth of the Empirical Turn in Bioethics", *Bioethics* 19 (1), 49–71.

Bosk, C.L. (1999), "Professional Ethicist Available: Logical, Secular, Friendly", *Daedalus* 128 (4), 47–68.

Bovenkerk, B. (2006), "Biotechnology, Disagreement and the Limitations of Pubic Debate", in Kaiser, M. & Lien, M.E. (eds.), *Preprints of the Sixth Congress of the European Society for Agricultural and Food Ethics: Ethics and the Politics of Food*, Oslo, 97–102.

Bovenkerk, B. and Poort, L.M. (2008), "The Role of Ethics Committees in Public Debate", *International Journal of Applied Philosophy*, 22 (1), 19–35.

Brody, B.A. (1993), "Assessing Empirical Research in Bioethics", *Theoretical Medicine* 14, 211–19.

Callahan, D. and Jennings, B. (eds.) (1983), *Ethics, The Social Sciences and Policy Analysis*. New York: Plenum Press.

Clouser, K.D. and Gert, B. (1990), "A Critique of Principlism", *Journal of Medicine and Philosophy* 15 (2), 219–36

Dawson, A. (2005), "Informed Consent: Bioethical Ideal and Empirical Reality", in: Häyry & Takala (2005).

Devries, R. and Subedi, J. (eds.) (1998), *Bioethics and Society: Contructing the Ethical Enterprise*. Upper Saddle River, NJ: Prentice Hall.

Düwell, M. and Steigleder, K. (2003), *Bioethik; Eine Einführung*. Frankfurt am Main: Suhrkamp.

Düwell, M. (2008), *Bioethik; Methoden, Theorien und Bereiche*, Stuttgart [etc.]: Metzler.

Düwell, M. (2005), "Sozialwissenschaften, Gesellschaftstheorie und Ethik", *Jahrbuch für Wissenschaft und Ethik* 10.

Fox, R. and Swazey, J. (1984), "Medical Morality is not Bioethics- Medical Ethics in China and the United States", *Perspectives in Biology and Medicine* 27, 337–60.

Goldenberg, M.J. (2005), "Evidence Based Ethics? On Evidence-Based Practice and the "Empirical Turn" from Normative Ethics", *BMC Medical Ethics* 6 (11).

Grunwald, A. (2004), "The Normative Basis of (Health) Technology Assessment and the Role of Ethical Expertise", *Poiesis & Praxis* 2, 145–93.

Gutman, M. (2005), "Disziplinarität und Inter-Disziplinnarität in Methodologischer Sicht", *Technikfolgenabschätzung- Theorie und Praxis* 2 (14), 69–74.

Haimes, E. (2002), "What Can the Social Science Contribute to the Study of Ethics? Theoretical, Empirical and Substantive Considerations", *Bioethics* 16 (2), 89–113.

Hartogh, G. den (1999), "Empirie en Theorievorming in de Ethiek", *K&M: tijdschrift voor empirische ethiek* 23, 172–7.

Have, H.A.M.J., and Lelie, A. (1998), "Medical Ethics Research Between Theory and Practice", *Theoretical Medicine and Bioethics 19, 263–76.*

Häyry, M. and Takala, T. (eds.) (2003a) *Scratching the Surface of Bioethics.* Amsterdam: Rodopi.

Häyry, M. and Takala, T. (2003b), "What is Bioethics All About? A Start", in: Häyry & Takala (2003a).

Häyry, M. and Takala, T. (eds.) (2005), *Bioethics and Social Reality.* Amsterdam: Rodopi Hedgecoe, A.M. (2001), "Ethical Boundary Work: Geneticization, Philosophy and the Social Sciences", *Medicine Health Care and Philosophy* 4, 305–9.

Hedgecoe, A.M. (2004), "Critical Bioethics: Beyond the Social Science Critique of Applied Ethics", *Bioethics* 18 (2), 120–43.

Heeger, R. and Willigenburg, T. van (1993), *The turn to applied ethics*. Kampen: Kok Pharos Hoffmaster (1990), "Morality and the social sciences", in: Weisz (1990). pp. 241–60.

Holm, S. and Jonas, M. (2004) *Engaging the world: The Use of Empirical Research in Bioethics and the Regulation of Biotechnology.* Amsterdam [etc.]: IOS Press.

Keulartz, J, Schermer, M., Korthals, M. and Swierstra, T. (2004), "Ethics in Technological Culture: A Programmatic Proposal for a Pragmatist Approach", *Science, Technology & Human Values* 29 (1), 3–29.

Krones, T. and Richter, G. (2003), "Kontextsensitive Ethik", in: Düwell & Steigleder (2003).

Levitt, M. (2003a), "Public Consultation in Bioethics. What's the Point of Asking the Public When They Have Neither Scientific Nor Ethical Expertise?", *Health Care Analysis* 11 (1), 15–25.

Levitt, M. (2003b), "Better Together? Sociological and Philosophical Perspectives on Bioethics", in: Häyry & Takala (2003a). pp. 19–28.
Levitt, M. (2004), "Complimentarity Rather than Integration", *Medicine Health Care and Philosophy* 7, 81–3.
Levitt, M. and Häyry, M. (2005), "Overcritical, Overfriendly? A Dialogue Between a Sociologist and a Philosopher on Genetic Technology and Its Applications" *Medicine Health Care and Philosophy* 8, 377–83.
López, J. (2004), "How Sociology Can Save Bioethics. . . Maybe", *Sociology of Health and Illness* 26 (7), 875–96.
Mays, N. and Pope, C. (2000), "Qualitative Research in Health Care: Assessing Quality in Qualitative Research", *BMJ* 320, 50–52.
Miller, A. (2003), *An Introduction to Contemporary Metaethics*. Cambridge [etc.]: Polity Press Mittelstrass, J. (2005), "Methodische Transdisciplinarität", *Technikfolgenabschätzung- Theorie und Praxis* 2, 18–23.
Molewijk, B., Stiggelbout A.M., Otten W., Dupuis H.M. and Kievit, J. (2003), "Implicit Normativity in Evidence-Based Medicine: A Plea for Integrated Empirical Ethics Research", *Health Care Analysis* 11 (1), 62–92.
Molewijk, B. (2004a), "Integrated Empirical Ethics: In Search for Clarifying Identities", *Medicine Health Care and Philosophy* 7, 85–7.
Molewijk, B., Stiggelbout A.M., Otten W., Dupuis H.M. and Kievit, J. (2004b), "Empirical Data and Moral Theory. A Plea for Integrated Empirical Ethics",*Medicine Health Care and Philosophy* 7, 55–69.
Morgan, G. and Smircich, L. (1980), "The Case for Qualitative Research", *The Academy of Management Review* 5, 491–500.
Musschenga, A.W. (1999), "Empirical Science and Ethical Theory: The Case of Informed Consent", in: Musschenga, A.W. & Steen, W.J. van der (eds.) (1999), *Reasoning in ethics and law*. Aldershot: Ashgate.
Pearlman, R.A., Miles S.H. and Arnold, R.M. (1993), "Contributions of Empirical Research to Medical Ethics", *Theoretical Medicine* 14, 197–210.
Putnam, H. (2002), *The Collapse of the Fact/ Value Dichotomy*. Cambridge MA [etc.]: Harvard UP. Ritchie, B.F. & Kaplan, A. (1940), "A Framework for an Empirical Ethics", *Philosophy of Science* 7 (4), 476–91.
Scheer, L. van der and Widdershoven, G. (2004), "Integrated empirical ethics: Loss of normativity?",*Medicine Health Care and Philosophy* 7, 71–9.
Skorupinski, B. and Ott, K. (2002), "Technology assessment and ethics", *Poiesis & Praxis* 1, 95–122.
Steen, W.J. van der (1995), *Facts, value and methodology*. Amsterdam [etc.]: Rodopi. Solbakk, J.H. (2004), "Use and Abuse of Empirical Knowledge in Contemporary Bioethics",*Medicine Health Care and Philosophy* 7, 5–16.
Sugarman, J. and Sulmasy, D.P. (2001), *Methods in Medical Ethics*. Washington DC: Georgetown UP. Twine, R. (2005), "Constructing Critical Bioethics by Deconstructing Culture/Nature Dualism", *Medicine, Health Care and Philosophy* 8, 285–95.
Velleman, D. (1989), *Practical Reflection*. Princeton, NJ: Princeton UP.
Vorstenbosch, J. (1993), "Four Ways of Leaving the Ivory Tower; Perspectives on Research in Applied Ethics", in: Heeger & Van Willigenburg (1993).
Vorstenbosch, J. (1999), "Open Deuren en Dichte Luiken", *K&M: Tijdschrift Voor Empirische Ethiek* 23, 178–85.
Weisz, G. (ed.) (1990), *Social Science Perspectives on Medical Ethics.* Dordrecht [etc.]: Kluwer. Welie, J.V.M. (1998), "Clinical ethics: Theory or Practice", *Theoretical Medicine and Bioethics* 19, 295–312.
Zussman, R. (1997), "Sociological Perspectives on Medical Ethics and Decision-Making", *Annual Review of Sociology* 23, 171–89.
Zussman, R. (2000), "The Contributions of Sociology to Medical Ethics", Hastings Center Report 30 (1), 7–11.

Chapter 8
Facts or Fiction? A Critique on Vision Assessment as a Tool for Technology Assessment

Nicole C. Karafyllis

Abstract This essay questions the concept of vision assessment as an appropriate tool for technology assessment on methodological, anthropological and ethical levels, and shows its epistemic neighbourhood to the scenario-techniques. In general, the central idea of a "future technology" is critically analyzed, backed with central examples drawn from nanotechnology, human doping techniques, and social neuroscience. Main concepts that are used for critique are (a) feasibility and desirability, (b) discourse and debate, and (c) vision and utopia. In addition, the essay reflects on the new genre of pop science, a mixture of science and popular writing on which the concept of vision assessment heavily depends.

Keywords Vision assessment · Pop science · Social neuroscience · Doping · Nanotechnology · Technology assessment · Utopia · Imagination · Desirability · Discourse

8.1 Methodological Inquiry on Vision Assessment

Vision assessment is a method that focuses on technologies of the future and debates about future(s) in general. The time seems to be right for "assessing visions" because at present, in the age of mass media culture and virtual simulation, the epistemological border between materially designing science futures and narrating science fiction has become more and more invisible (see, e.g., Pethes and Schicktanz, 2008).

This blending is a challenge for philosophers. In this article, some of the methodological, anthropological and ethical weaknesses of *vision assessment* will be outlined – stressing, however, that envisioning good futures remains at the core of ethical expertise. At the outset (Part 1), I will outline four basic arguments against vision assessment, particularly in relation to the field in which vision assessment has already been applied, i.e. nanotechnology. Then (Part 2), I will focus on related

N.C. Karafyllis (✉)
College of Humanities and Social Sciences, United Arab Emirates University, Al Ain (Abu Dhabi), United Arab Emirates
e-mail: karafyllis@t-online.de

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_8,

terminologies that deal with future(s), e.g. "technological future" and "future technology", in order to show their discontinuities with classical utopian thinking. In addition, I will ask about their different normative implications, particularly regarding modes of participation and approaches to objectify individual futures. In order to show the multirole purposes that the reference to "nature" and "technology" can fulfil within public discourses about the anthropological future of the human being (Habermas, 2001), I will focus on the example of *techno-doping* in Part 3. This should clarify the limits of a vision assessment that seeks to embrace a society's vision on technological progress. The latter section is followed by a second example (Part 5), which outlines an experiment in social neuroscience which is put forth for the sake of a better future society, emphasizing the early stages of experimental design in which values already enter without being made explicit. The overall aim of the second example is to show how the genre of *pop science* (Part 4) triggers (not only) neuro- and nanotechnology as "hot topics", thereby underlining that the strategic role of the scientists (the experts) and media within the methodology of vision assessment has to be thoroughly revised. Finally (Part 6), I will summarize my thoughts and ask about the opportunities that vision assessment might have to offer within established approaches of technology assessment.

That individuals proactively "react" on envisioned futures is one of the basic assumptions within *vision assessment* (Grunwald, 2004, 2007), a newly proclaimed tool within the old toolbox of technology assessment (TA). The latter is a conglomerate of methods enabling scientists to analyze the development and implementation of new technologies and their impact on society (VDI, 1991; Skorupinski, 2000). Aiming to gain knowledge for orientation and political decisions by interdisciplinary scientific research, TA is normatively based on both the *feasibility* and the *desirability* of new technologies. Whereas the study of feasibility concentrates on acquiring sufficient technical data of present prototypes and exploring possible breakthroughs in basic research, e.g. by interviewing experts, the study of desirability is even more complex. Desirability is based on the assumption of individual *potentials* – which follow psychological and social dynamics – rather than on technological *possibilities*, i.e. the measurement of desirability partly resists statistical methodology and even hard empirical analysis.

At this point, vision assessment sets in. Particularly the difficulties in assessing *nanotechnology*, assembling different techniques and disciplines (Nordmann, Schummer and Schwarz, 2006), gave rise to the idea that analyzing the *discourses* on these techniques might clarify its desirability.

First, let us ask if there already is something like a discourse on nanotechnology. In the case of nanotechnology, the wider TA-community including ethicists has faced and still faces a new situation. They are strongly encouraged (and financed) by national and international science policies to assess and evaluate nanotechnology even *before* the object of assessment is defined (see e.g. BMBF, 2004; EGE, 2005; ESF, 2007; Schmid et al., 2006; TA-Swiss, 2006). For instance, a thorough risk assessment is impossible at this stage. Nanotechnology is still in its infancy. Apart from nanomaterials (nano-tubes, nano-ceramics) there are hardly any products ready to be assessed. Of course, this is not a new situation for TA. However,

at the same time as the need for assessing nanotechnology was politically articulated, nano-particles were already aggressively marketed and they were even used in cosmetic products as carriers to transfer vitamins into deeper skin layers, without any TA before. This situation is new. Nevertheless, ethicists and TA-researchers publish excessively on nanotechnology, and even a new journal with the title *NanoEthics* was launched in 2007. There definitely is an *academic* discourse on the factual "non-presence" of nanotechnology (for a critique, see Nordmann, 2007, 2008; Karafyllis, 2008) – an academic discourse to which philosophers also contribute in terms of a negative self-fulfilling prophecy: the more you write about it, the more nanotechnology seems to already exist; but does this already count as a societal discourse?

To the contrary, nanotechnology hardly makes it to the front pages of daily newspapers or TV news stations (other than genetic engineering, neuroscience or nuclear energy) because the majority of people are not (yet?) interested in it. Similarly, most scientists themselves refuse the new labeling within academia, and they do so for good reasons. As it is far from clear which techniques or products might result from basic research on the nano-scale (i.e. $<10^{-9}$ meters), the majority of possibly involved scientists have already decided not to become "nanoscientists" or even "nanotechnologists". Instead, they still work under the disciplinary headlines of microelectronics, physics, biology, chemistry, genetic engineering, neuroscience, medicine, mechanics and material science. Whereas the diversity and interdisciplinarity of a nanotechnology in the making still imposes methodological problems for TA and ethics, the concept of vision assessment seems to offer at least a stopgap. If as of yet there are no prototypes to assess and evaluate, then can one not at least assess and evaluate visions of future applications – or even the future(s) in which they might be embedded?

Thus, second, the basic terms and methods of vision assessment have to be critically analyzed. Armin Grunwald's and his Karlsruhe colleagues' key steps of vision assessment concentrate on (a) collecting the visions on *one specific future technology* in the present debate, (b) analyzing their content and strategic role *within* the debate, (c) evaluating the normative implications involved, and (d) scrutinizing the practical impact of the transported visions on present society. Note that the protagonists first speak of "debate" in steps (a)–(c) rather than "discourse", a term that is reserved for (d), i.e. the societal level on which the findings are finally projected. At first sight, the thrilling idea is to *delegate the process of evaluation* directly to the public sphere, as *media analysis* is suggested as one instrument of vision assessment, though the methodology as a whole remains unclear (e.g., which specific media are analyzed?). Particular when expert knowledge and popular science writing merge into "pop science" (see Part 4), the analytical difference between descriptive and normative "facts" becomes blurred, though the resulting vision of the future might be alluring. Almost hidden, the strongly normative premise of Grunwald's approach is that the technology to be assessed (e.g., nanotechnology) actually *is* a "future technology". But how do TA-researchers, scientists, journalists, or "the public" know that? What, if the majority of people do not want this technology? What if it causes, for example, irreversible harm and/or huge social

disparities? What, if it turns out to be too costly? A "future technology" only is one if it will be bought and used by today's *and* tomorrow's people, meeting their needs and/or fulfilling their desires. Moreover, focusing on one specific "future technology" shows a methodological mangle of science practice, as any development of new technologies heavily depends on synergies with other fields. At least the scientific rhetoric already incorporated these possible synergies. For example, an all-embracing "nanoworld" is based on the premise that it will gain surplus value not only for material science but also for microelectronics, prosthetics, gene therapy or neuroscience. It is not far-fetched to assume that nanotechnologies will succeed to provide solutions for some fields of applied research, but not for all that are proclaimed today.

Vision assessment should be situated in a background theory of "prospective knowledge", or "outcome knowledge", as Grunwald (2007: 54) puts it (in the German original: "Folgenwissen"). This means that only individuals who can proclaim to have this knowledge, i.e. researchers, can contribute to this theoretical background. In the best of all worlds for TA-researchers, all members of society would follow the discourse rules set up by discourse ethics when debating about their competing visions of the "future" (ibid: 57). At first sight, this seems to be a participatory element. But upon closer inspection, the aim is to reach consensus on *one* future for one society, depending on one specific technology – a highly exclusionary approach which is, moreover, unlikely to be productive for political decision-making in democratic societies. As it seems, public participation in the societal discourse is substituted by persuasion from actors that are professionals in both mass communication and science. However, before this very special TA-vision of complete participation or persuasion might become true one day, at least the experts' visions should be analyzed, according to the protagonists of vision assessment. In this understanding, both the expert and the journalist are seen as a watchdog for society. This is a rather idealistic view of their professional roles. Scientists and journalists seem to be a perfect match, consisting of the *analyzer* (of empirical technological data *for* the future) and the *sensor* (of normative implications of technologies *in* the future, on a meta-level). For example, empirical socioeconomic data on present inequalities which are triggering a technology's desirability in the real world remain a blind spot from the very beginning of vision assessment – at least, if the interviewed scientist does not have a strong moral interest in these issues, commonly referred to as an "ethos". As the third party on a politically powerful meta-meta-level, the TA-researcher becomes the *interpreter* of this conglomerate of facts, visions, fears and wishful thinking. That all three also have to sell their products on linguistic and visual markets (Bourdieu, 2005) is tacitly ignored. Neither does the scientist, when interviewed, only *describe* technologies and thus not represent only the empirical level of research (moreover, he/she has only one single voice within the pluralist scientific community), nor does the journalist completely represent the normative level of interpreting future technologies, or societies' pluralist opinions on these technologies as a whole. In consequence, the very idea of powerfree discourse and achievable consensus within a pluralist society, most prominently put forward by Jürgen Habermas, is slimmed down to a low-calorie "discourse light"

for the leading elites of high-industrialized societies. One is reminded of Habermas' early work, *Toward a Rational Society* (engl. 1970, Chapter 6), in which he criticized science and technology as ideology: by reducing practical questions about the good life to technical problems for experts, contemporary elites eliminate the need for public, democratic discussion of values. Thus, the society becomes depoliticized by its elites which ignore the dialogical principle implied in practical reasoning.

Third, in order to understand the strategic value of the new method, let us look at the contemporary history of vision assessment. The roots of this idea can be found in the mid 1990s, where the concept of TA had to face a major crisis in political perception and implementation (first in the USA, then in Western-Europe) as it was regarded too critique-driven, thereby appearing to slow down technological progress. The concept of vision assessment heavily depends on the *scenario-approach*, developed newly in the 1990s as a less critical method of TA. Rather than criticizing a technology in the making, different scenarios should envision societal options in which the new technologies could make more or less sense and create different kinds of benefits. The scenario-approach has been used quite frequently for envisioning energy-scenarios of the future and determining benchmarks for funding research on, e.g., regenerative energies or nuclear energy. Scenarios of future technologies were envisioned even if experts doubted the technical feasibility, based on the premises of data present at the time when scenarios were created. There, the "fictionists" and the creators of "facts" have not been identical, which has been a reason why the scenario-approach did have some fruitful outcomes for designing socially acceptable futures (e.g., regarding energy futures that depend on renewable resources), and it still does.

Fourth, compared with the scenario-approach, in vision assessment the professional role of TA-researchers themselves vastly changes, and so do their modes of responsibility. They *select*, *aggregate*, *combine* and *interpret* both data and values from experts and media rather than designing their own scenarios, based on their own formulated values, for which they could be held responsible. The new element within this process is to assess the media, i.e. to distract oneself from the established meta-level of scientific discourse by reaching higher, on the level of a meta-meta-analysis, which appears to encompass society as a whole. Within the Karlsruhe approach of TA, we can thus find a normative shift towards the creation of *future facts* (empirically based on present data and values) instead of painting a picture of *possible futures based on certain and possible facts and values*. Possible value shifts are not taken into account at all, at least not explicitly. Overall, this tendency of hiding value-discussion is not only dangerous, but counterproductive for TA. Because: What we desire *for* the future has to be imaginable now – a reason why Grunwald (2007) speaks of the "immanence of presence". But what humans will desire *in* the future might be completely different. The same is even true for what humans decide for the future now, particularly, as the example of nanotechnology shows, when the basic premises for decision-making (e.g., consistent and coherent information) are not even given. Moreover, without explicitly formulating values, the process of TA will remain on the reduced level of a mere

assessment (implementing hidden values, though) rather than reach the stage of evaluation.

To sum up: In the case of nanotechnology, science and technology assessment is dealing with a bunch of techniques that relate to other fields and which are not yet (and maybe never will be) reality, such as "meta-brains" and "smart skin", and it often uses images and narratives of the science-fiction genre (Coenen, 2006). Communication on nanotechnology embraced "the future" much more than, for example, genetic engineering (see, e.g., Drexler, 1986). The concept of vision assessment goes further in normatively describing futures than the scenario-approach as it uses value-laden tools (media analysis and expert interviews) without highlighting these a priori imported values. Vision assessment seems to offer the long-wished-for tool of cultural appropriation of technological developments. Close up, neither the concepts of "public", "discourse", "culture", nor "society" are defined, nor is the theoretical problem solved that evaluation necessarily depends on assessment, and, last but not least, on a priori information ready to use for assessment. Fiction cannot be evaluated without facts; and facts on future technologies do not emerge without artifacts. *As it seems, the main thesis of vision assessment is that desirability can be assessed without sound knowledge about feasibility*.

Let us take stock in order to raise more methodological doubts. Is the concept of vision assessment applicable on technologies in general, or is it bound to the specific case of nanotechnology? Vision assessment, I argue, is epistemologically still linked to the case of assessment for which it was developed, i.e. nanotechnology. Therefore, it is not a general concept for TA. Under premises of ethical judgment (e.g., the precautionary principle), it might not even be a concept for evaluating nanotechnology. As a consequence, currently feasible (e.g. carbonfibres on the nano-scale) and non-feasible techniques (nano-robots) are united within the designed futures. However, vision assessment is also applied in the area of ubiquitous/pervasive computing. It might soon also be used in the field of neurotechnologies, as the related semantic field of one's own biographical future is more than alluring (see Part 5). Astonishingly, however, the method of vision assessment has not yet been used to design futures based on predictive genetic testing, an area that like no other influences the shape of future societies on a material level.

In the following, I will sketch other influential approaches dealing with "the future" and show their normative backgrounds. The overall aim is to critically reflect on the recent tendency to transform TA into vision assessment, thereby tacitly transforming the participatory element embodied by present social actors into technology- and economic-driven wishful thinking of academic experts.

8.2 The Pasts and Futures of the Future: Utopy and Vision

The idea that visions can be assessed and evaluated emerges in a specific cultural mentality, driven by the idea of (technological, economical, cultural, individual) *progress*. In modern times, the future is envisioned as both open and individually

designable, including the future of one's own body, brain and social relationships. The mentality of present Western societies is powered by the idea that individual potentials can be – and have to be – detected, explored and shaped for the sake of a "better" future, both in the individual and societal sense. In this mentality, long known time- and space-related categories ("world", "individual life") have changed and are becoming "more open" and sometimes even unlimited on the semantic level (e.g., the visions of outer space, cyberspace, physical enhancement and anti-aging). Examples are the continuous biomedical enhancement strategies for the human body and the concepts of life-long learning and emotional intelligence, just to name three out of many. The latter were triggered by brain research on neuroplasticity since the 1990s, and developed as psychobiological tools to train the brain in order to become economically and socially more successful (Karafyllis and Ulshöfer, 2008).

A crucial part of these "self-technologies" is to constantly assess one's own visions, compare them with those of others, and reflect on their future realizations. Underlying this are reality-checks which encompass the present potentials and achievements, including the access to technical means for enhancements. The modern subject seems to be individually responsible for her/his future and is encouraged to enhance "natural" potentials. Against the normative background of an assumed *competition* of both societal visions and individual potentials, this definitely can make sense. Futures still seem to be something subjectively defined, but thought as achievable against a socially objectified background.

On the other hand, the future itself has turned out to become an object, as Bill Joy's famous article "Why the future doesn't need us" (2000) might illustrate. The concept of a "technological future" meets this rhetoric. But how can "the future" have needs? Is "the future" a competitor, and with whom or what is it competing? With the presence? With the past? With nature? By proclaiming a technically upgraded future in which human beings (and their present needs) are dispensable, nothing less than the anthropological self-image of being human is under attack. Underlying this is a reference to a holistic concept of nature, and that this abstract nature does not necessarily need humans – an argument that, if taken seriously, would also account for the present. Thus, instead of debating individually competing futures within society (e.g., regarding job opportunities or access to healthcare and education), the mentioned terminologies allow for change of the normative battlefield to the species level, as if the whole human race is under threat when resisting technological progress.

It is precisely this mentality of self-objectification and the ideology of technical progress with a life of its own, the latter competing with a holistic nature with a life of its own, in which the concept of an overall *vision assessment* as a tool for TA has recently been able to emerge. At present, individual visions associated with future technologies, e.g. nano- and neurotechnology, not only become decisive factors for science policies but also a tradable commodity on established "vision markets" in which holistic (some might say totalitarian) concepts of nature, technology, society and also future are merged. This development has already hit the radar of Science and Technology Studies which scrutinize how "science futures" are made

up.[1] Not rarely, these envisioned science futures centre on only one specific "future technology".

Getting back to our initial example: The academic debate about nanotechnology was accompanied from its early beginning by visions and religious symbolisms comparable to the field of space flight, superconductivity, or nuclear technology (see Coenen, 2006 for details). Scientists were actively communicating with the media and delivered persuasive images of the future technology in the envisioned state of application (Lösch, 2006), for instance images of nano-bots cruising the inside of human veins and arteries. The latter were visualized as approachable and, more or less, empty spaces. The imagination of "another" world in relation to modern high technologies is often put forward in the spatial rather than the temporal sense. The viewer is persuaded to already *be* in this microcosmic world rather than *have* it one day. This approach of envisioning phenomenal corporeality is supported by metaphoric language such as "nanoworlds", "nano-landscapes" and "meta-brains".

This imaginary skeleton of science visions differs from the classic utopias of Francis Bacon and Thomas Morus (see Saage, 2001–2004, Schaper-Rinkel, 2004). The early modern utopias, most prominent in Bacon's *Nova Atlantis* (1627), were also backed-up with contemporary science and social experiences, but located on unknown places within *this* world and time. They were both there and not there. Theoretically, these utopias could be detected by a captain that finds the courage to sail in unknown waters and finds ways to deal with the many uncertainties and risks involved. At that time, the sailing ship, a long-known metaphor for individual human life, was semantically transformed for addressing the process of scientific endeavour, reaching out for the new worlds (in plural!) on the micro- and macroscale that are hidden parts of this world. For example: When the microscope was established as a new tool of scientific observation, one of Bacon's heirs and member of the Royal Society of London, the botanist Nehemiah Grew, introduced his groundbreaking *The Anatomy of Plants* (published: London 1672) with a dedication to King Charles II: "In sum, Your Majesty will find, that we are come ashore into a new World, whereof we see no end. It may be, that some will say, into another Utopia."

Classic utopias believed in the advancement of science and learning for the sake of a better society. On the contrary, however, they offered a chance for the critique of present boundaries that emerged due to imbalances of power and dominion (Saage, 2001, Vol. 1: 30). This is but one of the reasons why the utopian society had (and still has) to remain in this world. Hence, a technical artifact (e.g., the microscope) did not have a future of its own other than contribute to a society's future as a new method for learning about nature's secrets and finding tools to use them.

---

[1] Recent examples are the international conferences "Science Futures" at the ETH Zurich (6.–9. February 2008; http://science-futures.ch) and "Szenarien der Zukunft. Technikvisionen und Gesellschaftsentwürfe im Zeitalter globaler Risiken" at the RWTH Aachen (18.–19. October 2007; http://szenarien.rwth-aachen.de). (Last visit for both: 28 January 2008).

Time-related utopias emerged during Enlightenment. Modern utopias related to high technologies envision future worlds (often related to other planets or outer space) in future times, thereby transforming the idea of "future". The future is no longer something one hopes for today. As already mentioned, the future becomes part of an objectification process instead. In some scientists' visions, it seems to have become an agent of its own and implies the meta-perspective. In consequence, the human individual is terminologically merged into a human race (which also implies a meta-perspective) whose present outfit seems to be dispensable for both the subjectified futures and the objectified future as a whole. As Edmund Husserl once put it: "Mere fact-sciences make mere fact-humans" (Husserliana IV, 4; transl. N. K.).[2]

The question of the normative background of a future society, i.e. the wished-for social norms and modes of participation which might (de)motivate individual training at present is rarely being asked. It is an old topos of utopian thinking, though, ranging from the works of Francis Bacon to those of Karl Marx, Karl Mannheim and Ernst Bloch (Bloch, 1964; Zyber, 2007). From both an ethical and an anthropological point of view, social utopias and technological utopias cannot be discussed separately. As technological innovation itself is envisioned to pay off for society, its idealtypic member in modernity is constituted as both *Homo faber* and *Homo utopicus*. Thus, envisioning a better future and actively working for it go hand in hand (Ropohl, 2008). Conversely, denying the chance of a better future is accompanied by a loss of motivation to design, to work and to consume – an increasing phenomenon that is regarded as pathological and termed as "depression" in highly industrialized societies.

Therefore, envisioned futures are important also in a socioeconomic sense. In reality, better futures are not merely "there" for everyone, due to economic, political and social limitations; rather they are actively generated for what is regarded to be the normal consumer and market participant, mostly based on the data of present economies. This artificial genesis happens by means of science policies and technology marketing, often supported by science writers, fiction authors and film directors who are melting assumptions on "the average" (e.g. the future individual's annual income, age, body and health status and even gender) and "the exceptional" (e.g. the emergence of biotechnical enhancements). The main contents of envisioned futures (e.g., humans who live longer) are often embedded in elitist visions of the future as they are based on experts' knowledge of technologies' potentials and, e.g., the prognosis of future scarcities. This is well known from the scenarios developed with regard to sustainability issues, for instance on future resource scarcities and limits to growth.

However, there is an important difference. In the case of sustainability-related scenarios, the problem of a conglomeration of scarcities induced the setup of scenarios (and most often belong to the scenario-approach already outlined). The imported norms and values (e.g., intragenerational and/or intergenerational justice)

[2] In the German original: "Blosse Tatsachenwissenschaften machen blosse Tatsachenmenschen."

were formulated *before* the scenarios were modelled. As a consequence, the scenario allowed envisioning possible consequences of present activities under the premise of set values and empirical data (with the persistent problem of an overload of non-knowledge, particularly regarding future technologies). The result is a construct that allows to think the following: "*If* I want (= have the value that) other/future people to have at least as much freshwater and functioning ecosystems as I have at present, *then* I have to reduce my consumption of x." At the same time, at least a rudiment version of one's own values is transported onto other/future humans or generations. These values do not resist an ethical analysis.

In contrast, a scientist's scenario ("vision") which is motivated by basic research follows a different construction. There, the scenario is sort of "built around" his or her vision of future applications of this basic research, such as Eric Drexler's early vision of self-replicating and self-creating nanobots – an almost religious vision deeply connected to those of the artificial life-community. Neither one's own values nor the values of future societies necessarily have to enter; they can be an add-on, though. Conversely, they can enter unconsciously, as value-orientation is not a defined obstacle at the very beginning. Thus, visions from single experts are much more likely to be technologically determined than "public visions" of futures, or expert groups' visions. They become "technological futures", in contrast to futures bound to a value-framework in which a specific technology is embedded. Of course, a technological future affects the anthropological concept of the imagined human being ("Menschenbild"). In a technology-driven future, the human being, including his/her brain and body, will also have to be more "technically upgraded" – a fact that is supported by the imagery of science fiction.

As Grunwald (2007: 55) rightly argues, the problem of extrapolating technological and social determinisms on a time scale by means of scenario-techniques (to which the concept of vision assessment belongs) is evident. I would like to add the underestimated *biological determinisms* that can accompany both, as for instance in the case of biomedical enhancement and social neurosciences. With the assistance of experts and new technologies, one's life can, on the one hand, theoretically be enhanced in very different ways. On the other hand, the theoretical vision of one's own technically enhanced future has to be adaptive to specific social contexts which might not be covered by scientific debates. Present norms, related to normative concepts of nature, technology and society, can function as exclusionary filters for better futures for some members of society, as the following example shows.

8.3 Example: The Social Boundaries of Techno-Doping

Since August 2007, the South-African sprinter Oscar Pistorius has been in the public media nearly every day. The 22 year old Pistorius, also known as "the fastest man on no legs", was born without fibula bones. As a consequence, an amputation of his lower legs was performed while he was still a baby. Nevertheless, by means of specially designed carbon fibre prostheses, he is able to walk; he is even

able to run about as fast as "normal" sprinters who were at that time qualifying for the Olympic Games in Beijing 2008. Athlete Pistorius is holding the double amputee Paralympics' world records for 100, 200 and 400 meters. In 2008, he wanted to rise to a new challenge. However, the Olympic Committee and the public media have been raising doubts about this new kind of "techno-doping", enabling a disabled person of the Paralympics to become a sportsman within the Olympics, i.e. the contest for the "normal".

While I am writing this article, the discussion whether Pistorius should be allowed to take part in the Olympic Games or not is still vivid. A decision of the *International Association of Athletics Federations* (IAAF) on 14 January 2008 prohibited Pistorius from taking part in the Olympics. German scientists had analyzed his prostheses in order to find out if he might have an unfair advantage over able-bodied athletes (e.g. concerning the width of his steps). In fact, they found he indeed does have a "mechanical advantage", as Gert-Peter Brueggemann, professor for biomechanics in Cologne, had put it in his expertise in November 2007. On 16 May 2008, the IAAF-decision was revised by the *Court of Arbitration for Sport*, pointing out, however, a lack of sufficient evidence for the proclaimed mechanical advantage and thus allowing Pistorius to be eligible for the Olympics. In the remaining six weeks, Pistorius failed in qualifying for the Olympics 2008. He again took, however, the gold medals for the 100, 200 and 400 meters sprints in the Paralympics 2008.

Whereas most officials of sports, allied with scientists, argue that the purity of sports, its fairness, and its ideal of nature is endangered, the public opinion on the Pistorius-case is completely different. That a disabled person is able to overcome certain natural limits by means of technology and gain more ranges of personal freedom cannot be wrong – and: isn't this the crucial argument for the vision of technological progress and human enhancement? Obviously, public acceptance of technically enhancing potentials contrasts normative ideals of (professional) sports and nature, moreover starting at the time when the *Tour de France* continuously offered an unseen amount of doping cases (see Wehling, 2003). The societal consequences of this schizophrenic situation (not only) within sports, i.e. to reject specific bodily interfaces of humans and artifacts (Orland, 2005) but, at the same time, to establish human biofacts (Karafyllis, 2003) by constantly manipulating biological growth, fitness and development, remains to be seen. The same ambivalence is true for science and technology. Some scientists support highly normative concepts such as posthuman idealtypes of power, success and well-being on the species level, whereas in the individual case of Pistorius it is the scientists who are engaged in proving and maintaining an athlete to be "normal". One might argue that professional sport is a very special case. However, it functions as an illustrating example of how public acceptance in the case of technical innovations can differ from experts' and strategic elite's opinions, particularly based on normative grounds. This type of public disagreement that is not (or insufficient) backed-up by scientific analysis can be found in other technology discourses, not last in the ones about genetic engineering of food crops in Europe which differ from the ones in the USA (Levidow, 2001).

One reason for giving doping a preference above techno-doping is related to phenomenology, highlighting the immanence of the human body. As long as technology does not show *phenomenally*, as in the case of conventional doping and genetic engineering of living beings, it *might* still be nature. But techno-doping shows the artificial element, the prosthesis – a reason why Pistorius is also called "blade runner" by the public media, which reminds one of the famous 1982 science fiction movie with Harrison Ford as cyborg-protagonist. Science, including science-driven sports, seems to accept the aforementioned self-deception more than the (Western) public itself. In a Habermasian perspective, this special way of self-deception among, e.g. scientists, is one limit to public discourse as it undermines standards for a "self-correcting learning process" (Habermas, 2005: 89–91).

This relates to one anthropological thesis of this essay, highlighting that the phenomenal visibility of the border between nature and technology is anthropologically important; moreover, that neither the concept of nature nor technology alone can form a normative argument, rather than its specific relation when being applied to a social context. When vision assessment ignores this immanence of the visible it can hardly give recommendations for future public perceptions or even visions of future actors in science and technology. With its present methodology, vision assessment excludes the phenomenal sphere of the life world, where technology can affect the senses. In the case of nanotechnology, this invisibility of technical interference with nature might lead to a similar resistance as has already emerged against genetic engineering.

I would like to point out one sentence which Pistorius said quite angrily to the journalist Judith Reker (of the Swiss weekly DAS MAGAZIN 34, 2007),[3] after being asked if he identifies himself as a "cyborg". Pistorius responded: "*I hate Science Fiction*." For him, the problems are more than real, and his disability is a fact, no fiction. The open question is, how different parts of society deal with it in regard to different contexts. Pistorius argues that the prostheses enable him to walk in normal everyday life, so why can a high-tech version of this not be used to run in a sports competition? Why isn't he allowed to enhance himself?

As the Pistorius example should exemplify, in technology discourses the reference to "nature" is most often used in a twofold sense: to overcome the limits of nature and to maintain nature as an inert reference at the same time. So inferring about "nature" implies both inclusionary and exclusionary elements, the latter relating to social stratification and status that should appear to be "naturally" given. In consequence, one could estimate a future society that strictly votes for the visibility of technology as nature's "other" (Ropohl, 1983), and thus resists technologies on the nano-scale. Therefore, as philosophers engaged in technology assessment we have to ask reversely: Which and whose visions are *not* assessed by vision assessment, though they might be already debated in public and even in science? To answer this question in the next part, we will take a look at the role of journalists and science writers and, above all, at the successful genre of pop science.

---

[3] Judith Reker. "These und Prothese", *DAS MAGAZIN* 34, August 2007.

8.4 The Genre of Pop Science and the Genesis of "Public" Visions

Public media play a key role in communicating science and shaping public acceptance of specific scientific theories, as sociologist Peter Wehling (2006: 254f) argues. Media convey and transform the knowledge claims and ignorance claims of science. Thus, they model systemic knowns and unknowns at the public level by means of filtering what is, presumably for everybody, relevant to know and what is not. With reference to *memoro*-politics[4] (Hacking, 1998: 215ff): They also model what has to be remembered and what can be forgotten. According to S. Holly Stocking (1998: 169), journalists' own *interests* are a crucial part of this filtering process. Journalists are not only watchdogs but also gatekeepers, often even facilitators for transporting information to the public sphere. This personal handwriting which is normatively inscribed in a journalist's work is but one of the reasons why an assessment of visions that have been transported by journalists has to fail for the overall aim of TA.

On the content level, the modern idea of the *public relevance* of scientific knowledge and nonknowledge is based on heterogeneous concepts such as progress and innovation, economic well-being, health, social justice, security, and human nature, all of which have to be mediated as particularly relevant for individuals in order to be read and "taken in." Scientific knowledge of the human brain seems to assemble all of the above-mentioned target concepts for public relevance. Connecting scientific propositions with terms traded in already existing "linguistic markets" (Bourdieu, 2005) and cultural stereotypes (as part of an act called "framing") eases public acceptance by upholding the notion that laypeople and experts do not differ substantially in their scientific knowledge capacities. Referring to the institutional dimension of science, this notion can be used for conveying the impression of existing participatory elements in political decision making concerning new (and old) research areas in science and technology ("scientific citizenship"; see Wehling, 2006: 259). Since it itself already uses linguistic market terms (e.g., "cyborg", "future", "health"), symbolic images (e.g., neuroimages and images of nano-landscapes), and stereotypes (man/woman, female/male, black/white), the mediating process gains in efficiency and effectiveness, which is also necessary for successful science fundraising.

However, this efficiency contains reactionary elements, and it functions at the cost of envisioning a normatively "better" society by means of scientific progress. Hence, it undermines the idea of cultural progress. In addition: that many individuals resist the lure of a discourse on a "future technology" might not always be a result of ignorance. Public ignorance of science and technology can be a social reflection of this contradiction and normative disillusion rather than merely exemplifying the "knowledge deficits" of society. The crucial question for processing

[4] *Memoro*-politics relate to the cultural history of the concept of soul and are empowered by the introjectionistic idea that there is a deeper (and higher) knowledge of the self than is actually experienced in everyday life.

effective science policies is: How can these "ignorant" people be addressed in order to become part of a "public discourse"? One possible method is to blend the border between facts and fiction by authority, as is happening in pieces of work that experts themselves have written and that fall into the category of pop science.

The particular influence of pop-science literature – written by scientists rather than journalists – on the mediating process of scientific knowledge has not yet been sufficiently scrutinized, and I can do nothing but offer a rough sketch here. The genre of pop science developed back in the middle of the nineteenth century, and was related to the new concept of public life (and its counterpart: private life), as well as to the ideas of publicity and civil society. It was instantiated by an enormously expanding newspaper and journal market for mass distribution and the building of museums. From its very beginning, scientists trivialized pop science because of its simplifications and viewed pop science as an add-on within an imagined two-phase approach of scientific writing, which, nevertheless, was regarded as necessary for the advancement of science. In fact, science and pop-science literature offer complementary ways of understanding scientific knowledge. Pop science established unique transformations of knowledge (Daum, 2002: 26), particularly related to holistic ideas of "nature" and the "world," and it still does. According to Andreas Daum, European pop science started (at least in the German and French contexts) with Alexander von Humboldt's *Kosmos* (1 vol. 1845), at about the same time as the 1848 revolution was breaking out. At present, "nanoworlds" and the world of "the" brain is under pop-scientific scrutiny. Present examples of this genre are the pop-scientific essays on neuroscience, artificial intelligence, evolutionary theory and other "hot topics" which have been written by "big name" scientists and, for example, merged into John Brockman's editorial works (e.g, Brockman, 1995, 2006). The US-American literary agent and media activist John Brockman puts forth pop science by means of a specific marketing discourse concerned with two questions: First, what is science? Second, what is reality? He is on a mission to implement a "third culture" between sciences and humanities (the latter of which he ignores, however), and this third culture seems to be ruled by journalists and science writers. In terms of media studies, Brockman does *agenda setting*. One of his admirers, Frank Schirrmacher who is editor of the influential daily newspaper *Frankfurter Allgemeine Zeitung*, followed his example in the German contexts and published bestselling books on Darwinism, the aging society, and nanotechnology (Schirrmacher, 2001).

In the pop-science literature, the author speaks both as scientist and journalist. In so doing, he or she is able to perform a double filtering process of knowledge: First, the author speaks as scientist (backed by his or her personal authority), clearly describing what is known within his or her scientific community (and, implicitly, what is not known); second, the author gives voice to the lay reader, individually selecting what is important (according to the author's set of norms and values) to know and reframing this knowledge within the cultural perspective as to *why* it is worth knowing. Descriptive and normative arguments are intertwined. Typical writing styles of pop science include a first-person perspective (singular: "I") in the introductory passages to stress scientific authority, a third-person perspective

in the middle section as typical for classical science writing, implying an objective meta-level, and, again, a first-person perspective at the end (both singular and plural, and the "we" most often outweighs the "I"). As a result of this rhetorical strategy, both the "known unknowns" (disguised by generalizations and abstractions during scientific modeling) and the "unknown unknowns" of science (often related to scientific paradigms, in the tradition of Thomas S. Kuhn) are less likely to become part of public "ignorance claims" (Wehling, 2006), compared to the mediating process instantiated by serious science journalists. It is important to remember that science, in general, is not only exporting terms and symbols from laboratories into the public sphere but is also importing them from public issues which crop up in society, for example, in social movements (e.g., the animal rights movements, or the women's rights movements) or religions. In pop-science literature, both scientific knowledge and ignorance can be actively constructed in tacit accordance with features of existing social and political conditions – for instance, concerning issues of social exclusion and divisions of labor.

8.5 Example: Social Neurosciences and "the Future" of Society

The purpose of this second example is to broaden the understanding of strategies of defining and controlling *one* future for *one* society, and to show how pop science contributes to the legitimization of this totalitarian type of research. So far, there is no TA or Science Assessment of Social Neurosciences. However, there is a valuable TA-study on Neuroimaging (Hüsing et al., 2006). In addition, topics covered by Social Neurosciences receive vast media support. Stories about "the" brain – an entity described as a single world of its own – sell, as they seem to address social relationships like love, sexuality and stigmatization (see Karafyllis and Ulshöfer, 2008 for details). We might imagine the day when the visions of both experts and journalists about "the brain future" find their way into vision assessment. So, let me contribute to a "vision assessment of vision assessment" and look at the normative settings of today's research in the Social Neurosciences.

Social neuroscience can be highlighted as an example of basic research, where the design of a *future society* is already inscribed in early stages of research. Other than in the case of nanotechnology, there definitely is a discourse on "the brain", supported both by scientists, the media, and researchers from the humanities. As in the case of nanotechnology, however, scientists continue to provide colorful and symbolically contaminated images (mostly deriving from the method of fMRI – functional magnetic resonance imaging) that are printed in daily newspapers, weekly journals and shown on television. However, an academic debate on the visions of researchers engaged in Social Neuroscience, that is *their visions* of how a future society should be, is still lacking.

This lack does not become obvious because the brain seems to be omnipresent in public debate already. Claudia Wassmann argues that the brain became an *icon* as a normative instance especially in the years 1984–2002, due to several highly

recognized TV programs on brain scans: "a gap has opened between the representation of brain imaging in the lay press and the properties brain scans acquired within the neurosciences. This gap has widened since the beginning of the new century" (Wassmann, 2007: 153). It is the idea that one is responsible for both one's brain future and one's own social future, for instance by means of lifelong learning and training of one's emotional intelligence, which is normatively relevant.

Within the same time period, the brain as icon has also become influential in the humanities and social sciences, as the emergence of new sub-disciplines show (e.g., neuroethics, neurophilosophy, neurotheology, neuroeconomy). There, the relationship between emotion and intelligence, the former topoi of the (potentially) irrational and the rational, are being (or already have been) reconfigured. According to neuroscientists (e.g., Damasio, 1994), emotions now seem to have an original cognitive content and ensure rationality, at least in the brains of "normal" people. "Cognition", in the cognitive sciences, has a meaning which is quite different from its typical understanding within philosophy (i.e. a conceptual and propositional structure). Rather, in the cognitive sciences, it is "used for any kind of mental operation or structure that can be studied in precise terms" (Lakoff and Johnson, 1999: 11). In this view, a "cognitive unconscious" exists, which, moreover, opposes psychological traditions and their ontologies of soul and mind. In recent years, models and terms from the field of neurosciences and cognitive sciences have colonized the epistemic cultures of many other disciplines, in the process transforming some of their ideas about what is normal, what is human, and, not least, what determines a functioning society.

For instance, in a recent study entitled *The good, the bad, and the ugly: An fMRI investigation of the functional anatomical correlates of stigma*, neuroscientist Anne C. Krendl and her colleagues analyzed how feelings of disgust towards socially stigmatized groups are represented in neuroimages with reference to the brain's capacity for controlling this disgust (Krendl et al., 2006). I explicitly chose this study to exemplify how transformations of models and terms from both sociology and the social world take place in the laboratory of the social cognitive neurosciences, as its experimental design is very thoughtfully conceived. Although these researchers are particularly sensitive to the underlying biases of Social Neurosciences and are aware of the impact of these biases on society, the study shows how difficult it is, methodologically, to keep to one's own normative premises. In general, for social neuroscience the problem arises that if you want to measure the process of stigmatization (i.e. a categorization) in the brain, you must define a priori what stigma (i.e. a category) is and in which brain area(s) it might show, for instance, in the area responsible for feeling disgust.

The general hypothesis of Krendl's et al. study is that control and disgust refer to two separate neural systems. The amygdala, the "organ" of emotions, is involved in the areas responsible for "feeling" disgust. The scientists were interested in both modeling and understanding how the process of social categorization takes place. Twenty-eight students were recruited from Dartmouth College (New Hampshire, USA). The implicit architecture of the experimental design, e.g. regarding the level

of students' familiarity with the stigma-type, which I will not discuss here, was quite challenging.

Photographs of persons who had one of the following attributes: *being obese*, *extensively face-pierced*, *transsexual*, *generally unattractive* were selected and shown to students (both men and women) at random. They were supposed to rank the intensity of their feelings of disgust for each face in the photos on a special disgust scale. Previously, a scale for general attractiveness ("likeability") had been developed for every single volunteer, based on individual evaluative ratings of photographs with "control faces", in order to compare the brain condition for each individual when looking at photographs of "normal" people with the condition which developed while looking at the photos of the stigmatized. *That* these people are generally stigmatized was the scientists' decision (i.e. it was their categorization), even if Krendl et al. claimed that the chosen stigma categories are "widely acknowledged" (Krendl et al., 2006: 7). Their assessment was accompanied by the decision to take the photographs from social platforms of self-addressed groups, such as webpages of piercing artists or a dating website for overweight people. This means that they selected photographs of people who identified themselves as obese, pierced, transsexual, or unattractive, which does not necessarily mean that they view themselves as socially stigmatized.

Put in philosophical terms, in the study of Krendl et al., *acceptability* was modeled neuroscientifically, and this is a category of ethics. The category of "stereotype" (i.e. *acceptance*), on which, according to Krendl et al., social stigma is based, was considered as given. Here, already, the terms used are important for modeling knowledge, as the experimental use of "stereotype" provides a reference to previous neuroimaging studies on the stigmatization of race (to which Krendl et al. refer). This is what Ian Hacking – referring to Nicholas Jardine – called the "calibration" of instruments within scientific developments (Hacking, 1998: 98), i.e. that every new method introduced for measurement has to be calibrated against the old one, including the evaluation of how adequate the old one was. In psychological terminology, the concept of the "calibration of instruments" (e.g. clocks) is known as *validation* (of tests and questionnaires, i.e. constructs), which leads to other problems. Obviously, there is no awareness on the part of some scientists within the social neurosciences to calibrate their instruments, meaning *social concepts* like stigma and stereotype, with sociology or philosophy.

I argue that concepts like stereotype can be *instruments* in social cognitive neuroscience. They make it possible to technically generate hypotheses in the context of cognition and emotion, and these concepts differ from the concepts which result from the experiment. It is important to notice that this transformation differs from metaphorical use and the science/society cross-border trade of metaphors. How can concepts be instruments? Because within the experiment, their real meaning seems to be irrelevant. Instead, they just lend the experiment their linguistic skeleton, purified of metaphorical and social meanings and implications, for the purpose of social cognitive science. A cultural and social concept, like stereotype, which is binary coded, can be an extremely useful instrument on the practical level of doing science, because it can be combined with attributes which are also binary coded.

There is a clearly-established cultural stereotype of black and white, likewise of man and woman. Concerning the "old" and still intensively used instrument, i.e. the stereotype black/white applied in stigmatization studies of race, the concept of stereotype somehow made sense (this is not to suggest, however, that it made sense in the studies in which it was employed). But a stereotypic structure is not obvious in fat/slim, extensively pierced/not at all or not extensively pierced, unattractive/ attractive. They relate to aesthetic categories, which are highly heterogeneous. There is no objective beauty, moreover which is not related to a type, and even the idea of ugliness does not contradict the idea of attraction. All chosen types involve *continua*, and are not discrete attributes. Of course, black/white also involves continua of color, but color *can be more easily stereotyped*, e.g. by scientists' choice of photographs, than attractiveness. And what is the binary other of transsexual? Not transsexual? Heterosexual? Same sex sexual? Taking a closer look, it appears that this study does not focus on stereotypes but on *normality* and its opposite, the construction of abnormality. This is an important difference. And it makes all the difference regarding the question of *who* shapes this normality – science or society, the scientist with her abstract categories or the individual within her life world of personal experiences?

During the experimental process, and by means of several abstractions and generalizations, predefined attributes of individuals' faces on photographs were converted into properties of members of social groups. On the other hand, the idea of a social group emerged because one single attribute was seen as principal and thus made the essence of this social group. The social world was remodeled. Within the laboratory context, the individuals in the photographs became "targets" of social stigma, whereas the members of the indicated social groups became "bearers" of social stigma. The volunteers in the laboratory became "perceivers" of social stigma, and the photographs themselves "stimulus materials". This setup is not an exception but the normal approach and terminology for social cognitive neuroscience.

According to Krendl and her colleagues, the disgust inspired by obesity is much more controlled within the students' brains than disgust towards transsexuality, i.e. seeing a photograph of a transsexual feels more disgusting than seeing a photograph of an overweight person. The most disgusting of all is to perceive general unattractiveness.

We can explore Krendl et al.'s study a bit further, with regard to the methodology and epistemology of social neuroscience. They assert that their study is, firstly, inspired by the awareness that social neuroscience studies on social categorization and stigmatization have predominantly dealt with race differences (see Phelps and Thomas, 2003 for a critical overview). Secondly, their study represents a critique of previous studies of social stigma which had resulted in the concept that theoretically controllable stigmas (such as obesity) lead to more negative feelings in the receiver than uncontrollable stigmas (e.g. blindness). Again, in philosophical terms Krendl et al. tried to reject the idea that *blame* and *guilt*, which have been lurking behind the concept of controllability, are involved in stigmatization. This complex field of guilt and visible stigma, which was analyzed in the laboratory, imported a specific Christian tradition into the scientific modeling. The differentiation between

controllable stigma and that which is "given" has a long religious history, distinguishing a stigmatized person from others on the basis of a bodily wound. Having biblical origin, stigmata in the Catholic tradition refer to marks on the body which resemble the wounds of Jesus received while hanging on the cross – like wounds on feet and hands. In religious tradition, stigmata are not a sign of guilt but show a co-suffering with Jesus Christ and the capability of bearing the sins of others.

According to Krendl et al., they chose the attributes of obesity and transsexuality because these attributes are ambiguous, whereas piercings are clearly controllable, and unattractiveness is determined by genetics and is thus uncontrollable. The category of "control" is used variously in the social neurosciences: first, regarding personal and social behavior; second, regarding the control of emotions and emotional areas inside the brain by other, more "intelligent" areas (mainly the prefrontal cortex); and, third, regarding the present and future social response (of groups) and emotional response (of individuals) to the perceived behavior of others.

A scientific target of social neuroscience is the unveiling of political correctness rhetoric, or other forms of learned control, and to determine whether the evaluation given in the questionnaire (behavioral data) contradicts the neuroimages (fMRI data) or confirms them. These two different measurement methods are newly referred to as "explicit measures" (questionnaires) and "implicit measures" (neuroimages), thereby tacitly abolishing or at least reducing the implicit measurement methods referring to unintentional bias from psychology (see also Phelps and Thomas, 2003: 756). Unconsciousness wins over unintentionality. In the end, this issue is about the definition of truth and which science holds the greatest claim to defining it. And since emotions still seem to evidence the innocent nature of the uncivilized animal in us, which cannot lie, the neuroimages are thought to represent the "original" and true feeling of the reptile mind. It's a jungle out there, in the brain. Of course, the wildlife of the amygdala can be tamed by the civilized brain areas responsible for evaluation and emotional learning. As a consequence, *social norms* seem to be inscribed in the brain, somehow governing its cruel "nature". The dialogic nature of social norms which are practiced in social life is completely ignored. Above all, the *vision of a better future* in which today's stigmatized persons are no longer victims of social exclusion is dispensable.

Not surprisingly, in the discussion section of their paper Krendl et al. offer the opinion that, "Over the course of evolution, the avoidance of those possessing stigma may have been adaptive" (Krendl et al., 2006: 12). As neuroscience is rooted in biology, and biology in Darwinism, this phrase "Over the course of evolution x may have been adaptive" seems to be unavoidable. It is like the final scene in a Western movie regardless of the story; the cowboys are on their horses, riding into the sunset: their work is now done.

In the mentioned study, there are no visions assessed. Rather, the scientists' visions of a present society (including stigmatized people) and of a future society (with, according to biological determinisms, a smaller number of stigmatized people than today) are inscribed into the very heart of the experimental setup. As in the case of vision assessment of nanotechnology, exclusionary mechanisms regarding participatory elements manifest at the very beginning, i.e. already in the first hypothesis

of the experiment. Could vision assessment, based on expert interviews and media analysis, detect these types of primary normative settings?

An ethicist might ask: For what purpose is this kind of research carried out and funded? Does it matter that the technology for neuroimaging already exists, and that financial investments have to pay off, i.e. that as many people as possible should be brainscanned, no matter what the reason? Or is the aim to control society for the sake of a better future for the "normal" majority of its members, reversely putting pressure on the present individual not to become a member of a (possibly) stigmatized group?

Ethicists' aim should be to reflect on a new area instead of labelling themselves as "neuroethicist" or "nanoethicists", thereby accepting the propria "neuro" and "nano" as essential. *The area encompassed could be described as the brain culture of individuals who rationalize their "self" while still believing in their unique personality*.

To sum up: In the mentioned experiment of social neuroscience, neither the individual nor the society is envisioned to have an open future. "Future" only means the future presence of the human species, organized in distinct populations. Social determinisms and biological determinisms are fused and thereby the human ability to make moral judgments is being abolished.

8.6 Outlook: The Future of Vision Assessment

The main argument of this essay was that envisioning one's own future is based on the belief in individual potentials and sharply contrasts the idea of artificially designing "the" future of "the" society. As up to now there is no definite methodology for vision assessment, it allows for thinking of methodological instruments ranging from *discourse analysis* (a concept widely used in the humanities) to *simulation* (a concept widely used in engineering) to politically initiated, *public discourses* which follow defined and institutionalized discourse rules. Admittedly, none of these instruments can empirically certify full inclusion and equality, but regarding both they offer different shades of grey between the black pole of expert-elitism and the white pole of politically engaging the whole society. Naturally, the sources and media that are being used in these approaches substantially differ, as do the chances for addressing pluralism. At present, vision assessment-techniques combine media analysis (of daily newspapers, cinema productions and TV programs), pop science writings, and expert interviews. From a systematic point of view, each of these sources is already a meta-source and has to be evaluated with regard to its unique normative settings and implications. From a sociological point of view, each of them has to be analyzed as to which social groups (or even strata) are included and excluded (or purposely exclude themselves) from the media, thereby also taking gender and race aspects into account. From a historical point of view, vision assessment is unlikely to serve as a prognostic method. As historians of science (e.g., Hård and Jamison, 2005) point out, technological innovations of the past were rather supported by "small narrations" than by big visions or projected futures.

When vision assessment is used as a tool for TA, the question as to whether the "imagined" technology to be assessed should fall into the TA-category of "technology-induced" or "problem-solving" is likely to be ignored. This omission imposes ethical problems and normative short-sights. In the particular case of nanotechnology, the question *which* problems should be solved with the help of nanotechnological means is not even asked. As a theoretical consequence, the question of present and future needs, and the turn to possible alternatives, is also neglected in this approach. Instead of correlating the envisioned means with preferable ends in the ethical sense (e.g., taking arguments of justice and participation into account), vision assessment paves the way for science actors that search for accepted ends even before the means (i.e. the technical prototypes) allow them to deliver reliable data if they in fact *are* reliable, i.e. if they in fact *are* means. By vision assessment, public acceptance is empirically tested even before a normative judgment on the acceptability of a technology could have taken place. Therefore, it remains unclear if the concept of vision assessment is part of a science *of* (mentally) designed futures or *for* (really) designing *a* future that is envisioned to be "better" for all.

This methodological weakness might turn out as strategic back-up for a science policy that forces scientists who are engaged in basic research into early visions of application and marketing. At the same time, such a science policy would weaken the position of ethicists and TA-researchers. It does not lack irony that ethicists are weakened by the encouragement to assess and evaluate something which is still a vision. In fact, by helplessly claiming (in printed form) that they cannot assess or even evaluate nanotechnology, ethicists open a fake-discourse which gives the public the impression that ethicists *can* be whistle-blowers in the nano-field already. In the long run, ethicists and TA-researchers can thus only disappoint the public – and thereby reduce the general trust in ethics and TA. On the contrary, ethicists can also say "No" to science funding and resist the lure of assessing and evaluating pure visions of powerful science actors. As a positive side effect, risk research would still be kept in the stronghold of the engaged scientists itself (who are best informed about the "facts", i.e. the malleability of the generated epistemic objects) rather than being delegated to TA-researchers who deal with categories of imagination all too soon.

Regarding the innovative process of how scientists and engineers actually create and design technologies, there is more theoretical work to be done, combined with laboratory studies (see, e.g., Banse et al., 2006; Karafyllis, 2004, 2006; Stiftung Brandenburger Tor, 2007; Poser, 2008). Basic research does not primarily seek for inventions; they more or less crop up by creative acts of tinkering. However, nowadays possible applications are also imagined "technically" (by means of simulation), thus partly reducing the imaginary element of a scientist's mind. In the 1990s, the terms "imagination" and "engineering" were coined to form the neologism "imagineering". The latter addresses artificial worlds of imagination, in which new products and production processes should be visualized (Bürdek, 2007: 348). There, ad hoc-premises are encrypted in programs which contain data for future supplies and demands. In all kind of simulation techniques, including vision assessment, the anthropological self-image is handled as astonishingly inert. Thinking of the growing

anthropological need for understanding oneself as part of nature (Karafyllis, 2008), one can doubt if a technologically upgraded individual of the future is the best guess. In general, anthropologies have a dialectic structure (Latour, 1995), i.e. one cannot proclaim a "technological future" without referring to a "natural future". In addition, psychological boundaries for imagining and enduring possible futures have not yet hit the radar of the proponents of vision assessment.

Vision assessment as a concept for TA might be useful in combination with media ethics (Funiok, 2007), science ethics (Düwell, 2004) and business ethics. Then, for instance, the normative questions would arise as to *who* is setting the agenda and *who* is funding agenda-setting for what purpose. Moreover, one might ask *who* develops persuasive images and narratives for future technologies for what purpose. Only insofar as this information is provided is a public debate about futures possible that would deserve the name "discourse". Moreover, only then can individuals differentiate between facts and fiction and are able to reflect on their own values that are basic for a better future. Therefore, vision assessment could help to deconstruct ideologies.

Obviously, nanotechnology is triggered by basic research. This insight would provide a chance for classic TA in the problem-oriented mode: asking about people's present needs and wants and how technologies are able to support and stabilize them in order to generate a better future for (ideally) everyone. Idealizing the possible extent of participation on a theoretical level, as Habermas did, has an *operative* outcome for the actual discourse; as the process of decision-making is only publicly accepted as reasonable if it is based on the avoidance of deception and exclusion. However, this insight also allows for imagining that (not only) nanotechnology might be dispensable for a better future – depending, of course, on who you ask.

Acknowledgment I thank the editors of this volume, two anonymous reviewers and my colleague Dr. Duane Lacey for helpful suggestions.

References

Bourdieu, Pierre. *Language and Symbolic Power*. Cambridge: Harvard University Press, 2005 (Reprint).

Banse, Gerhard, Armin Grunwald, Wolfgang König and Günter Ropohl (eds.). *Erkennen und Gestalten. Eine Theorie der Technikwissenschaften*. Berlin: edition sigma, 2006.

Bloch, Ernst. *Geist der Utopie*. Bearbeitete Neuauflage der zweiten Fassung von 1923. In: Gesamtausgabe, Vol. 3, Frankfurt am Main: Suhrkamp, 1964.

(BMBF) Bundesministerium für Bildung und Forschung (ed.). *Nanotechnologie pro Gesundheit: Chancen und Risiken*. Berlin 2004 (http://www.bmbf.de/pub/nano_pro_gesundheit_bericht.pdf)

Brockman, John (ed.). *The Third Culture: Beyond the Scientific Revolution*. New York: Simon & Schuster, 1995.

Brockman, John (ed.). *What Is Your Dangerous Idea? Today's Leading Thinkers on the Unthinkable*. London: Simon & Schuster, 2006.

Bürdek, Bernhard. “Was beim Prozess des Gestaltens wirklich geschieht”. In: *Bedingungen und Triebkräfte technologischer Innovationen*. Stiftung Brandenburger Tor (eds.). Berlin 2007, 337–348.

Daum, Andreas. *Wissenschaftspopularisierung im 19. Jahrhundert. Bürgerliche Kultur, naturwissenschaftliche Bildung und die deutsche Öffentlichkeit, 1848–1919*. 2nd ed. Munich: Oldenbourg, 2002.

Coenen, Christopher. “Der posthumanistische Technofuturismus in den Debatten über Nanotechnologie und Converging Technologie” In: *Nanotechnologien im Kontext*. Alfred Nordmann, Joachim Schummer and Astrid Schwartz (eds.). Berlin: Aka, 2006, 195–222.

Damasio, Antonio R. *Descartes’ Error: Emotion, Reason and the Human Brain.* New York: Grosset/Putnam, 1994.

Drexler, Eric. *Engines of Creation. The coming Era of Nanotechnology*. New York: Anchor Books, 1986.

Düwell, Marcus. “Research as a challenge for ethical reflection.” In: *Ethics of Life Scientists*, M. Korthals and R. Borgers (eds.). Dordrecht: Kluwer, 2004, 147–155.

(EGE) The European Group on Ethics in Science and New Technologies to the European Commission. *Opinion on the ethical aspects of nanomedicine*. Opinion Nr. 21 (17. Jan. 2007). Brüssel: European Commission.

(ESF) The European Science Foundation. *ESF forward look on Nanomedicine*. November 2005. http://www.esf.org/publication/214/nanomedicine.pdf (Last visit: 9 January 2006).

Funiok, Rüdiger. *Medienethik – Verantwortung in der Mediengesellschaft*. Stuttgart: Kohlhammer, 2007.

Grunwald, Armin. “Vision Assessment as a new element of the Technology Futures Analysis Toolbox”, In: *Proceedings of the EU-US Scientific Seminar: New Technology For sight, Forecasting & Assessment Methods*, Seville, May 13–14, 2004. http://www.jrc.es/projects/fta/index.htm (Last visit: February 12, 2006).

Grunwald, Armin. “Nanotechnologie als Chiffre der Zukunft.” In: *Nanotechnologien im Kontext*. Alfred Nordmann, Joachim Schummer and Astrid Schwartz (eds.). Berlin: Aka Verlag, 2006, 49–80.

Grunwald, Armin. “Umstrittene Zukünfte und rationale Abwägung. Prospektives Folgenwissen in der Technikfolgenabschätzung.” *Technikfolgenabschätzung – Theorie und Praxis* 2007, 16, 1, 54–63.

Hacking, Ian. *Rewriting the Soul: Multiple Personality and the Science of Memory*. 2nd and corr. ed. Princeton: Princeton University Press, 1998.

Habermas, Jürgen. *Toward a Rational Society*. Boston: Beacon, 1970.

Habermas, Jürgen. *Die Zukunft der menschlichen Natur.* Frankfurt am Main: Suhrkamp, 2001.

Habermas, Jürgen. *Zwischen Naturalismus und Religion*. Frankfurt am Main: Suhrkamp, 2005.

Hård, Mikael, and Andrew Jamison. *Hubris and Hybrids: a Cultural History of Technology and Science*. London: Routledge, 2005.

Hüsing, Bärbel, Lutz Jäncke, and Brigitte Tag. *Impact Assessment of Neuroimaging*. Zürich: vdf Hochschulverlag AG, 2006.

Husserl, Edmund. *Gesammelte Werke (Husserliana)*. Den Haag: Nijhoff 1950ff. (Vol. VI: *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*).

Joy, Bill. “Why the future doesn’t need us.” *Wired Magazine* 2000, 8, 4, 238–262.

Kapp, Ernst. *Grundlinien einer Philosophie der Technik*. [1877] Reprint Düsseldorf: Stern-Verlag Janssen, 1978.

Karafyllis, Nicole C. (ed.). *Biofakte. Versuch über den Menschen zwischen Artefakt und Lebewesen*. Paderborn: Mentis, 2003.

Karafyllis, Nicole C. 2004. “Zum Systemcharakter von Leitbildern in der Technikentwicklung und –gestaltung.” In: *Technik – System –Verantwortung*. Klaus Kornwachs (ed.). Münster: LIT, 2004, 485–498.

Karafyllis, Nicole C. “Notwendigkeit, Möglichkeiten und Grenzen einer Cultural Philosophy of Science.” *Erwägen Wissen Ethik* 2006, 17, 4, 613–633.

Karafyllis, Nicole C. “Bacillus nanotechnensis: Der Erreger einer ‘neuen’ Epidemie im Labor der Gegennatur-Gesellschaft und dessen Grenzen der Biofaktizität.” In: *Mensch ohne Maß? Reichweite und Grenzen anthropologischer Argumente in der biomedizinischen Ethik*. Giovanni Maio, Jens Clausen, Oliver Müller (eds.). Freiburg: Alber, 2008, 356–387.

Karafyllis, Nicole C. and Gotlind Ulshöfer (eds.). *Sexualized Brains. Scientific Modeling of Emotional Intelligence from a Cultural Perspective*. Cambridge, Mass.: MIT Press, 2008.

Krendl, Anne C., Neil Macrae, William M. Kelly, Jonathan A. Fugelsang, and Todd F. Heatherton. “The good, the bad, and the ugly: An fMRI investigation of the functional anatomic correlates of stigma.” *Social Neuroscience* 2006, 1, 1, 5–15.

Lakoff, George, and Mark Johnson. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books, 1999.

Latour, Bruno. *Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie*. Berlin: Akademie-Verlag, 1995.

Levidow, Les. “Precautionary Uncertainty: Regulating GM Crops in Europe.” *Social Studies of Science* 2001, 3, 842–874.

Lösch, Andreas. “Anticipating the Futures of Nanotechnology: Visionary Images as Means of Communication.” Technology Analysis & Strategic Management (TASM) 18, 3/4, 2006, Special Issue on the Sociology of Expectations in Science and Technology, 393–409.

Nordmann, Alfred, Joachim Schummer and Astrid Schwarz (eds.). *Nanotechnologien im Kontext*. Berlin: Aka Verlag, 2006.

Nordmann, Alfred. “If and Then: A Critique of Speculative NanoEthics.” *NanoEthics* 2007, 1, 1, 31–46.

Nordmann, Alfred. “No Future for Nanotechnology? Historical Development vs. Global Expansion.” In: *Nanotechnology and Nanoethics: Framing the Field*. Fabrice Jotterand (ed.). Dordrecht: Springer, 2008 (in press).

Orland, Barbara (ed.). *Artifizielle Körper – Lebendige Technik. Technische Modellierungen des Körpers in historischer Perspektive*. Zurich: Chronos, 2005.

Pethes, Nicolai, and Silke Schicktanz (eds.). *Sexualität als Experiment? Identität, Lust und Reproduktion zwischen ‚Science’ und ‚Fiction’*. Frankfurt am Main: Campus, 2008.

Phelps, Elizabeth, and Laura A. Thomas. Race, behavior, and the brain: The role of neuroimaging in understanding complex social behaviors. *Political Psychology* 2003, 24, 4: 747–758.

Poser, Hans. *Herausforderung Technik*. Frankfurt am Main: Peter Lang, 2008.

Ropohl, Günter. “Technik als Gegennatur.” In: *Natur als Gegenwelt*. Götz Großklaus, Ernst Oldemeyer (eds.). Karlsruhe: von Loeper 1983, 87–100

Ropohl, Günter. “Homo Faber: Die Kunst des Machens.” In: *Der Mensch – ein kreatives Wesen?* Heinrich Schmidinger, Clemens Sedmak (eds.). Darmstadt: WGB 2008, 259–274.

Saage, Richard. *Utopische Profile*. Vol. I–IV. Münster: Lit, 2001–2004.

Schaper-Rinkel, Petra. “Technik, Wissen und Macht in Utopien und Zukunftsvorstellungen der Frühen Neuzeit.” In: *Technik in der Frühen Neuzeit – Schrittmacher der europäischen Moderne*. Engel, Gisela and Karafyllis, Nicole C. (eds.). Frankfurt am Main: Vittorio Klostermann, 2004. 245–259.

Schmid, Günther, et al. (eds.). *Nanotechnology. Assessment and Perspectives*. Berlin and Heidelberg: Springer, 2006.

Schirrmacher, Frank. *Die Darwin AG. Wie Nanotechnologie, Biotechnologie und Computer den neuen Menschen träumen*. Cologne: Kiepenheuer & Witsch, 2001.

Skorupinski, Barbara and Konrad Ott. *Technikfolgenabschätzung und Ethik*. Zurich: VDF, 2000.

Stocking, S. Holly. “On drawing attention to ignorance”. *Science Communication* 1998, 20, 165–178.

TA-SWISS (ed.). *Nano! Nanu? Informationsbroschüre zum publifocus ‚Nanotechnologien und ihre Bedeutung für Gesundheit und Umwelt’*. Bern: BBL, 2006.

VDI (Verein Deutscher Ingenieure). *Technikbewertung – Begriffe und Grundlagen*. VDI Report 15. Düsseldorf: VDI, 1991.

Wehling, Peter. “Schneller, höher, stärker – mit künstlichen Muskelpaketen. “ In: *Biofakte*. Karafyllis (ed.). Paderborn: Mentis 2003, 85–100.

Wassmann, Claudia. “The Brain as Icon – Reflections on the Representation of Brain Imaging on American Television, 1984–2002.” In: *Tensions and Convergences. Technological and Aesthetical (Trans)Formations of Society.* Reinhard Heil et al. (ed.). Bielefeld: Transcript, 2007, 153–162.

Wehling, Peter. *Im Schatten des Wissens?* Konstanz: UVK, 2006.

Zyber, Erik. *Homo utopicus. Die Utopie im Lichte der philosophischen Anthropologie*. Würzburg: Königshausen & Neumann, 2007.

Chapter 9
Exploring Techno-Moral Change: The Case of the ObesityPill

Tsjalling Swierstra, Dirk Stemerding and Marianne Boenink

Abstract Technology is a major force in modern societies, co-shaping most of its aspects, including established moral norms and values. Technology Assessment aims to explore the consequences of New and Emerging Science and Technology [NEST] in advance, to help create better technology. This article develops a method for enhancing our moral imagination with regard to future techno-moral change. At the core of this method lies so-called NEST-ethics, the argumentative patterns and tropes that constitute the 'grammar' of ethical discussions about emerging technologies. This grammar can be applied to explore at forehand the moral controversies and even the moral changes that are provoked by these technologies. In the form of alternative techno-moral scenarios these explorations can be used to inform and enhance public deliberation on the desirability of the NEST in question. This results in a type of ethical TA that is self-reflective regarding its own moral standards. To illustrate our method, we offer 'fragments' of a techno-moral scenario on the moral consequences of the introduction of a future ObesityPill.

Keywords Techno-moral scenarios · Ethical technology assessment

Technology has developed into a major force in modern societies, now co-shaping most aspects of it. In the words of the American pragmatist John Dewey: 'Steam and electricity have done more to alter the conditions under which men associate together than all the agencies which affected human relationships before our time.' (Dewey, 1954) (p.323) Sometimes technology leads to happy results, sometimes to less happy ones, most often to an ambiguous mix of both. From the standpoint of modern democracy, it is important that those living with the consequences of technology, the citizens, have at least some say in the direction of its further development (Bijker, 2001; Feenberg, 1999; Sclove, 1995). This citizen participation can both be argued for as being of intrinsic value – people have a right to exert democratic influence over the powers that bind them – or in more instrumental terms: mobilizing different points of view leads to better knowledge and thus to

T. Swierstra (✉)
Faculty of Behavioural Sciences, University of Twente, Enschede, The Netherlands
e-mail: t.e.swierstra@utwente.nl

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_9,

better technology. Although some argue that the technological issues and (parts of) the assessment of risks to our safety or to the environment are best left to the experts (Ferretti, 2007), few would argue that we should leave assessing the broader social and cultural impacts of new and emerging science and technology [NEST] in expert hands too. With regard to such 'soft' impacts, the input of all stakeholders is needed.

Unfortunately, we cannot foretell the future impacts of technology. Humbled by many failed attempts in the past, we have by now learned that the future is impossible to predict. Not only do we lack the necessary knowledge, but the future is essentially open and contingent on our choices, as is clear from phenomena like *self-denying* or *self-fulfilling prophecies*. Still, we cannot help preparing for it. Purposeful action assumes some degree of speculation about future impacts. A popular method of preparing for the future while at the same time acknowledging its essential uncertainty and openness, is by creating diverse scenarios: narratives evoking alternative future worlds. These scenarios are then used – among other things – to spark discussion amongst stakeholders about the (un)desirability of the NEST in question, and may help us to devise strategies that are robust in as many of the possible worlds as possible (Notten et al., 2003). Many consider the interactive *exercise* of devising such scenarios with relevant stakeholders to be even more important than its eventual *outcome*. What matters is creating a proactive and sensitive attitude amongst relevant actors.

Scenarios pertaining to NEST cover a broad range of consequences, for instance impacts on health, safety, environment, quality of labour, legal and social consequences, political and even cultural consequences. One particular type of consequence, however, has as yet received little systematic attention: the fact that *NEST regularly leads to moral change*. Technologies help change the societies in which they are introduced. This is true, even when the opposite also holds: technologies also change due to social pressures. But this so-called *co-evolution of technology and society* (Rip and Kemp, 1998) does not halt at the door of morality. Of course, if one believes that morality is fixed, universal and unchanging, technological change cannot be accompanied by moral change. But that belief is hard to sustain in the light of the historical experience. We constantly see NEST uprooting established moral routines. These disturbances manifest themselves as controversies about how to re-establish a 'fit' between NEST, our moral world and us. This can be done by adapting the technology to the relevant moral norms and values; by adapting relevant morality to the NEST in question; and by a negotiation resulting in mutual adaptation. In this way NEST can lead to moral change, although it never *determines* whether such a change will occur or the direction of that change. Given a specific moral environment, NEST can make some moral options easier to argue for and others less easy. For example: a birth control technique like the pill allows technical control of female fertility. It is clear that this has made it easier for Western women to claim their sexual autonomy, and less easy for others to argue against it. Because in this moral environment autonomy was already available as a moral value, the pill could uproot traditional sexual morals by providing women with the technological means to actually practice their sexual autonomy. This is context dependent,

of course, for in a patriarchal society, the same technology could easily worsen women's sexual subjection by making their bodies permanently available.

Please note that such *moral* consequences differ from *morally relevant* consequences. The latter do receive attention insofar many scenario exercises are im- or explicitly *normative*, aiming at some common good. When we focus on the *moral* consequences, by contradistinction, we are not primarily interested in applying moral standards to NEST, but in the opposite, *descriptive*, question: how might the NEST affect current moral standards and practices? Of course, these moral consequences will be of interest from a normative, ethical point of view too, but it is important to distinguish *describing* possible moral change from *evaluating* it.

Why is it important to explore techno-moral change in advance? What is the practical-normative relevance of techno-moral scenarios? We can see at least three reasons. It is important to explore the emotions and controversies some technologies are bound to stir up. Instead of being taken by surprise, as in the case of the unexpected European distrust of Monsanto's GM crops, policy makers can set themselves to create conditions and procedures for these ethical controversies to unfold in a fruitful way so that they will benefit collective deliberation. Furthermore, techno-moral scenarios deepen discussions about the desirability of a NEST by stimulating us not to deliberate on a NEST in isolation from society, but instead on NEST in its socially – and thus morally – embedded form. Only by looking at technology in this broad, embedded, form, we can truly evaluate its desirability. Techno-ethical scenarios can thus contribute to societal learning in relation to the introduction of new technologies. Finally, techno-moral change is a defining feature of modernity, as nineteenth century visionaries like Marx, Baudelaire and Nietzsche already stressed. Modernity is characterized by its *dynamism*, by the acute sense of everything being in a flux. In the famous words from the *Communist Manifesto*: 'All that is solid melts into air.' (Berman, 1983) Moderns are left without solid foundations or Archimedean points. Techno-moral scenarios provide citizens with valuable training to accept, and learn to deal with, this important feature of their existence.

This article then aims to answer two questions. The first one is descriptive: *how* best to explore future techno-moral change? We consider our effort to be part of what Don Ihde has dubbed a 'material hermeneutics': the exploration of future worlds co-shaped by NEST to assess whether present-day morals would still fit the new socio-technological reality (Ihde, 1998). Exploring techno-moral change turns out to be an extra-ordinarily difficult undertaking. We are so immersed in current morals, our identities are so deeply entwined with them, that most people find it hard, and distasteful or even frightening, to imagine them to change. The issue, then, is how to stimulate our *moral imagination* so as to be able to jump our 'moral shadows' as far as we can.[1] Moral imagination commonly refers to the mental faculty that allows us to empathize with other people's feelings and to assess the possible consequences of our actions (Coeckelberg, 2007; Fesmire, 2003; Johnson, 1993). We use the term in a more radical sense: the evocation of moral worlds differing

[1] Which, with the benefit of hindsight, proves to be never very far.

from ours. *Science (or technology) fiction* is well established. We are aiming for *morality fiction*.[2]

Of course, here we lack the space to develop a full techno-moral scenario, let alone several alternative ones. Instead we focus on introducing a method for, as dutch technology researcher Arie Rip calls it, 'controlled speculation' about techno-moral change. We aim for a proof of principle, rather than for full scenarios. We investigate this question in Sections 9.1, 9.2, and 9.3. In Sections 9.1 and 9.3 we introduce theoretical reflections on the relations between technology and morality to get our moral imagination going. In Sections 9.2 and 9.4 we illustrate our approach by applying these reflections to a specific case: the ObesityPill.

Even if one succeeds in developing plausible ideas about the moral consequences of a particular emerging technology, how will this then help to establish its (un)desirability? This second question is normative rather than descriptive and is dealt with in Section 9.5. We argue that imagining techno-moral change can in fact help to improve our moral deliberations about technological development. We defend a type of ethical TA that is self-reflective regarding its own moral standards.

9.1 Exploring Techno-Ethical Controversies

9.1.1 Morality and Ethics

To investigate how NEST affects morality, we first need to define 'moral' and 'ethical'. We follow John Dewey, according to whom humans are first and foremost active beings. Theoretical reflection only comes in second, when provoked by practical problems that cannot be dealt with by sticking to the established practical routines. Reflection is an instance of the typically human capacity to adapt to changing environments (Dewey, 1922; Logister, 2004).

Morals pertain to either the relations we entertain with other (usually human) beings, or to conceptions of what constitutes a good life. In everyday situations, morals exist as practical routines that are considered to be so self-evident as to hardly deserve reflection or even only articulation. As a consequence people are largely unaware of their existence and influence. These moral routines once started their existence as conscious solutions for conflicting stakeholder interests/rights or as answers to the question: what would be a good life to lead, as an individual and/or as a community? But afterwards, we obey these tacit norms and pursue these tacit values unthinkingly. For example, as Bernard Williams (1985, p.185) once pointed out, 'normal' people do not consciously decide that it is immoral to kill an obnoxious colleague. The thought does not even cross their minds. If it does, this indicates abnormality. We only become aware of moral routines when people disobey them, when conflicts between routines emerge and a moral dilemma arises,

[2] Of course, this is exactly what good SF also is.

or when they are no longer able to provide satisfactory responses to new problems. At this moment morality turns into ethics, the latter being the critical reflection on (and discussion about) the former. Whereas morality is characterized by unproblematic acceptance, ethics is marked by explicitness and controversy. Ethics is 'hot' morality; morality is 'cold' ethics. We *do* ethics when we put up moral routines for reflection, discussion and reassessment. For example: in discussions about emerging technologies, values like health, safety, sustainability and economic growth are usually 'cold'; the medical use of embryonic stem cells or the possibility of human enhancement are typically 'hot.'

9.1.2 NEST-Ethics

To avoid empty speculation, our moral imagining of the techno-moral future has to be grounded in the present. We therefore start by assuming that in future controversies the same argumentative patterns can be discerned that mark previous and current discussions about NEST. Swierstra and Rip (2007) have produced an overview of this so-called NEST-ethics. They distinguish a (synchronic) taxonomy (or grammar) of arguments and argumentative patterns, and a (diachronic) account of how NEST-controversies typically – but by no means always – unfold. In this section we first summarize part of their findings (adding some minor refinements of the original scheme along the way). We then use these patterns to imagine a future ethical controversy over an as yet non-existent medical technology: a genomics based obesity pill. To facilitate this exercise in moral imagination, we attached a code-number to each argument or trope so it can be used as a heuristic device to stimulate our imagination.

Ethical reflection on our mutual relations is usually done by looking at the consequences of actions or policies, at basic rights, duties and responsibilities and/or at the different criteria to distribute costs and benefits justly. The good life ethics is about the values that are important in life and the qualities or character dispositions one needs, to have a fair chance of realizing these values (Swierstra and Rip, 2007).[3] The argumentative patterns of NEST-ethics can thus be summarized under the headings of consequences, rights and principles, justice, and the good life.

---

[3] This distinction between four types serves the analytical purpose of identifying ethical arguments as used by real life actors, when discussing the pros and cons of a NEST. It is not meant to take a side in the debate amongst professional ethicists, since – at least – Kant, whether one monolithic ethical theory can do justice to all these types of ethical argument (for example consequentialism, or deontology) or that we need to accept moral heterogeneity in the sense that different ethical theories highlight irreconcilable dimensions of moral experience. For the latter position, which we adhere to, by the way, see: (Larmore, 1987). Because NEST-ethics serves a descriptive goal primarily, and first of all seeks to stimulate our moral imagination, the question whether different positions taken by (future, imagined) parties are rationally justified, need not concern us.

The birth of a NEST is typically heralded by [**1.a.1**] arguments that point to hoped for consequences, in the form of visions about how the NEST will increase our control over the world and thus our well-being. These promises reflect the enthusiasm of the technology developers and are functional in mobilizing financial, political and public support. Sceptics then typically question these promises along four axes [**1.b.1**]. Is the promise plausible, or are we dealing with hype and overpromising? [**1.b.2**] Even if the promise will be fulfilled, won't there be so many adverse side effects that in the final count the costs more than neutralize the benefits? These side effects are to be expected, amongst other things because we know from history that technologies are often put to quite unforeseen uses (Mackenzie and Wajcman, 1999). (The opponent will typically counter with adding a new promise: the unforeseen problems will be solved with the help of unforeseen technical solutions.) [**1.b.3**] Is there not an alternative, better, way to realize the envisioned good? [**1.b.4**] Is the envisioned good really to be considered as a good?

The second category of arguments stress fundamental principles, rights and obligations – typically siding with the individual who is in danger of being sacrificed to the collective good. Of course, rights can and are be mobilized in favour of a NEST. Some will argue [**2.a.1**] that people (or other stakeholders, like animals) possess a positive (claim)right to the new technology; others will claim more modestly [**2.a.2**] that people possess a negative right to it: they should be allowed to purchase/apply the new technology because it does not harm others. But a more typical pattern is that opponents of the NEST mobilize rights and duties to counterbalance the promises of the proponents. When rights and principles become the subject of a ethical controversy, they are typically contested along four axes. [**2.b.1**] In the first place sceptics can point out that the principle is wrong. This is very rare, but sometimes occurs in debates where multiculturalism is an issue (e.g. gender specific types of 'honour' that are crucial in one culture, null and void in the other) [**2.b.2**]. Secondly, the principle can be acknowledged in abstract, but then denied to be applicable to the NEST in question [**2.b.3**]. Thirdly, the principle can be acknowledged, but then turned around to *oppose* the conclusion of the other party [**2.b.4**]. Fourthly, the principle can be acknowledged, but deemed less pressing than another, conflicting, principle.

NEST also often raises justice issues: how to distribute the consequences – costs and benefits? Of course, this justice issue typically comes up after accepting – albeit in general terms – the NEST under discussion. The discussion focuses on the question what criterion should be accepted for this distribution: [**3.1**] equality, [**3.2**] merit, [**3.3**] need, or [**3.4**] chance. However, the promises accompanying the introduction of a NEST are as a rule couched in either (implicitly) egalitarian terms ('This is a benefit to humankind!' 'Human progress!') or in terms of need ('It is immoral to stop this technological development because it will benefit the sick and the starving.'). The discussion usually concentrates on the question how to ensure that the NEST will indeed benefit [**3.1**] all or [**3.3**] those in dire need, instead of [**3.4**] those who are simply lucky enough to afford it. Two conflicting positions can be discerned: those arguing [**3.1.a & 3.3.a**] that – after a while – the trickle down effect will ensure that the benefits will reach all/the needy, versus those arguing

[**3.1.b and 3.3.b**] that this trickle down effect does not occur without political help, at least not in such a degree that it enhances the relative position of the majority/the needy in relation to those reaping the benefits first.

Finally, NEST often also touches on issues pertaining to the good life. These are hard to categorize, because the issue of the good life is so very complex, involving core-values, virtues, expectations, ideas about fate, religion, and so on. (Dohmen, 2002) However, there is a general issue in good life ethics, and in NEST-controversies in particular, that generates recurring images and tropes. This is the issue of how to value (technical) mastery over our fate, over the external forces currently outside our grasp. As we saw above, the introduction of a new NEST is accompanied by a general promise of increased control. This is supposedly a good thing, improving the quality of our lives. Although this claim may seem commonsensical at first glance, it is hotly contested. Of course no one goes so far as to denounce all technical control. But ideas do conflict over the right degree or form of control, or over what topics one should (not) want to control.

The promoters of the NEST often draw in general terms on [**4.a.1**] a Promethean imagery, stressing that humankind should always strive forward and upward. In this vision there is little or no patience with fixed *limits*; all that is acknowledged are *frontiers* that should be *transgressed*. Part and parcel of this attitude is that one should not try to avoid all risk; some uncertainty has to be accepted. Opposing this forward-pressing spirit are sceptics and conservatives who stress the importance of knowing when and where to stop. A theme shared by these critical voices is obeisance to pre-given limits, in opposition to the ideal of technical mastery and perfection. (Kass, 1997, 2002; McGee, 2002; Sandel, 2004) This obeisance, this plea to accept as being fundamental at least some of the limits placed upon us, is argued for in various ways. Some stress the *religious* character of these limits, urging us not to play God [**4.b.1**]. Closely related are the warnings that we should stick to *natural* limits, lest we create monsters [**4.b.2**]. Others point to what one could call *anthropological* limits, stressing that complete mastery over our environment would in the end dehumanize us, draining our lives of meaning, because human beings can only flourish when reality puts up some resistance to our desires and aims [**4.b.3**]. A fourth set of limits has to do with a different type of order: the split between the social and the technical. Referring to this *ontological* difference, critics argue that it is wrong to give a technical solution to what is essentially a social problem [**4.b.4**]. Finally, *cognitive* limitations are foregrounded in the metaphor of the sorcerer's apprentice, who unleashes powers that he then is unable to control [**4.b.5**].

These are all typical and recurring ethical arguments pertaining to NEST. However, there is another cluster of NEST-arguments that does not deal directly with the new technology, but deals in more general terms with the relation between technological and moral development. First, in NEST-discussions, technological *determinists* face technological voluntarists. The first group maintains that morals cannot influence the course of technological progress, either because [**5.a.1**] its course is preordained by an internal logic, or [**5.a.2**] because of the restrictions

international competition forces upon societal actors. *Voluntarists* counter by arguing that [**5.b**] technology is influenced constantly by societal forces, and can thus be steered in morally desirable directions. Second [**6.a**], technology *optimists* confront [**6.b**] technology *pessimists*. The first see technology as the solution, the second see it as the cause of our problems. Thirdly, technologically induced moral change is differently perceived and appreciated by the proponents and opponents of the NEST under discussion. The first will stress that [**7.a.2**] the NEST in fact does not raise novel ethical issues, because it is essentially similar to technologies that have already been accepted by society. This is the argument by *precedent*. If such a precedent cannot be found, however, proponents will argue that indeed the NEST does cause moral unease, but that society will *habituate* itself quickly to the now novel, and sometimes unnerving, NEST. The opponents will argue, by contrast, that [**7.b.1**] the NEST is already *immoral as it is*, or that [**7.b.2**] it will manoeuvre us upon a *slippery slope* to moral decay.

9.2 The Obesitypill, Part I: Ethical Controversy

We realize that this quick summary of NEST-ethics is awfully abstract. We will therefore now add flesh and blood to it by imagining the ethical controversy that could plausibly accompany the birth of the *obesity* pill, a genomics based drug that would allow people to consume all they want without gaining body weight.[4] How this pill would work need not concern us here. Suffice it to say that pharmaceutical industries around the globe are frantically researching such a drug, because it would generate billions in income given the growing obesity epidemic worldwide. Again, the limited space of this article allows us only a brief sketch, as a kind of 'proof of principle' of our method. In an actual scenario study one would develop plural scenarios by modifying some key uncertainties. Furthermore, it is only to be expected that applying the NEST-ethical patterns will invide different imaginations in different people. Because the aim is to stimulate the moral imagination, not to produce truth (which is impossible anyway where the future is concerned), this diversity is to be considered an asset of our method rather than as a weakness.

The unfolding of an ethical controversy concerning a NEST can be most fruitfully constructed as a narrative with four stages: status quo, novelty, conflict (action-reaction dialectics), and closure. In this section we explore the first three stages. We discuss the fourth stage in Section 9.2.

9.2.1 The Moral Routines Before the NEST is Introduced

A NEST like the ObesityPill can be expected to uproot established moral routines pertaining to our bodyweight. So, our first step should be to map existing moral

[4] To avoid making the narrative illegible, we have deleted the codes for each argument.

routines in the relevant area.[5] In the Netherlands, two discourses currently compete for hegemony.

The first discourse stresses the responsibility of the individual, highlighting duties and the good life. Obesity is presented as an indication of a weak will as a lack of virtue and as proof that one values the wrong things in life. Each individual is considered to be under a moral obligation to maintain a healthy body weight. Other people are not required to help you. Consequentialist arguments play a subordinate role in this discourse. If justice is an issue at all, it is only raised to deny the obligation of the community to pay for diseases of weak willed individuals.

The alternative discourse shuns moralizing and is popular in policy and science circles. Now 'structure' is highlighted instead of 'agency'. The discourse is predominantly consequentialist: how to best protect people against obesity? The answer: modify the (social, economic, material) environment into one that facilitates and stimulates a healthy lifestyle. This consequentialist orientation is backed-up by considerations pertaining to social justice: the state should provide everyone with a health enhancing socio-economic-cultural-material environment. Individual duties and virtues play a subordinate role in this discourse.

From an ethical perspective both discourses mirror each other. The discourse of individual responsibility stresses duties and virtues; the rivalling discourse stresses consequences and justice. Policy makers argue against the moralizing stress on individual autonomy, as people's lifestyles are largely determined by their environment. Or, alternatively, they argue that autonomy and free choice are indeed important, but that in this case the right to lead healthy lives overrules them. Or, finally, they argue that their 'environmental' policies will in the end create the conditions necessary for individuals to act autonomously. Vice versa, proponents of the moralizing discourse argue that whatever the socio-etcetera circumstances, some people will always act wrongly because of their weak will. Or that the policymakers' cure, restricting people's autonomy by manipulating their environment, is worse than the disease because freedom outweighs physical health. Or they challenge the claim that the results of their opponents' policies are indeed valuable, e.g. by arguing that 'undeserved' leanness is not a worthy goal.

9.2.2 The Introduction of the NEST, Accompanied by Enthusing Arguments from the Technology Promoters

Now consider the introduction of the ObesityPill. How might the proponents and opponents of this particular technical device fill in the argumentative patterns characteristic of ethical debates about NEST? Let's put our moral imagination to work!

The introduction of a NEST usually rests on an implicit, but remarkable, asymmetry. Whereas many applaud scientific and technical revolutions, with moral

[5] This part of the analysis is based on Swierstra (submitted): 'From Gluttony to Obesity: Three constellations'.

revolutions the reverse is true. As pointed out above, people commonly perceive moral change as a threat to their identity. Therefore, if we may believe most participants in NEST-ethical discussions, nothing or little ever happens or should happen in the moral world. Thus, techno-scientific discontinuity is asymmetrically paired to moral continuity. This helps to explain why the instrumentalist vision of technology as a neutral means to our goals is still dominant, notwithstanding the philosophical critiques – from Heidegger to Latour – of this vision.

A new technology is typically heralded by (consequentialist) visions, expectations and promises. Plausibly then, the ObesityPill will be presented to the general public as the solution to their problems. Its proponents will stress the pill's beneficial consequences, e.g. that it will cure the obese and prevent others from becoming obese, so that society no longer will be burdened by costs generated by obesity. They might then also add some arguments pertaining to rights and obligations: whatever one's personal stance, there is no ground to deny others their right to use the pill. Some might go on by arguing that justice requires enabling everyone to purchase the pill. Of course, they will say, in the short run the pill will be costly and probably only available to the rich. However, if more and more people will start using it, prices will drop and in the end almost everybody will be in the position to purchase the ObesityPill – which will finally put an end to the obesity-gap between rich and poor. Finally some arguments from good life ethics will be appealed to. It is a common motif that humanity gradually emancipates itself from nature's shackles. By freeing us from the whims of our bodies, so they say, the pill simply adds another chapter to this long and glorious history of human emancipation.

By thus mobilizing current moral convictions and routines in favour of the new medical device, proponents help create the impression that the pill is only a neutral instrument to help realize pre-given goals. But this is not really the case, in so far as the introduction of the ObesityPill does displace the two earlier discourses about obesity. In that sense, the pill is not morally neutral at all. It mediates, if introduced on a large scale, the relation between humans and their world in a novel way.

9.2.3 Conflict

The moralizing and individualizing discourse relies heavily on arguments pertaining to duties and to the good life. So, it is easily imaginable that its defenders will be outraged by the possibility created by the ObesityPill to control our body weight through medical technology, making willpower superfluous. According to them, the pill turns a healthy (and aesthetically pleasing) body into a consumer-good, to be purchased rather than laboured for and earned. The pill provides us with technological mastery over nature, but they consider this to be the wrong type of mastery.

Adherents of this discourse might experience more difficulty with countering the argument that consumers should have the (negative) right to purchase the pill. After all, they themselves stress individual autonomy and free choice. A possible argumentative stratetegy for them would then be to deduct opposing conclusions

from the general principle of autonomy, e.g. by arguing that taking the ObesityPill contradicts one's autonomy rather than being justified by it. Autonomy manifests itself in our reason and free will, not in substituting one's dependence on the body for a new dependence on (medical) technology.

They might also point to some undesirable consequences. Even if the pill will probably reduce body weight for many, it will inevitably have nasty side effects, for example an increase in empty hedonism. People will no longer need to exert self-control now they have delegated this to the ObesityPill. The NEST will leave them morally weakened. Or they might devalue the goal of being lean. Before the introduction of the pill, they might argue, having a healthy body weight was honourable because it designated self-control. Now that it can be purchased and consumed, it loses meaning and value.

It is less likely that justice arguments will figure prominently in this critique of the ObesityPill, because the focus of this discourse is on individuals and because this discourse despises the pill and so couldn't care less about its 'fair' distribution.

This is different for the policy discourse, with its focus on consequences and justice. Its adherents might start by questioning the claim that the ObesityPill will result in fighting obesity. They might point out that the lower classes, who are most affected by obesity, cannot afford the pill, or that they lack the self-discipline to take the pill. Furthermore, they might argue, the pill is likely to have unwanted side-effects. For example, because the rich can now safeguard themselves from the negative effects of over-consumption the pill might detract policies to change people's environments. In this way the pill will indirectly contribute to the prolonging of perfidious socio-economic inequalities. Or they might argue that the kind of health generated by the ObesityPill is not really a valuable consequence at all, because in fact it is not health, but a form of disease dependent on permanent medication. According to this party, the pill leads to a further medicalisation of society.

As stated before, rights and principles do not figure prominently in this discourse. Of course, its adherents will have trouble denying that in modern societies people should be allowed their own autonomous choices. However, they can object that autonomy as yet hardly pertains to the victims of socio-economic-material injustices, who lack the necessary cultural competences to really make autonomous lifestyle choices. Or they might mobilize the principle of autonomy against the ObesityPill by arguing that people first of all have a right to an environment that 'enhances' their autonomy. In so far as the pill draws away resources from policies directed to creating such an environment, it conflicts with the principle of autonomy. Some hardliners might even go further and argue that autonomy is of course important, but the right to live a healthy life is more important. From this they might conclude that the pill has to be outlawed, at least until people live in healthy environments. Or they might argue for the moral duty to show solidarity with those worse off than yourself, a duty that then conflicts with the right to choose autonomously for using the ObesityPill.

Finally, like rights and principles, the good life plays a subordinate role in this discourse. But it will probably make its presence felt in the form of a critique of the attempt to solve, what adherents of this discourse perceive as, a social problem

by technical means. This attempt could then denounced as contributing to a 'false consciousness' about the 'true causes' of obesity.

We now turn to the plausible replies of the pill's proponents. They are faced with a two-faced enemy and therefore have to develop two lines of defence. They will probably attack the adversaries adhering to the individualizing discourse with their own weapons. This discourse puts a lot of weight on the principle of autonomy. Of course, the defenders of the ObesityPill see little cause to fight this venerable moral principle. However, they do interpret it quite differently The pill is not incompatible with autonomy, but people should be granted their autonomy, now interpreted as the right to make their own (consumer) choices. Even if one frowns upon the pill, it is bad taste to paternalistically pressure this private conviction on one's fellow-citizens, whose ability to make autonomous choices should both be assumed and respected. Furthermore, on the level of good life ethics, the pill promoters argue that the whole idea of mental self-control is out-dated by the new medical technology. Indeed, the Greek, Roman and Christian moralists stressed the necessity of mental self-control, but this was only for lack of a more effective way of self-control through medical technology. Now the pill has been made available, this type of morality should be considered as out-dated. Mental self-control should be respected as the best solution available in pre-technological times, but now it has become superfluous – at least where our body weight is concerned.

Now they turn to their adversaries from Discourse B. What matters, and this they share with their opponents, are results. Sterile moralizing should be avoided. Why, then, not allow experimenting with the obesity pill, with those wealthy enough to afford this device as voluntary guinea pigs? If the pill then proves to be an effective measure against obesity, the rationale for paternalistic policies is undermined. Paternalism can only be justified – if at all – by its results, part of which is educating individuals into autonomous citizens. If medical technology now provides us with results without paternalism, isn't that preferable? And if it works for the wealthy, isn't it a requirement of justice to make the ObesityPill available to all? Especially when we realize that a larger market will help to make the pill cheaper anyway?

9.2.4 Some Intermediary Reflections

We want to conclude this section with a few reflections on our 'imagined controversy'.

First: the discussion outlined above is of course speculative, but it is not idle speculation. The combination of previously existing moral discourses with the NEST-ethical patterns generates suggestions that both stimulate our moral imagination *and* ensure that this imagination remains grounded in reality. The ethical controversies we sketch have at least some degree of plausibility.

Second: it is interesting to notice that the two initial moral discourses on obesity come up with different arguments against the pill. The core values that constitute the core of each provoke these different reactions. Both perceive the pill as a threat,

but as a threat to different values, even if the debate is of course stylized and if the application of the NEST-ethical patterns might lead others to imagine other arguments that we have not mentioned.

Third: neither of the two 'threatened' discourses perceives the novelty of the ObesityPill to be a morally neutral means to an uncontroversial goal. The relation between means and goal embodied by the pill (healthy weight through medical technology) contrasts sharply with the means/goal relation characteristic of the individualizing discourse (healthy weight through a strong will) and with the means/goal relation characteristic of the policy discourse (healthy weight through a healthy environment). Goal and means are entwined. A major, although often overlooked element of moral change, is that a NEST can break up existing connections between goal and means.

9.3 The Closure of Techno-Ethical Controversies

The argumentative patterns of NEST-ethics can help us to imagine future controversies and how these might unfold. But the story has as yet no end. We have not yet attempted to imagine any closure of the (imagined) techno-ethical controversies. After the NEST disturbs moral routines, stakeholders will attempt to forge closure by creating a new fit between technology and morality, adapting the technology, the morals, of both. By speculating about possible closures, we find ourselves on very thin ice indeed, and in any real scenario exercise one would offer different scenarios, e.g. one where the technology is adapted to conform to current moral standards, one where the morals are adapted to the NEST, and one where the two strategies are combined. Because scenarios can be developed with different aims in mind, the *plausibility* of such an outcome need not be the first concern. It is perfectly legitimate, and common, to develop a scenario for highly implausible developments – like nuclear reactors exploding – when such developments are important (e.g. disastrous) enough. However, *all* scenarios have to be plausible to a certain degree, because they only work if the readers are seduced to momentarily suspend their disbelief. How to achieve such a minimum level of plausibility, when imagining closures to techno-ethical controversies? Here one can draw some guidance from the following two observations.

In the first place, some norms, values or principles can be plausibly expected to determine the outcome because they are more *robust* and thus resistant to (technologically induced) change than others. For example, it was not an unexpected coincidence that it was the no-harm principle that blocked experiments in human reproductive cloning. This principle is very robust and will weigh heavily in the coming years. How to determine which moral elements are robust? For an answer, we can draw on Dewey's pragmatist ethics.

According to Dewey, norms and values enter the world as solutions to practical *problems*. Because those problems are situated in time and space, the same holds for morality: it is deeply entwined with that world of everyday practice, because it

is meant to guide our action. If practical requirements change, so do morals. History is full of examples of norms, values and principles being moved, transported, reinterpreted and translated to help solve other and novel practical problems (Keulartz et al., 2002, 2004). Take for example the principle of autonomy. This principle was first coined to elucidate the precarious political status of the fifteenth century Italian city-states within the Holy Roman Empire; then played an important role in religious controversies; resurfaced in Rousseau's political philosophy; was elevated by Kant to take centre stage in his moral philosophy; and has in the last decades finally reached public prominence in the field of medical ethics[6] (Schneewind, 1998; Skinner, 1990). Other examples are the principles of non-maleficence and benevolence, or the Golden Rule. Such norms, values and principles have proven their worth over and over again, in many different contexts. Let us locate such abstract elements on a *macro* level, together with similarly fundamental processes like secularization, individualization and democratization (Trappenburg, 2003). These principles and processes may be trusted to remain relevant in tomorrow's ethical controversies. On the *meso* level we then locate the concretizations of these abstract principles and processes, adapted for specific practices. For example, the idea that the autonomy of patients requires their informed consent. On the *micro* level we then locate concrete ethical questions and answers like: should we ask ten year old patients for their informed consent or not?

We contend that the solutions on the macro level are the most robust and thus the least subject to change, the solutions on the micro level the least robust and the most subject to change, with the solutions on the meso level somewhere in the middle. Macro-elements are robust in the sense that they have proven to remain recognizable in many different practical contexts. The 'successes' of principles on the macro level don't make them immune to the pull of change, but they can be expected to change only slowly.[7] Similarly, although there is no guarantee that processes on this level will continue to unfold, it seems a reasonable bet that they will keep doing so in the foreseeable future. Decisions on the micro level, by contrast, depend heavily on contextual factors and are thus prone to change. On the meso level, rivalling theories argue for different interpretations of 'autonomy' (Schermer, 2001) and in different practices the same principle is often differently interpreted and enacted. These different interpretations keep the principle, as it were, constantly slightly destabilized. We can sum up our approach using the image of looking at a landscape from a

---

[6] One of the primary tasks for a pragmatist ethics is to study these transports, and evaluate them. In particular, such an ethics should pay attention to the problem that we cannot but meet new problems with old instruments, even if these instruments are poorly equipped for the new tasks ahead, never being designed therefore in the first place. The pragmatist ethicist excavates the original problem context in which a vocabulary came into existence, to show its particularity and situated character After that, the question can be put on the agenda and assessed, whether or to what extent this vocabulary can or should be fruitfully transported to other problem areas. (T. Swierstra, 2002)

[7] There are exceptions to this rule-of-thumb. For example: ideas about the perversity of homosexuality or the 'natural' hierarchy between men and women have at least such a venerable past, and they have considerable force in the past few decades – not the least thanks to the influence of the pill, an example of a NEST.

speeding train: the micro level in the foreground speeds by, the middle plan of the meso level changes considerably slower, whereas the macro level at the horizon moves almost imperceptibly[8] (Swierstra, 2004b).

These differences in robustness can guide us in our speculations on the plausible outcomes of ethical controversies. For outcomes that embody robust norms and values, are more plausible than outcomes that lack this type of moral backing. There is another way to differentiate plausible from less plausible outcomes. Although norms and values are almost by definition counterfactual, it is equally true that they must possess some degree of realism to gain practical relevance. '*Ought implies can*' as David Hume put it succinctly. In practice, whether we 'can' do something, is more often than not a matter of degree and of reasonableness. If it is very difficult to comply to a norm, it will dismissed by most as over-demanding, utopian and/or moralistic, only fit for the few moral heroes in our midst. Norms and values thus gain motivational force and influence to the extent that more people are convinced that these norms can indeed be put into practice without too much costs, and when the values are deemed to be realizable at least to some degree.

At exactly this point technology regularly interferes with morality. By opening up new practical avenues, technology can make some norms and values more realizable, and thus help them gain popular support. Thus, technology does not simply provide us with more practical opportunities, they at the same time kiss to life dormant obligations and responsibilities by supplying new 'cans' which result in new 'oughts'. Vice versa, technology often also closes off certain avenues, making actions more difficult or even impossible to perform. (Try doing without a cell phone for a while). This results in a weakening of values and norms that correspond with these marginalized actions. One could compare these mechanisms to that of a genome. Like genes, the 'expression' of norms and values is influenced by other norms and values, but also by external – in our case: technological – factors. Pursuing these technology/morality interactions, is the second way we can speculate in a controlled fashion on what are more and less plausible outcomes of future controversies.

9.4 The Obesity Pill, Part II: The Liberation of Fun

To summarize: speculations on techno-moral change can and should be guided by the identification of (a) robust moral elements that have proven themselves in a variety of times and locations, and (b) the influence of the NEST strengthening or weakening of the 'expression levels' of the various norms and values in play. Now let us return to our example of the ObesityPill controversy. What kind of closures are plausible? What kind of techno-moral change might plausibly result from the introduction of this biotechnological artifact? We will sketch one scenario,

[8] Note: of course there are no sharp, uncontroversial borders between the different levels.

even though alternative outcomes can be plausibly imagined, that highlights moral change instead of technological change or a combination of both.

The drug is initially marketed to the small group of patients who are pathologically obese, and for who no other therapies have proven successful. The principle of benevolence requires this, as does the no-harm principle: withholding available medication to patients is considered as harming them. So, a niche for the ObesityPill is quickly established, as soon as the pill has become available.

In a second stage, conflicts occur at the fringes of these patient-groups. No sharp medical boundaries can be drawn between those with a real disease and those with simply bad habits. This opens a space for constant negotiations and for shifting boundaries. The availability of this biotechnological artifact alone shifts the balance between diseased and weak-willed. Especially because moral condemnation of obese people is a social reality, it is rational for them to apply for the qualification 'diseased'. Some pressure their general practitioners into proscribing them this medication, even though their medical condition does not unequivocally qualify them for a prescription. In doing so, they mobilize an egalitarian conception of justice to get their way, arguing that their condition is essentially the same as that of the accepted in-group.

Furthermore, because the distribution of the pill is hard to control in practical terms, thanks to the Internet, the ObesityPill will quickly transform from a cure for the really diseased into a life style drug. But what is diseased anyway, especially now being weak willed is also partly determined by one's genetic make-up? Those taking the pill as enhancement argue that they don't harm anyone else, and thus they should be free to exercise their autonomy.

The availability of a cure for obesity thus profoundly challenges the established moral outlooks on obesity. The first victim is the moralizing discourse on obesity that stressed strong will and abstention. Because medication is now available, obesity further transforms from a sin into a disease. The result is a rapid a-moralisation of obesity, which further paves the way for allowing everybody to freely purchase the pill. Physical exercise is increasingly seen as a waste of valuable time, better spent in productive or fun areas. Many even go further and claim enthusiastically that for the first time in human history hedonism itself is liberated. At last having fun is separated from the punishment of disease. Nor is their any reason for guilt: the individual is not burdening the collective with the costs of her medication. Finally life style has really become a matter of free, autonomous choice, they cheer.

Opposing the (consequentialist) policy discourse on obesity, people start questioning the previous attempts to hold corporations accountable and to remove environmental causes of obesity. Thanks to this pill, they argue, the rationale behind the blunt, undiscriminating policy of environment-change ceases to exist. There is no longer need for state interventionism, now people can effectively control their body weight with the help of medication. The human organism, they say, is programmed to save energy. No one *likes* to climb a stairs when they can avoid it. According to them, the new medication offers us a way to obey our natural impulse to laziness, instead of being forced into doing unnatural exercises.

Due to the obesity pill, consequences change, rights and obligations are redirected, justice requires new things, and conceptions of the good life – i.e. pertaining to what is ill, what is weak, and what is fun – change. The new technology shapes its own, conducive, moral environment. That it is able to pull this off, is due to the fact that some very robust (macro)elements of morality – the principles of no harm, benevolence, justice, autonomy, the dividing line between natural and unnatural – are smoothly reinterpreted on the meso-level to endorse the new technology. On the other hand, the rivalling discourses and practices regarding obesity are weakened by the availability of a technical short-cut to a goal that everybody endorses: healthy bodies. Of course, many adherents of these discourses will stick to their positions, deploring what they perceive as moral and political decay. Of course, they will be branded as conservatives. Due to the availability of a technical alternative, the importance of self-control turns into a quaint form of masochism. Those adhering to their social solutions, are accused of a 'social fix'.

So, this particular scenario makes it easy to imagine how the ObesityPill might result in a moral change that can be christened *the liberation of fun*. Of course, a complete victory resulting in total closure is unlikely. Some adherents of the two rival discourses will stick to their basic contentions, and seek to further develop their moral positions, for instance by attaching more importance to having a 'natural' weight, or by stressing that weight problems are only symptoms of underlying socio-economic injustices. The controversy will continue, in all likelihood. Be that as it may: the scenario does provide us with a glimpse of techno-moral change. To what use?

9.5 How Techno-Moral Scenarios Can Enhance Our Moral Judgement

If we want to guide technological development in beneficial directions, we have to take into account that moral change will in all likelihood accompany it. Scenarios are a good way to explore such techno-moral change. If such a scenario convinces us that certain moral consequences might very well occur, we seem condemned to a choice between two options.

From what standpoint can one judge future morals? Here we have to avoid two extremes: relativism and transcendentalism. Moralities change when their environments change, although admittedly some moral elements are fairly robust. But this very general acceptance of moral change does not amount to moral relativism. Such relativism typically comes in two forms, either favouring the present or the future. Moral *presentism* simply favours current morals over the future ones we imagined as part of the scenario exercise. This precludes the possibility that our future selves, or our children, might have learned something worth knowing and applying in the present.

Neither should we passively *accept* the moral changes we imagined. There is no need to follow Baudelaire who hailed the 'extraordinary delight of celebrating

the advent of the *new!*' (Berman, 1983) (quoted on p.143) or the Italian futurists: 'Comrades, we tell you now that the triumphant progress of science makes changes in humanity inevitable, changes that are hacking an abyss between those docile slaves of tradition and us free moderns who are confident in the radiant splendor of our future.' (Berman, 1983) (quoted on p.24–5) This moral *futurism* rests on the opposite mistake of precluding the possibility that we presently possess a sharper insight in rights and wrongs than our future selves. The moral vision of our future selves will be bound with countless ties to the then existing practices and technological artefacts. Once a technological opportunity exists, it is hard to pause and reflect on – for example – the desirability of the novel rights and duties that this opportunity calls into existence. The reason is that when technology and morality co-evolve, they both start out fluid and flexible – hot technology as it were, and hot ethics – but in the course of their progressive realization they both solidify and become reified. They tend to become self-evident. Furthermore, the new technology will have created new interests and new (claim)rights which make the newly evolved and solidified techno-moral constellation difficult to change in practical terms. So, if we manage to imagine this new techno-moral constellation before it has become socially embedded, our reflection is less restricted by those facts and practicalities. Because in the present we still possess alternatives that in the future will be closed off, our current ethical reflection is more open and free compared to the cold morality of the future. Here exists an important parallel with Rawls' construction of the original position: it is easier to deliberate in an open and rational way about a (future) just society when one has still a degree of distance to it and does not identify yet with its biases. This is why our present selves might have a sharper vision than our future selves, and thus have something to teach.

One might object that our current moral deliberation is as determined by current technologies and practicalities as the moral deliberation in the future will be. True. Our present moral imagination is situated in ways that can only be clearly perceived with the benefit of hindsight. Ultimately, we are unable to jump our moral shadows. *But it is still worthwhile to try*. By using our moral imagination, by developing narratives about the future co-evolution of technology and morality, we broaden our mind. We seek out new experiences and travel to different – even if only imagined – moral cultures. For example: when our imagination evokes the possibility of self-mastery through a medical device like the ObesityPill, we are invited to reflect upon the pro's and con's of traditional and technical self-mastery. If the ObesityPill is likely to undermine policies that are (rather indiscriminately) directed to groups instead of individuals, we can now ponder the costs and benefits of both approaches. Discussions like these will enormously enhance our ethical technology assessment, because they deal with the ObesityPill and all its ramifications, instead of with a pill that is stripped of its moral context.

That we reject relativism, does not necessitate us to embrace the idea of transcendent, eternal and universal, moral standards. Like relativism, transcendentalism is in danger of deflating our openness, curiosity, reflexivity, creativity and willingness to learn. These are our core values, because they reflect the dynamism that defines our technological culture. We do not need an Archimedean point to decide on 'the'

best morality. In our search for (moral) truth it suffices, as Hans Georg Gadamer pointed out decades ago in his book on philosophical hermeneutics, to seek out conflicting perspectives that invite us to question our prejudices (Bernstein, 1983; Gadamer, 1986). Moral learning can occur where and when people are confronted with 'strange', new, conflicting morals. Even when, as in art, we have to devise these conflicting perspectives in our imagination. Moral plurality invites reflection, (self)criticism, dialogue and the open exchange of ideas.

By developing techno-moral scenarios, we travel to future worlds where different technologies and morals prevail. It is by seeking this confrontation between present and imagined morality, that we learn to guide technological change in a manner both reflective and flexible, without reifying either the present or the possible future.

Acknowledgment We gratefully acknowledge the support of the Netherlands Organisation for Scientific Research.

References

Berman, M. (1983). *All That Is Solid Melts Into Air. The Experience of Modernity* (3 ed.). London, New York: Verso.

Bernstein, R. (1983). *Beyond Objectivism and Relativism: Science, Hermeneutic and Praxis*. Oxford: Basil Blackwell.

Bijker, W. (2001). The Need for Public Intellectuals. A Space for STS. *Science, Technology & Human Values, 28*(4), 443–450.

Coeckelberg, M. (2007). *Imagination and Principles. An Essay on the Role of Imagination in Moral Reasoning*. Basingstoke, New York: Palgrave MacMillan.

Dewey, J. (1922). *Human Nature and Conduct*. New York: H.Holt & Co.

Dewey, J. (1954). *The Later Works*. Athens OH: Swallow Press Books.

Dohmen, J. (Ed.). (2002). *Over levenskunst. De grote filosofen over het goede leven*. Amsterdam: Ambo.

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

Ferretti, M. P. (2007). Why Public Participation in Risk Regulation? The Case of Authorizing GMO Products in the European Union. *Science as Culture, 16*(4), 377–395.

Fesmire, S. (2003). *John Dewey and Moral Imagination. Pragmatism in Ethics*. Bloomington & Indianapolis: Indiana University Press.

Gadamer, H.-G. (1986). *Wharheit und Methode: Grundzüge einer philosophischen Hermeneutik* (Vol. Band 1). Tübingen: J.C.B.Mohr (Paul Siebeck).

Ihde, D. (1998). *Expanding Hermeneutics*. Evanston, IL: Northwestern University Press.

Johnson, M. (1993). *Moral Imagination*. Chicago: University of Chicago Press.

Kass, L. R. (1997). The Wisdom of Repugnance. *The New Republic, 216*(22), 17–26.

Kass, L. R. (2002). *Life, Liberty, and the Defense of Dignity*. Encounter Books.

Keulartz, J., Korthals, M., Schermer, M., and Swierstra, T. (eds.). (2002). *Pragmatist Ethics for a Technological Culture*. Deventer: Kluwer Academic Publishers.

Keulartz, J., Schermer, M., Korthals, M., and Swierstra, T. (2004). Ethics in a technological culture. A programmatic proposal for a pragmatist approach. *Science, Technology and Human Values, 29*(1), 3–29.

Larmore, C. (1987). *Patterns of Moral Complexity* (2 ed.). Cambridge, New York: Cambridge University Press.

Logister, L. (2004). *Creatieve Democratie. John Deweys pragmatisme als grondslag voor een democratische samenleving*. Budel: Damon.

Mackenzie, D., and Wajcman, J. (Eds.). (1999). *The Social Shaping of Technology* (2 ed.). Buckingham: Open University Press.

McGee, G. (2002). *The Perfect Baby: A Pragmatic Approach to Genetics*. New York: Rowman & Littlefield.

Nietzsche, F. (1955). *Werke in drei Bänden,* (Vol. III). München: Carl Hanser Verlag.

Notten, P. W. F., Rotmans, J., and Asselt, M. B. A. V. (2003). An updated scenario typology. *Futures, 35*, 423–443.

Rip, A., and Kemp, R. (1998). Technological change. In S. Rayner & L. Malone (Eds.), *Human choice and climate change. Volume two: Resources and technology* (pp. 327–399). Columbus, Ohio: Battelle Press.

Sandel, M. J. (2004). The Case Against Perfection. *The Atlantic Monthly* (April).

Schermer, M. (2001). *The different faces of autonomy. A study on patient autonomy in ethical theory and hospital practice.* Universiteit van Amsterdam, Amsterdam.

Schneewind, J. B. (1998). *The invention of Autonomy. A History of Modern Moral Philosophy* (1 ed.). Cambridge, New York: Cambridge University Press.

Sclove, R. E. (1995). *Democracy and Technology*. New York, London: The Guilford Press.

Skinner, Q. (1990). Political Philosophy. In C. Schmitt and Q. Skinner (Eds.), *The Cambridge History of Renaissance Philosophy* (2 ed., pp. 389–453). Cambridge, New York: Cambridge University Press.

Swierstra, T. (2002). Moral vocabularies and public debate. The cases of cloning and new reproductive technologies. In J. Keulartz, M. Korthals, M. Schermer and T. Swierstra (Eds.), *Pragmatist ethics for a technological culture.* Deventer: Kluwer Academic Publishers.

Swierstra, T. (2004a). Een tumor is ook collectief bezit. Het afstaan van lichaamsmateriaal ten behoeve van DNA-banken. *Krisis. Tijdschrift voor empirische filosofie, 5*(4), 36–54.

Swierstra, T. (2004b). *Slachtoffer of burger? Een essay over het nader gebruik van lichaamsmateriaal ten behoeve van genomics onderzoek*. Amsterdam: NVBE.

Swierstra, T., and Rip, A. (2007). Nano-ethics as NEST-ethics. Patterns of Moral Argumentation About New and Emerging Science and Technology. *NanoEthics, 1*(1), 3–20.

Trappenburg, M. (2003). Darwin in de medische ethiek. In M. Adams, J. Griffiths and G. d. Hartogh (Eds.), *Euthanasie. Nieuwe knelpunten in een voortgezette discussie* (pp. 237–254). Kampen: Kok.

Williams, B. (1985). *Ethics and the Limits of Philosophy*. London: Fontana Press/Collins.

Part III
Evaluating New Technologies: Uncertainty and Precaution

Chapter 10
On Uncertainty in Ethics and Technology

Paul Sollie

Abstract The article aims to examine uncertainty in technology development and its subsequent ramifications for ethical technology assessment. Although uncertainty is a pivotal feature of complex technologies, its importance has not yet been fully appreciated within the field of ethics. Hence, the purpose of this paper is to study uncertainty in technology development and its consequences for ethics. Going on the insight of other scientific disciplines such as environmental studies or economics the concept of uncertainty is disentangled and a typology of uncertainty is proposed and introduced to ethical theory. The uncertainty typology results in a series of questions with regard to the collection of information about the object of assessment (i.e. complex technologies and their development) and the framework of assessment (i.e. ethical theory and its practical aim of guiding the assessment of technology development). What is more, the uncertainty surrounding technology development has ramifications for ethical technology assessment. Any attempt to provide an account of ethics of technology may seem daunting given the fact of uncertainty, because uncertainty results in a lack of information to guide our moral decision-making. In order to deal legitimately with uncertainty, I claim that any adequate ethics of technology needs to account for *both* substance and procedure. The paper concludes with requirements for any future ethics of technology under uncertainty.

Keywords Technology development · Technological innovation · Uncertainty · Typology of uncertainty · Uncertainty management · Ethical technology assessment

10.1 Introduction

Going on the insight of various scientific disciplines, such as economics and environmental studies, I will argue that uncertainty is a pivotal concept in many contemporary debates, notably that of technology development and its ethical evaluation.

P. Sollie (✉)
Ethics Institute, Utrecht University, Utrecht, The Netherlands
e-mail: paulsollie@gmail.com

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_10,

However, within the field of moral philosophy the concept of uncertainty is a notion not fully explored and scrutinised. This is noteworthy as we observe that technology has become increasingly more complex, pervades nearly every sphere of life, and involves far-reaching and often unforeseen and unanticipated consequences for human beings and the environment. Consider, for instance, environmental concerns like non-degradable plastics or carbon dioxide emissions; IT-related issues such as the influence of the Internet on communication and social relations; issues with regard to surveillance-enabling technologies like CCTV and RFID. Another example is the technological programme of 'Towards Ultra-fast Communication' (hereafter TUC), which aims at improving and increasing the speed of data transmission. (See Chapter 2) The speed of data transmission, which has increased over the past years, is however determined and constrained by aspects like the speed of the client, the capacity and the use of the data line, the routing, and the speed of the server. With current technologies sending packets of information over the Internet is rather unproblematic. However, two trends might compromise Internet traffic. First, there is an increase in data transmission over the Internet as new possibilities open up, such as streaming media and 3D movies. Second, the number of people engaging on the Internet is still significantly growing. For example, Africa, the Middle East, and Asia, which are in terms of percentage regions with the lowest penetration of Internet usage but have enormous populations, display a usage growth between 2000 and 2007 of 643, 495, and 282% respectively. (Internet World Stats, 2007). These trends will ultimately impact on the capacity of the Internet. While sending packets of information over the Internet at very high speed is unproblematic with current technologies, restraining factors are the nodes at which information is processed. At this moment these nodes are incapable of processing information optically. These nodes route information electronically, which is significantly slower than photonics. In order to overcome future problems of congestion engineers have engaged in this technological program of developing optical switches to provide for ultra-fast communication in the future without congestion. TUC aims at creating faster and more effective use of information and communication networks. This endeavour does not seem risky in itself or to raise any moral dilemmas at the outset. Nevertheless, what about future applications or consequences of such a complex technology? The issues mentioned, which surround complex technology development, have led to many a debate on the desirability and moral acceptability of new, complex, and possibly disadvantageous or even catastrophic technologies. Since the issue of uncertainty is underappreciated in such debates, the question of how we should deal ethically with new, uncertain technology developments is the focal point of this scrutiny.

The aim of this paper is to study uncertainty in technology development and its subsequent ramifications for ethical technology assessment. Uncertainty is to be distinguished from risk. Risk, which I leave outside this paper, refers to situations in which probabilities can be assigned to known possible future states of the world. Conversely, uncertainty has to do with situations in which probabilities cannot be attributed to future states, which are often indeterminate themselves. In this paper I will argue for the necessity of ethics to reflect upon uncertainty. This will involve a theoretical analysis, conceptualising uncertainty via a typology formulated in other

scientific disciplines. I will then show that this scrutiny is insightful for the debate on ethics of technology. Uncertainty is paramount, particularly in the field of technology development and its ethical assessment. Although the typology also applies to existing complex technologies, the paper focuses on the specific endeavour of technology development. The uncertainty surrounding technology development has ramifications for the ethical assessment. Any attempt to provide an account of ethics of technology might seem daunting given the fact of uncertainty, because uncertainty results in a lack of information to guide our moral decision-making. In order to deal legitimately with uncertainty, I claim that any adequate ethics of technology needs to account for *both* substance and procedure.

In Section 10.2 I attend to the underlying question of how to deal ethically with complex, uncertain technology developments. This brings to the front fundamental questions about ethics, ethical theory, and its aims. Section 10.3 is concerned with uncertainty in the field of technology development. Uncertainty within technology development relates to the characteristics of complex technology development, which include among other things multi-agency, opaque R&D trajectories, and substantial indeterminacy regarding use and impact. The complexity and uncertainty of developments stand out as special compared to ideal-typical cases in ethics that are straightforward, linear-causal, unequivocal, and calculable. As a systematic reflection on uncertainty is lacking in ethics of technology, I discuss and introduce in Section 10.4 a typology of uncertainty that differentiates between its nature, levels, and locations. In Section 10.5 I argue that by introducing this typology into the field of technology development we arrive at a clearer picture of uncertainty in technology development and its ethical assessment. The different dimensions of the typology prove to be insightful for working towards a methodology for morally evaluating complex technology developments. In the concluding Section 10.6 these insights will facilitate the identification of problems and requirements for any adequate ethics of technology.

10.2 The Aims and Adequacy of Ethics

The rationale behind this scrutiny is the difficult issue of how to deal with complex, new technologies that are surrounded by uncertainty. This difficulty underlies the practical and wide-ranging question ‘How are we able to deal proactively with complex technology developments that are characterised by uncertainty from an ethical perspective?’ that I take to be one into the nature and scope of ethical theory. This question can be decomposed into several subquestions of which I will address two. First, the question, which I only briefly touch upon, is a research into ethical theory as a philosophical discipline that aims at discovering, justifying, and applying right-making criteria. What are the nature and scope of these right-making criteria within ethical theory? Are these criteria coherently presented and how are they accounted for and justified? How are the principles related to the object of which they purport to say something about? The object of evaluation pertains to the second

subquestion and entails research into the uncertainty of new and emerging technologies. In this paper I concentrate on the second question, but I will commence with briefly reflecting upon the first question.

The central question is one that goes straight to the heart of the nature of ethical theory and raises questions on different levels as it regards the *question of what agents which interests should be taken and how this can be justified*. With Timmons I distinguish two complementary and necessary aims of ethical theorising that make up for the *adequacy* of ethical theory – the theoretical and the practical aim. Concerning the first aim, I recognise, in accordance with Gewirth, that the 'most important and difficult problem of philosophical ethics is whether a substantive moral principle can be rationally justified.' (1978, ix) The *theoretical aim* of the justification of right-making criteria generates three central questions: the distributive, the substantive, and the authoritative question. (Gewirth, 1978, 3) The distributive question relates to the question of whose interests other than and in addition to their own agents should consider. Any ethical theory should have a justified theory with regard to what agents they include in the realm of relevant agents of which account should be taken. Questions in this area concern issues such as whether 'I' am the only agent whose interests count as egoists will argue, whether foetuses have moral status, or whether we should include animals in the domain of relevant agents. The substantive question concerns the issue of which interests of relevant agents account should be taken. Do all preferences of relevant agents count, only particular preferences, or perhaps even none? The authoritative question of moral philosophy aims at tackling the question why one should be moral, in which the 'should' refers to a criterion other than the criterion of moral rightness whose obligatoriness is in question. (Gewirth, 1978, 3) What is more, with Parfit I agree that we need reason-giving facts for following certain rules and requirements set forth by an ethical theory. (Parfit, 2006, 72) Where there are no reasons given *why* we are obligated and required to follow certain moral rules, we will be reluctant to give them priority over other considerations, even non-moral considerations. 'To be able to make significant claims about the relative importance of moral requirements and requirements of [. . .] other kinds, we need some non-moral, neutral criterion.' (Parfit, 2006, 73)

The second level pertains to what Timmons calls the *practical aim* of ethics, namely applying right-making criteria to judge and evaluate persons, actions, state of affairs, and the like (Timmons, 2002, 3). It aims at demonstrating *how* these right-making criteria are practicable in moral issues, which is different from us being able *to* apply these criteria, for instance, under conditions of uncertainty. The latter entails, as I argued elsewhere, that the adequacy of ethical theory is independent from a contingent aspect or circumstances such as the uncertainty of technology developments (See, further, Sollie, 2008). For an adequate theory there is no problem *how* to apply right-making criteria, but there may be a *problem for us to apply* these criteria, which is ultimately a problem of our epistemic indeterminacy, our limited knowledge and capabilities. So, it is one thing — although certainly not a minor and easy thing — to satisfy the theoretical objective of ethical theory by forwarding and justifying right-making criteria, yet it is quite another thing to meet the practical

aim of ethics by guiding correct moral reasoning. According to Gewirth, 'an ethical theory should not only set forth a justified principle that grounds moral rights and duties; it should also show how the principle serves to resolve moral conflicts, including conflicts between the duties it grounds.' (Gewirth, 1984, 249) Both the requirements of justification and practicability need to be satisfied for an ethical theory to be considered as adequate.

The application of right-making criteria, often formulated as principles such as the Kantian Categorical imperative or the Consequentialist principle of utility, is not without problems as I will demonstrate more extensively below. For the present purpose, it suffices to state two main issues in view of the practicability requirement. First, one of the aspects that I will highlight at this point concerns the disagreement between proponents of different ethical theories. They may disagree about the legitimacy of the right-making criteria forwarded in a theory. How should we deal with the fact that agents may hold diverging beliefs about morality when trying to resolve moral issues in practice? This issue bears directly on the theoretical objective of ethical theory, namely the underlying justification of right-making criteria. In this sense the practical aim is dependent on the theoretical aim. Second, even if all agents endorse one particular ethical theory, this theory still may leave room for conflicting requirements or duties within that specific moral framework. For this reason, any adequate ethical theory must specify *how* it will adjudicate such theory-intrinsic conflicts.

To sum up, any adequate ethical theory should be able to deliver the theoretical aim of ethics by answering the authoritative, distributive, and substantive question as well as the practical aim of guiding correct moral reasoning by demonstrating *how* right-making criteria apply. The *adequacy* of ethical theories is measured by the extent in which both the theoretical and practical objectives are satisfied. *Both* requirements of justification and practicability need to be satisfied for an ethical theory to be considered as adequate.

10.3 Complex Technology Development

In this section I take up the issue of uncertainty that surrounds many new and complex technology developments. Uncertainty is a concept less scrutinised in ethics than risk, of which exists an extensive literature. Despite the apparent reluctance to address this question, uncertainty proves to be a matter central to many a debate on ethics and technology.

Complex technology developments are characterised by uncertainty. Uncertainty seems to be reducible to empirical observations such as the changing nature of technology development over the course of history or the complex development trajectories with unpredictable outcomes. Without presenting a historical perspective on technology development, it is not unreasonable to argue that contemporary technology development is fundamentally of a more complex nature than in the past. One of the main reasons is that in the past technologies were to a great extent within

human comprehension. Technology was related to core human actions and its main purpose was to relieve or better coordinate certain human activities. Technologies merely comprised of the same activities as humans were concerned with, but technologies made these easier, lighter, faster, and so on. For example, the wheel enabled people to travel or transport things faster and more easily; a hoisting apparatus to move heavy objects with less force; or, windmills to manufacture certain products more efficiently and faster. Without attempting to trace a specific point in history, over time technology development became more complex. TUC, for instance, is a difficult to comprehend technological program as it starts off from fundamental research on materials to discover the material or constellation of materials that best suit the goal of switching light optically. New scientific discoveries, such as Newtonian mechanics, molecular physics, and quantum mechanics, definitely contributed to this change and opened up the possibilities for new sorts of technologies, which are more complex of nature and exceeding human measure and intellect.

The most important factors of uncertainty that surround technology development pertain to the unpredictable, unforeseen, and unanticipated nature of complex technology development trajectories, that is from research and development to the subsequent user application and consequences of the artefacts (See Healy 2006; Tenner, 1996; Brown et al., 2000; Cilliers, 2005; Verbeek and Slob, 2006; Von Schomberg, 2007). Telling examples are the Internet (developed for decentralised and secluded military and academic information transfer and now utilised as a multipurpose worldwide network), the telephone (originally designed for business purposes — Bell could not imagine people having and using such a noisy apparatus at home — and now every single person has a cell phone and can be contacted anywhere and anytime), or SMS (developed for telemetric purposes, i.e. the measuring and reporting of user information to providers and/or designers, while it is now being used by people to text messages to each other). The uncertainty arising from unpredictable, unforeseen, and unanticipated nature of technology development has many reasons (Sollie, 2005). To begin with, the examples indicate that, while technology is designed for specific purposes, it often ends up being used for completely different activities. This also relates to what Moor (1985, 269) has coined the logical malleability of technology; technology can be shaped to do any activity that can be characterised in terms of logical operations. Moreover, the development trajectories are often opaque (see, for instance, Brey, 2001, 52) for a variety of reasons. To begin with, there is the fact of multi-agency. The research and development trajectories consist of many agents from different disciplines who are all involved in various parts of the process. Each individual phase of development might be so recondite that other agents cannot fully comprehend what has been done in previous phases or what will be done in next phases. Not infrequently, agents take over products or artefacts from previous phases as a black box, i.e. a closed system of a technological artefact, which they then continue to work on. Next, uncertainty also arises from the unprecedented pace of technological developments (See, for instance, Johnson, 1997, or Osborn, 2002, 37). There is a continuous drive from science, the market, politics, and so on, to develop new and improve on existing technologies; for instance to produce faster technologies, to increase efficiency, or to innovate and introduce new technologies. This technological race obstructs our

trying to react adequately to and to get grip on these developments. Moreover, uncertainty is triggered by the scale on which technology is being developed; global and international developments interact with local and national developments. These are complex processes in which influence is exerted on different levels, in which local and global treaties and laws play important roles; not to mention cultural and moral pluralism.

These developments and observations are also what David Collingridge infers in his Control Dilemma: 'attempting to control a technology is difficult, and not rarely impossible, because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow.' (1980, 19) In the early phase, during the design and development of a specific technology, it is still possible to control and steer this development. Controlling and steering necessarily presupposes relevant information on for instance application, consequences, and impact. However, the information on basis of which one can perform a (moral) evaluation and subsequently steer the technology is lacking in early phases of development. Only in later phases of technology development and during its embedding and stabilising in society it is possible to collect relevant information on the application and consequences, but at that point it is extremely difficult and expensive to adjust an already existing and embedded technology.

A contemporary example of uncertainty of technology development concerns nanotechnology, which is an umbrella term for the fabrication of devices on atomic and molecular scale. Nanotechnology is an interdisciplinary field of disciplines like physics, engineering, chemistry, biology and computer science. It is a trumpeted area of technology development that is accompanied by huge promises such as powerful quantum computers and long term life preservation. It is nevertheless also fraught with uncertainty. As nanotechnology is in its early phase, 'it is still quite uncertain which options will be developed and which applications can and will be materialized. Nanotechnology is mostly promise, and sometimes pure speculation.' (Rip, 2006, 270) '[T]he future science and technology is still uncertain, let alone the future world in which these functions may have effects.' (Rip, 2006, 275) It is far from certain what will be possible in nanotechnology, which applications will be developed, how they will be used, and how they will impact on human beings and the environment. Nanotechnology is a particularly illustrative example of an uncertain technology development because of its complexity, its current merely speculative state, our inability to predict future applications and consequences, and the connected fears of catastrophic scenarios that it is accompanied by. (See for such possible applications and catastrophic scenarios, e.g., Weckert and Moor, 2006 and Clarke, 2005).

To conclude, many new technologies are surrounded by uncertainty due to opacity, complexity, multi-agency, long development trajectories, orientation at the future, global character, impact, and the fact that technology often shifts in application. These reasons contribute to the fact that uncertainty is an essential concept in modern, complex technology development and, consequently, it needs serious attention from those involved in ethics of technology.

10.4 A Typology of Uncertainty

Although uncertainty is central to complex technology, the question of what uncertainty entails is still unexplored. Over the course of the past decades, the concept of uncertainty has gained a lot of attention in many different scientific disciplines, like economics and environmental, management, and innovation studies. Numerous articles and books have been published trying to describe and conceptualise the concept of uncertainty. For example, one of the first major contributions to the study of uncertainty was by Frank Knight trying to incorporate risk and uncertainty into economical theory in his seminal 1921 *Risk, Uncertainty, and Profit*. In moral philosophy or, more specifically, ethics of technology, the importance and relevance of uncertainty for its discipline has not yet been fully recognized. Uncertainty has a rather new and un(der)developed position within ethical theorising. Ethical theories often put forward ideal-typical cases to show how these right-making criteria can be applied in the practice. The ideal-typical cases are generally linear-causal consisting of well-defined borders, a small and fixed number of stakeholders, and so on. For example, is it morally justified to steal bread to survive when poor? Should I kill one Indian to rescue nineteen others? Is it morally permitted to treat a mother with uterine cancer if this results in the loss of her embryo – when not treating her would result in the death of both? Is it allowed to steal medicine from an apothecary to save my sick mother? What these classical cases have in common is that they are rather surveyable. The cases consist of rational, acting, and evaluating agents that are capable of overlooking, calculating, and evaluating from a normative framework a limited subset of future states of the world of which one has substantial information regarding the range of possible actions and their consequences.

Although ideal-typical cases are well suited to show the adequacy of an ethical theory for guiding our moral reasoning in practice, this is only so theoretically. Complex, real life cases are of a different, more intricate nature and are misrepresented if treated as ideal-typical and linear-causal. As a result, ethical theories should be more robust and be tested against more complex issues. More specifically, coinciding the advent of complex modern technologies, uncertainty entered the domain of ethics when people found themselves called upon to morally evaluate these developments. If ethics aims at evaluating technology (not only retrospectively, but also proactively), then it has to account for the inherent uncertainty of technology development. This concerns an intrinsic problematic aspect of ethical theory, namely the application of theory to real life cases, which are always contextual and exceed the complexity and linearity of scholarly cases. I argue that, whereas many traditional theories have presented right-making criteria (e.g. principles) that were well suited for the assessment of relative well-calculable, unequivocal, and transparent objects of evaluation, complex technologies thwart both the practical and theoretical aims to a greater extent, because of the features surrounding complex technology development. Complex technology developments do not suffice the descriptions of traditional, ideal-typical ethical cases due to the uncertainty of future consequences, impacts, and applications. The following complex questions should be dealt with. Who are the agents that are involved in a particular technology development? What

are the consequences of particular decisions during the R&D trajectories? What are possible applications and consequences of new technologies? Who is affected and to what extent? What status do stakeholder values and opinions have and how are these integrated in ethical analysis? For example, although engineers of TUC have specific goals in mind, various decision during R&D, such as about the materials, model structure, or algorithms, influence implicitly or explicitly the final product and thereby open up or close down potential applications. Moreover, it is not yet determined how results of TUC will materialise, how it will be implemented, how it will be applied by end-users, or what consequences it will have for society. Unlike traditional ethical cases, the variables for the ethical evaluation of new, complex technologies are vague and unclear, which poses serious problems for ethics when trying to evaluate these technologies. Since uncertainty is un(der)developed in ethical theory, it is worthwhile to study how uncertainty has been conceptualised in disciplines other than ethics and how ethics might benefit from this.

10.4.1 Three Dimension of Uncertainty

Before turning to a conceptual analysis of uncertainty it is important to recognise that uncertainty is a fact of life that we have to deal with when making decisions in all sorts of domains. The fact that certainty is a utopian perspective might yield to inertia, but there is no need to throw in the towel too quickly. In case of decision-making, therefore, we would better try to understand what uncertainty is, because when we have an enhanced understanding of its dimensions and implications for decision-making, we are opening up the way for constructively dealing with a complicating fact of life. If we acknowledge this, then the first step is to conceptualise uncertainty. With Walker et al. I take uncertainty as 'any departure from the unachievable ideal of complete determinism.' (Walker et al., 2003, 8) For various reasons, uncertainty is not simply the absence of knowledge. First, uncertainty can prevail even in situations where a lot of information is available. Second, new information does not necessarily increase certainty, but might also augment uncertainty by revealing the presence of uncertainties that were previously unknown or understated (See, e.g., Beck, 1999, 6). Third, there might even be situations of uncertainty that are indeterminable and which for the reason of the nature of that situation cannot be reduced by acquiring knowledge. Such indeterminable situations are related to, for instance, the behaviour of other agents. These aspects show that uncertainty is more a multifaceted concept than it might seem at first glance. For that reason, I will advance by elucidating its different dimensions by reconstructing a typology of uncertainty. This typology of uncertainty (Walker et al., 2003) involves a three-part distinction between the nature of uncertainty, the levels of uncertainty, and the sources or locations of uncertainty.

With regard to the first dimension of uncertainty, the *nature* of uncertainty, Walker et al. distinguish between two types of uncertainty, namely knowledge or epistemological uncertainty and variability or ontological uncertainty (2003, 13–14).

This distinction contributes to identifying and assessing the nature of uncertainty of the phenomena that are studied, in casu technology development. Epistemological uncertainty is uncertainty that is related to the properties of agents and the collection of knowledge. Epistemological uncertainty results from the lack of information or the complexity of the situation that needs to be assessed. In principle, this type of uncertainty can be reduced by acquiring knowledge, more measurements, and better models. Other terms used for epistemological uncertainty include weak, internal, secondary, or substantive uncertainty (See Meijer et al., 2005, 8). Ontological or variability uncertainty pertains to the uncertainty that is inherent to the variability of the system and which cannot be reduced. Many empirical variables that are the input for the calculus or assessment are instable. They change and vary over space and time and are beyond human control, simply due to the nature of the phenomena involved (Walker et al., 2003, 13). Ontological uncertainties are attributes of reality. Decisions within a framework or system generate a large number of potential and possible outcomes, decreasing the confidence in predictions. Strong, fundamental, stochastic, random, primary, external, aleatory or procedural uncertainty are synonyms for ontological uncertainty (See Meijer et al., 2005, 8). According to Walker et al., one can identify four aspects that might contribute to ontological uncertainty (2003, 13–14). First, the inherent randomness of natural processes. Second, the unpredictability and variability of human behaviour, such as irrational behaviour or discrepancies between what people say, feel, think, and what they actually do. Third, the unpredictable nature of societal processes stem from social, economic, and cultural dynamics. Fourth, technology development not rarely entails technological surprise, such as breakthroughs in technology, unexpected consequences, and side effects (See also Tenner, 1996). Meijer et al. contend (2005, 8) that the distinction between epistemological and ontological uncertainty is not a sharp one. Ontological uncertainty contributes to knowledge uncertainty, because, due to its variability, perfect knowledge and certain predictions are anyhow unattainable. But epistemological uncertainty can also exist in deterministic processes, e.g. due to a lack of communication, inexact measurements or too high a complexity.

The second dimension of uncertainty involves the *levels* of uncertainty, which according to Walker et al. (2003, 11) pertain to the question of *how* uncertain something is. The uncertainty of something can be classified in different stages ranging from deterministic understanding, also called the 'know,' to complete ignorance, the 'no-know'. They identify four levels of knowledge on the spectrum, from determinism to indeterminism. (Walker et al., 2003, 11–13) First, statistical uncertainties entail uncertainties that can adequately be expressed in terms of probabilities. Second, uncertainties that cannot be depicted adequately in terms of probabilities, but which can only be specified in terms of possible outcomes are called scenario uncertainty. The methods for assessing the probabilities are, however, not correctly understood, which makes it more indeterminate than statistical uncertainties. Third, recognised ignorance involve uncertainties with regard to the relationships and mechanisms of which we realise in some way or another that they are present, but of which we cannot establish any useful estimate; for instance, due to limits of predictability, knowledgeability, or due to unknown procedures. The fourth level connects to total

or complete ignorance. These uncertainties are most indeterminate; the so-called 'unknown unknowns'.

The third dimension comprises the *sources or locations* of uncertainty that are an identification of where uncertainty manifests itself within the model or framework. This dimension, Walker et al. argue (2003, 9), refers to the model in which it is possible to pinpoint the various sources of uncertainty for a specific domain. The source or locations might yield to uncertainty in two ways; uncertainty in systems, frameworks, or theories and uncertainties arising in particular domains or spheres of human life and action. On the one hand, the locus of uncertainty is based on the uncertainty in systems or theories. Subsequently, the location of uncertainty is an identification of where uncertainty manifests itself in the complex of decision-making. Walker et al. distinguish the following locations of uncertainty with respect to the model (2003, 9). First, the context, which has to do with the framing of the problem, including the choices determining what is considered inside and outside the system boundaries, as well as the completeness of this representation in view of the problem at hand. This involves the identification or setting of the boundaries of the system to be modelled and, accordingly, the aspects of the real world that are inside the system, the aspects that are outside, and the completeness of the system. Second, Walker et al. identify model uncertainty, which is associated with both the conceptual model (i.e. the variables and their relationships that are chosen to describe the system located within the boundaries and thus constituting the model complex) and the computer model, in casu the practical framework of ethics. Model uncertainty can be divided in model structure uncertainty (uncertainty about the form of the model or framework itself) and model technical uncertainty (uncertainty arising from the practical implementation of the model). Third, inputs that are associated with the description of the reference system and the external forces that are driving changes in the reference system. It is sometimes useful to divide the inputs into controllable and uncontrollable inputs, depending on whether the decision-maker has the capability to influence the values of the specific input variables. Fourth, parameter uncertainty relates to the data and the methods used to calibrate the model parameters. Fifth, model outcome uncertainty is the accumulated uncertainty associated with the model outcomes of interest to the decision-maker.

On the other hand, the sources or locations of uncertainty relate to the specific domain about which the decision-maker is uncertain. The source of uncertainty depends on the context. Different sources of uncertainty are distinguished by Meijer et al. (2005, 10), such as technological uncertainty, resource uncertainty, uncertainty regarding labour and market, consumer uncertainty, competitive uncertainty, political uncertainty, and supplier uncertainty. I will focus on the domain of technological uncertainty. Technological uncertainty can be subdivided in three elements, Meijer et al. say. (2005, 11) First, uncertainty about the technology itself. Decisions we make are influenced by the perception of the technology itself. Moreover, the uncertainty concerning individual technological innovations is influenced by the complexity of the technology. Second, uncertainty about the relation between the technology and the technological system, i.e. the complex of technologies that it is part of. Third, uncertainty about the availability of alternative technological

solutions, which connects with both technologies that are already available as well as technologies that might become available in the future.

In the final two sections I will illustrate how this typology might improve on the understanding of uncertain technology developments and their ethical evaluation. Introducing the typology of uncertainty sets the scene for two important insights and conclusions that will be addressed. First, in Section 10.5 it is demonstrated that uncertainty relates to the whole of the situation that is to be evaluated, namely the object of evaluation (i.e. the technology), the subject of evaluation (i.e. the actor), and the framework of evaluation (i.e. ethical theory). Section 10.6 concludes with some closing remarks that purport to be an outline for any future ethics of technology that is able to deal with uncertainty.

10.5 The Uncertainty of Ethics

In this chapter I observed that uncertainty is a concept un(der)developed in ethics. As argued in Section 10.2 the practical aim of ethics is to guide correct moral reasoning by demonstrating *how* right-making criteria apply to objects of assessment, in casu complex technology developments surrounded by uncertainty. Where ethical theories generally focus on rather static and calculable ideal-typical situations to demonstrate their practicability, complex, real life cases seem to obscure and frustrate this aim. Complex technology developments do not suffice the descriptions of ideal-typical ethical cases due to the uncertainty of future consequences, impacts, and applications. In view of this uncertainty, complex technology developments lack substantial information that is needed for the ethical evaluation. This deficiency in information fundamentally frustrates the practical aim of ethics. Should we infer from this that only a retrospective type of ethics of technology is feasible for technology development, or is there any potential to work towards an ethics of technology that allows for proactive, *ex-ante* moral evaluations of technology development? Is an ethics of technology able to incorporate a proactive perspective on uncertain situations; and, if so, to what extent? In retrospect of this study the prospects for an ethics of technology seem discouraging, but the question raised might benefit from insights of other scientific disciplines, in which uncertainty is a central topic of debate and reflection. The typology of uncertainty is useful for ethical theorising in relation to technology development (but might also prove to be interesting for existing technologies and other complex cases).

This typology and its three dimensions will be examined in relation to the case of complex technology development. At the start-off of every new development, an agent x (either individual or collective) is called upon to morally evaluate and justify the proposed development. Hence, at time t^1 agent x is to decide whether to develop a specific new technology, such as TUC, for which different development trajectories (o^1, o^2 … o^n) are possible. It is at t^1 expected of x to provide a proactive moral evaluation of an intended new technology which entails an overview of the different, potential development trajectories, the possible applications or uses,

and its impact and consequences at t^n, resulting in a well-considered judgment on its moral permissibility. The uncertainty of such situation relates to at least three aspects of ethics; namely to the agents or subjects (they are not omniscient), to the object of assessment (the uncertainty of future use and impact of technologies), and to theoretical issues (the uncertainty of the ethical framework used).

The first factor of uncertainty connects to *agents* who are not omniscient persons, but who are nevertheless confronted with the need to evaluate a new technology on basis of the information available; the information of which it is also the question whether agents correctly understand and interpret it. As argued in Section 10.3, facts like the involvement of many agents with different skills and from different disciplines or the long duration and opacity of development trajectories are central to many technology developments. In TUC many agents are involved in R&D; some are concerned with the conceptual design of optical switches, some with fundamental research in materials, and others with making proto-types, and so on. Accordingly, technology development is often too complex to be fully understood by a single agent. This first feature of uncertainty lies within agents themselves. The nature of this uncertainty is epistemic.

The second factor of uncertainty pertains to the *framework* of evaluation. In case of ethical theory the location of uncertainty is an identification of where uncertainty manifests itself in the framework of evaluation. The evaluative framework, consisting of rules, principles and/or values, is the starting point for reflection and judgement. This framework is not an objective or a perfect representation of reality. By this I mean that, although ethical theory purports to set forth statements with truth value about for instance the nature and status of moral judgments, it has an inherent uncertainty attached to it. For example, which data or inputs are accepted as relevant for the evaluation within a specific theory or meta-theory? In consequentialist theories only data are accepted that relate to the consequences of the actions and the central good that needs to be maximised. With regard to the theory itself, uncertainty might arise in applying certain principles and values to evaluate new technology developments. In connection with meta-theory, uncertainty relates to the foundations of theory. Choosing a meta-theory, in casu meta-ethics, necessarily implies certain presuppositions on the nature and status of moral judgements. It is important to clarify and explicate this uncertainty. The nature of this uncertainty is ontological, i.e. it is inherent to a specific framework and/or reality.

The third factor of uncertainty concerns the object of assessment, such as possible *development trajectories*, which connect to the levels of uncertainty. As we have seen, technology developments, especially complex variants, are characterised by uncertainty. This uncertainty is, however, not always of the same level. For certain developments it might be reasonably argued that trajectories o^1, o^2, o^5, and o^7 are likely on basis of available information, but that o^3, o^4, o^6, o^8, and o^9, are improbable. For example, research in optical switching and ultra-fast communication is a specific area in complex technology development in which engineers are able to foresee that in the future Internet traffic will increase (notably with the advent of (3D) video streaming). This specific information implies that certain trajectories are likely, but that others are moot. Other technology developments might involve

greater uncertainty, or might even be indeterminate. Fundamental research into materials to investigate which materials best conduct current under suboptimal circumstances is such an example. In this case we are confronted with a high degree of indeterminacy. The level of uncertainty is not an absolute measure, but represents different stages of (in)determinacy; on the one side information is available (determinacy) and on the other it is lacking (indeterminacy).

Uncertainty is wrongfully underappreciated in ethics, but by introducing this typology and its distinctions we arrive at a clearer picture of the concept of uncertainty for ethics. This is not to imply that uncertainty has become unproblematic. For a variety of reasons, this typology is a first contribution in overcoming this deficiency. It contributes to conceptual clarity about the dimensions of uncertainty and it shows the nuances of uncertainty in order to prevent uncertainty to be black boxed and reduced to risk. It shows where and how uncertainty manifests itself in the complex of technology development and its ethical assessment. It results in a better comprehension and overview of the role of uncertainty in ethics and technology. Uncertainty is, for instance, not simply the lack of knowledge, but it also relates to the inherent uncertainty in frameworks and reality. These insights about uncertainty raise fundamental questions — for instance with regard to the subject, the object, and the framework of evaluation — for any ethical framework that purports to assess new technology developments. Some of these questions with regard to the framework of evaluation will be outlined in the final section.

10.6 The Ethics of Uncertainty – An Outline for Any Future Ethics of Technology

Based on the previous sections, I will propound a number of questions as to what consequences uncertainty yields for the framework of evaluation, in casu ethical theory in its practical aim of proactively evaluating new technologies and in its theoretical aim of justifying theories and principles. One of the most pregnant questions relates to how uncertainty, that is among other things a lack of information, bears upon the practical aim of guiding correct moral reasoning. I will scrutinise the possibilities for ethical theory to deal legitimately with uncertainty, that is, even in the absence of relevant information for the ethical evaluation. I will argue that both pure substantive and pure procedural theories fail to guide proactive moral reasoning under conditions of uncertainty and that, as a result, we should endorse a substantive theory that incorporates a procedural solution.

To arrive at this conclusion I will first discuss uncertainty in relation to substantive and procedural theories.[1] So, how does the distinction between substantive and procedural ethical theories bear upon our aim of ethically evaluating uncertain

[1] For a more detailed exposition of (1) the relation between uncertainty and ethics and (2) the nature of and differences between substantive and procedural ethics, see Sollie (2008) and, especially, Sollie (forthcoming 2009).

technology developments proactively despite our lack of information? First, the lack of information, as a resultant of uncertainty, to apply right-making criteria foremost connects to *substantive* ethics. Substantive ethical theories specify what is right in terms of duties, rights, and values in relation to the object of assessment that are to be satisfied independent of any decision-making procedure. Substantive theories are theories for which *we* are in need of information to be able to evaluate technologies. Information is a prerequisite for us to be able to generate determinate answers when applying substantive right-making criteria to objects of assessment — provided the ethical theory has been able to demonstrate its practicability of *how* the right-making criteria apply to the object of evaluation. Without information substantive accounts are rather unsuccessful to guide our proactive assessments. Pure substantive theories fail in this regard, not because they are inadequate (they need not be), but because *we*, as subjects of ethical theories, require information to apply right-making criteria in practical moral reasoning. Hence, it is not so much a problem of the inadequacy of substantive ethics as it is a problem for us to apply substantive principles under conditions of uncertainty; there is no inherent failure or shortfall in substantive ethics in this regard. Any attempt to provide such an ethics of technology based on a substantive ethical theory may seem daunting given the fact of uncertainty. The lack of information frustrates the aim of guiding proactive moral reasoning. Without information we are rather handicapped in using substantive accounts to guide pro-active assessments. For example, consequentialists require information for calculating and maximising the utility of different, possible states of the world. However, if in circumstances of uncertainty substantial information is lacking about the possible states of the world then we face serious problems in trying to realise our practical aim of guiding correct moral reasoning via consequentialism. Under uncertainty it is not only consequentialist theories that are subject to the limitations of a substantive approach, any substantive ethical theory is.

Does this imply that we should turn to pure procedural ethical theories in situations of complexity and uncertainty? Pure procedural ethical theories solely attend to the process of moral decision-making instead of focusing on the substance. One could therefore argue that in absence of information we should resort to pure procedural theories. The problem with using pure procedural theories to assess uncertain technology development concerns the fact that such theories only focus on the procedures of communication and debate without regard to the substance. It cannot guide action without being supplemented with a material principle, which, in fact, draws our attention to the need for a substantive account. In the end, it is only able to guide the way in which the debate is carried out by providing communicative tools and instructions for moral deliberation, but these do not bear on the substance of the debate. Despite the fact that the lack of information is less a concern for procedural ethics as it is for substantive ethics, pure procedural ethics can only say something about the structure of debate and not about the substance. It is not able to make moral judgments by guidance of moral principles based on the content of the subject matter. An additional problem accompanying pure procedural theories is of a moral-psychological nature. Why would people agree to the outcome via a procedural solution that is independent from any content? How will we trust a procedure that

is not grounded in any substance? We have no reason to trust a procedure without substance.

These are major deficits in procedural approaches and therefore I maintain that we should strive for a substantive theory that is able to justifiably and legitimately include a procedural approach. The substance is required for guiding moral judgments, while the procedures can be invoked in cases in which we are confronted with uncertainty. Any future ethics of technology should therefore also be procedural of nature, not only substantive. A proactive evaluation under uncertainty necessitates a procedural turn. The upshot of this analysis is that pure substantive accounts are unsuitable for *us* to evaluate uncertain technology developments due to the lack of information. This still leaves unanswered the question as to what procedure is to be invoked and how this is legitimised and authorised. The problem with uncertainty *is* its uncertainty, its resultant lack of information and, for this reason, it is not so much a question about finding the right answer to uncertainty as it is to discovering an authorised and legitimate way of dealing with uncertainty within ethical theory that incorporates both a substantive and procedural approach. As said, uncertainty is not an intrinsic problem of the adequacy of ethical theory, but it prevents *us* from applying right-making criteria to complex cases *directly*. Consequently, the previous argument for a dual approach of substance and procedure is to present an *indirect solution*.

For this indirect solution to work and to be legitimate, any ethics of technology needs to show how procedures are justified within the ethical framework and how it conceives of the relation with substantive principles. Moreover, it needs to be demonstrated which procedure best fits the specific nature of technology development. In view of the central research topic, the scrutiny so far amounts to the formulation of three questions that any future ethics of technology needs to address and cope with in order to be adequate. The first two questions address the adequacy of ethical theory and the third questions stems from our aim of finding a legitimate approach that allows for a proactive assessment of technology developments that are characterised by uncertainty.

1. Theoretical aim of ethics: What right-making criteria underlie ethical theory and how are these right-making criteria justified?
2. Practical aim of ethics: How do these right-making criteria guide practical moral reasoning?
3. How does ethical theory deal with complexity and uncertainty, that is, how does it construe a justified relation between a substantive and procedural approach that is able to legitimately deal with uncertainty in a proactive manner?

This is the direction I am convinced an ethics of technology should be directed, because substantive accounts by themselves will inevitably prove to be insufficient to guide the assessment of complex and uncertain developments. Hence, what is required is a theory that satisfies all the demands put forward in this analysis. The aim of this article is not to present an answer — although I believe Gewirthian ethics to be a promising candidate for such an approach (Gewirth, 1978; Beyleveld, 1991; Beyleveld and Brownsword, 2007) as is further explored in Chapter 13 — but to lay

down arguments for the necessity of this research, the inevitability to scrutinise and take serious uncertainty in ethical theorising, and the subsequent consequences and requirements for any future ethics of technology.

Acknowledgment The author would like to thank the anonymous reviewers for their comments and suggestions as well as Deryck Beyleveld and Marcus Düwell for fruitful discussions that have helped to improve this paper. This article is a modified and reworked version of: Sollie, P. (2007) Ethics, Technology Development and Uncertainty: an outline for any future ethics of technology. In: *Journal of Information, Communication and Ethics in Society*, 5(4), 293–306.

References

BECK, U. 1999. *World Risk Society*. Cambridge: Polity Press.

BEYLEVELD, D. 1991. *The Dialectical Necessity of Morality: An Analysis and Defense of Alan Gewirth's Argument to the Principle of Generic Consistency*. Chicago: University Press of Chicago.

BEYLEVELD, D. and BROWNSWORD, R. 2007. *Consent In The Law*. Oxford: Hart Publishing.

BREY, P. 2001. Disclosive Computer Ethics. *In:* R.A. SPINELLO and H.T. TAVANI, (eds). *Readings in Cyberethics*, Sudbury, MA: Jones and Bartlett Publishers Inc. pp. 51–62.

BROWN, N., RAPPERT, B., and WEBSTER, A. 2000. Introducing Contested Futures: From *Looking into* the Future to *Looking at* the Future. *In:* N. BROWN, B. RAPPERT, and A. WEBSTER, (eds). *Contested Futures. A Sociology of Prospective Techno-Science*. Aldershot: Ashgate Publishing Ltd. pp. 3–20.

CILLIERS, P. 2005. Complexity, Deconstruction and Relativism. *In: Theory, Culture & Society*, **22**(5), pp. 255–267.

CLARKE, S. 2005. Future Technologies, Dystopic Futures and the Precautionary Principle. *In: Ethics and Information Technology*, **7**, pp. 121–126.

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

GEWIRTH, A. 1978. *Reason and Morality*. Chicago: The University of Chicago Press.

GEWIRTH, A. 1984. Replies to My Critics. *In:* REGIS, E. Jr. (ed.). *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*. Chicago: Chicago University Press, pp. 192–255.

HEALY, T. 2006. *The Unanticipated Consequences of Technology* [online]. [Accessed 18th October 2006]. Available from World Wide Web: <http://www.scu.edu/ethics/publications/submitted/healy/consequences.html>

JOHNSON, D.G. 1997. Ethics Online. *Communications of the ACM*, **40**(1), pp. 60–64.

KNIGHT, F.H. 1921. *Risk, Uncertainty, and Profit*. London: The Riverside Press.

MEIJER, I.S.M., HEKKERT, M.P., FABER, J., and SMITS, R.E.H.M. 2005. Perceived Uncertainties Regarding Socio-Technological Transformation: Towards a Typology. *In: Proceedings DRUID (Danish Research Unit for Industrial Dynamics) Winter Conference on Industrial Dynamics, Innovation and Development*, Skørping, Denmark.

MOOR, J.H. 1985. What is Computer Ethics? *Metaphilosophy*, **16**(4), pp. 266–275.

OSBORN, D. 2002. Stretching the Frontiers of Precaution. *Ethics in Science and Environmental Politics*, pp. 37–41.

PARFIT, D. 2006. *Climbing the Mountain*. (accessed through the Internet: http://individual.utoronto.ca/stafforini/parfit/parfit_-_climbing_the_mountain.pdf, last accessed: 1 October, 2008).

RIP, A. 2006. The Tension between Fiction and Precaution in Nanotechnology. *In:* E. FISHER, J. JONES and R. VON SCHOMBERG, (eds). *Implementing the Precautionary Principle: Perspectives and Prospects*. Cheltenham, UK and Northampton, MA, US: Edward Elgar. pp. 270–283.

SOLLIE, P. 2005. Technology and the Control Dilemma. Exploring an Ethics of Technology. *In:* P. BREY, F. GRODZINSKY, and L. INTRONA, (eds). *Ethics of New Information Technology. Proceedings of the Sixth International Conference of Computer Ethics: Philosophical Enquiry*, 17–19 July 2005, Enschede. Enschede: CTIT. pp. 331–343.

SOLLIE, P. 2008. Don't blame it on the principles! Uncertainty and uniqueness in ethical technology assessment. In: T.W. BYNUM, M. CALZAROSSA, I. DE LOTTO and S. ROGERSON, (eds). *Proceedings of the Tenth International Conference. Living, Working and Learning beyond Technology, Ethicomp 2008*. Mantua: Tipografia Commerciale: pp. 430–439.

SOLLIE, P. (forthcoming 2009). *The Uncertainty of Ethics and the Ethics of Uncertainty*. (PhD-thesis). Utrecht: Utrecht University.

TENNER, E. 1996. *Why Things Bite Back. Technology and the Revenge of Unintended Consequences*. New York: Random House.

TIMMONS, M. 2002. *Moral Theory: An Introduction*. Lanham: Rowman & Littlefield Publishers, Inc.

VERBEEK, P.-P. and SLOB, A. 2006. Technology and User Behavior. An Introduction. *In:* P.-P.VERBEEK and A. SLOB, (eds). *User Behavior and Technology Development. Shaping Sustainable Relations Between Consumers and Technologies*. Dordrecht: Springer. pp. 3–12.

VON SCHOMBERG, R. 2007. From The Ethics of Technology Towards an Ethics of Knowledge Policy & Knowledge Assessment. A working document of the services of the European Commission. Research Directorate General of the European Commission, Brussels.

WALKER, W.E., HARREMOËS, P., ROTMANS, J., SLUIS, J.P. van der, ASSELT, M.B.A. van, JANSSEN, P. and KRAYER VON KRAUS, M.P. 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-based Decision Support. *In: Integrated Assessment*, **4**(1), pp. 5–17.

WECKERT, J. and MOOR, J. 2006. The Precautionary Principle in Nanotechnology. *In: International Journal of Applied Philosophy*, **20**(2), pp. 191–204.

Chapter 11
New Technologies, Common Sense and the Paradoxical Precautionary Principle

Steve Clarke

Abstract I examine different forms of the Precautionary Principle (PP) to see if these are suitable, *inter alia*, for the regulation of new technologies. Weak versions of the PP may be suitable, but are not importantly different from Cost-Benefit Analysis (CBA). Strong versions of the PP are importantly distinct from CBA but are not a suitable basis for regulation because they lead to paradoxical outcomes if applied consistently. I consider three different lines of response to the change of paradox and argue that all three are unsatisfactory. First, I argue against Sandin's (2007) suggestion that we should be optimistic about finding a solution to the paradox on the grounds that the PP appears to be embodied in common sense reasoning. Second, I consider Weckert and Moor's (2006) attempt to resolve the PP paradox by appealing to the distinction between positive, negative and intermediate duties. I argue that this does nothing to resolve the PP paradox in many crucial cases. Furthermore, even when it can be used to resolve the paradox, it does not provide a satisfactory resolution. Third, I argue that Gardiner's (2006) attempt to recast the PP as a form of maximin is unsatisfactory because, although it resolves the PP paradox, it can only be successfully applied in a range of cases which is much narrower than the range in which advocates of strong versions of the PP typically attempt to apply the PP.

Keywords Common sense · Cost-benefit analysis · Duties · Maximin · Paradox · Precautionary principle · Risk · Uncertainty

11.1 Introduction

Recent developments in a range of areas of science including artificial intelligence, nanotechnology, biochemistry and photonics promise to enable us to develop new technologies that may improve our lives dramatically. However, the use of these new technologies may involve risks to our health and safety as well as risk of damage

S. Clarke (✉)
Faculty of Philosophy, University of Oxford, Oxford, United Kingdom
e-mail: stephen.clarke@philosophy.ox.ac.uk

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_11,

to the environment. How should governments and regulatory agencies attempt to formulate policies to manage the risks involved in the use of such new technologies? One influential approach that they may take is to apply cost-benefit analysis (CBA) to different available policy options. Under CBA we attempt to determine the probability of benefits occurring as well as the probability of costs being incurred, given a particular form of regulation of a new technology. We then compare the relative balance of costs and benefits for that policy option with other viable alternatives and select the outcome with the best overall balance of costs and benefits, adjusting for the probabilities of these occurring.

A very different approach is to apply the precautionary principle (PP) to policy options. The PP is a conceptual tool that was applied in environmental law, but which is now applied in a variety of contexts including healthcare and the regulation of new technologies. It made its initial appearance in Sweden and the former West Germany in the late 1960s (Sunstein, 2005: 16) and has become increasingly influential throughout Europe.[1] Indeed, it is sometimes held that there is a fundamental divide between America and Europe in respect to risk management; Americans typically applying CBA and Europeans applying the PP (Sunstein, 2005: 13–14).[2] CBA is not without is problems.[3] However, the focus of this paper is not on these problems, but on a cluster of problems that affect influential versions of the PP. Although it is common to hear references made to *the* PP, there is no one PP. There are many different formulations of the PP. The following three are influential examples of quite different formulations of the PP.

[1] Principle 15 of the 1992 *Rio Declaration on Environment and Development*:

> In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation. (United Nations Environment Programme, 1992)

---

[1] The PP was referred to in 27 resolutions of the European Union parliament between 1992 and 1999, is referred to in the 1992 Maastricht treaty on the European Union and has appeared in a draft constitution for the European Union (Sunstein, 2005: 17). According to Majone (2002) the European commission actively promotes the PP as a 'key tenet' of European community policy and as a general principle of international law.

[2] Although use of the PP is much more prevalent in Europe that in the US, this is not because Americans do not employ precautionary reasoning. Americans are more risk averse than Europeans when it comes to reasoning about particular issues, such as unemployment and the cost of energy (Sunstein, 2005: 14). Also, the American response to terrorist threats, in the wake of the events of September 11th 2001 amounts to an application of the PP (Stern and Wiener, 2006).

[3] CBA has been subjected to a wide range of criticisms, especially in environmental economics (e.g. Bromley and Paavola, 2002), but also in law and philosophy. In particular it has been extensively criticised for requiring the commensuration of apparently incommensurable costs and benefits (e.g. Adler, 1998) and for involving the monetary valuation of human life (e.g. Anderson, 1988). For a defence of CBA against a variety of criticisms see Schmidtz (2001). Hansson (2007) provides a recent catalogue of philosophical problems that are relevant to the application of CBA.

[2] Final Declaration of the First European 'Seas at Risk' Conference, 1994.

> If the 'worst case scenario' for a certain activity is serious enough then even a small amount of doubt as to the safety of that activity is sufficient to stop it taking place.[4]

[3] The Wingspread Statement:

> Where an activity raises threats of harm to the environment or human health, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically. In this context the proponent of the activity, rather than the public, should bear the burden of proof. (Wingspread, 1998).

[1] is an example of what is sometimes referred to as a weak version of the PP (wPP). It is not genuinely an alternative to CBA. In fact it is compatible with CBA. The main purpose of [1] appears to be to ensure that CBA is not used in a biased manner in which only those risks that are established with 'full scientific certainty' are considered.

[2] is an example of what is sometimes referred to as a strong version of the PP (sPP). It is incompatible with CBA. It stipulates conditions under which we should regulate on the basis of consideration of the potential costs of a policy, regardless of the potential benefits of that policy or the potential costs of alternatives. So it is antithetical to CBA, which involves a weighing of costs and benefits.

[3] does not focus directly on action but on identifying the 'burden of proof'. It can be understood as being supplementary to CBA rather than an alternative to CBA. Depending on what is required to meet the burden of proof it can resolve into a weak or a strong version of the PP, or something in between. If it does not ask more of the proponents of the activity than that they present the relevant evidence for the conclusion that the potential benefits of the activity in question exceed its potential costs then it resolves into a weak version of the PP. It effectively resolves into a strong version of the PP if it requires the proponent of an activity to establish that that activity carries no risk whatsoever of serious harm to the environment or to human health.

I've characterised CBA and the PP as competing approaches to risk management. In doing so it might be supposed that I have mischaracterised the PP. Defenders of the PP sometimes argue that the PP should be applied in situations of uncertainty,[5] whereas CBA is designed to be applied in situations of risk. The classic distinction between risk and uncertainty goes back to Knight (1921). On Knight's usage, risk refers to circumstances where the probabilities of potential outcomes can be specified, on the basis of reliable evidence, whereas uncertainty refers to circumstances where the probability of potential outcomes cannot be specified, on the basis of reliable evidence.[6] The rolling of an unbiased six-sided die is a clear case of risk without uncertainty as we appear to know what all the possible outcomes are

---

[4] Cited in Sunstein (2005: 29).

[5] See, for example, Sandin (1999: 892–894).

[6] This is the standard interpretation of Knight (1921). For a different view see LeRoy and Singell (1987).

and we have good grounds for specifying the probability of all of these occurring. Speculation about the possibility of an afterlife is a case of uncertainty without risk. We can imagine various possible different afterlives, but we have no obvious basis for determining the probability of any of these eventuating.

CBA is grounded on risk assessment, as is wPP, however, sPP involves no consideration of risks, but relies of consideration of the mere possibility of uncertainties. So, it might be thought that sPP is not in direct competition with CBA at all. However, there are very few cases of pure uncertainty and there are very few cases of pure risk, in Knight's senses. When an army goes to war its generals typically have a rational basis for assigning subjective probabilities to the likelihood of different strategies succeeding, even while they will acknowledge that war is beset by many uncertainties. And when an insurer writes a car insurance policy they do so on the basis of an assessment of the risks of a particular driver having an accident, even though they are aware that the act of driving is beset by many uncertainties. Most real world circumstances involve both risks and uncertainties.[7] CBA and sPP are employed by considering different aspects of real world circumstances, but they can also be applied to many of the same real world circumstances. So, despite appearances, they are often in direct competition.

The presence of a variety of different formulations of the PP is responded to very differently by advocates of the PP. Some such as Gardiner (2006) and Weckert and Moor (2006), set themselves the task of identifying a core formulation of the PP which captures the guiding idea behind the PP, while being most able to respond to the various critics of the PP. However, others rejoice in the diversity of formulations of the PP. Jordan, Andrew and O'Riordan (1999) have the following to say about the PP:

> Like sustainability, it is neither a well-defined nor a stable concept. Rather, it has become the repository for a jumble of adventurous beliefs that challenge the status quo of political power, ideology, and environmental rights. Neither concept has much coherence other than is captured by the spirit that is challenging the authority of science, the hegemony of cost-benefit analysis, the powerlessness of victims of environmental abuse, and the unimplemented ethics of intrinsic natural rights and intergenerational equity. It is because the mood of the times needs an organising idea that the Precautionary Principle is getting attention.
>
> (Jordan, Andrew and O'Riordan, 1999: 16).

The inclusiveness of a vague, shifting concept may be useful as an organising idea for a diverse protest movement, but once the PP becomes used widely as a tool of policy formation it becomes important that the PP – or at least particular formulations of the PP – are well-defined. If policies are to be guided by the PP, and make reference to the PP, then it is important that the PP can be understood in a consistent manner. If not then we will be unable to secure agreement as to how to implement any particular policy that invokes the PP.[8]

---

[7] I've argues that real world circumstances often involve both risks and uncertainty. Many commentators argue, on somewhat similar grounds, that there is no well-founded distinction between risk and uncertainty. See for example, Friedman (1962).

[8] I argued similarly in Clarke (2005: 125). See also Gardiner (2006: 40).

11.2 The PP Paradox

If proponents of the PP are to identify a core formulation of the PP, for the purposes of enabling policy formation, then it looks as if they will have to make a basic choice between wPP and sPP. Because variants of wPP simply offer guidance regarding the proper use of CBA, effectively the core formulation of wPP will be a core formulation of CBA. There may be interesting issues about the proper formulation of CBA, but they are beyond the scope of this discussion. The latter choice requires that the proponents of the PP overcome what has emerged as the most serious objection to sPP, which is that, if properly applied, it produces paradoxical outcomes. In the remainder of this section and also in the next one, I will explain the PP paradox and discuss reasons why people are often unconcerned by the paradoxical nature of sPP. In the final two sections of the paper I will consider two recent attempts to resolve the PP paradox, due to John Weckert and Jim Moor (2006) as well as Stephen Gardiner (2006).[9]

Strong versions of the PP, if applied rigorously, lead to the paradoxical situation where we are neither permitted to perform nor permitted to fail to perform an action under consideration. This well known paradox is expressed somewhat differently by different scholars. One formulation is due to Manson who considers the PP paradox to be analogous to the 'Many Gods' objection to Pascal's Wager (2002: 272).[10] He argues that any precautionary measure that we might take to avoid harm, itself involves risk of harm. Consider, for example, a situation in which the US Government is deciding what to do to respond to the threat of potentially disastrous environmental damage, due to climate change. A possible response would be to reduce carbon emissions significantly, as per the Kyoto Protocol. And it seems plausible to think that an application of sPP would lead to the recommendation that the US Government accepts and applies the Kyoto protocol. However, enforcement of the Kyoto protocol is not without the risk of harms itself. A significant reduction in carbon emission in the US cannot plausibly be achieved, in the near future, without significant changes to the lifestyles of ordinary citizens. There are risks involved in such social and economic change. Perhaps global economic depression will be caused, chaos will ensue, and the current democracy in America may be replaced by a dictatorship. So applications of sPP lead to the contradictory recommendation

[9] In Clarke (2005: 123–124) I examined two earlier attempts to address the PP paradox, due to Sandin et al. (2002) and Saunders and Ho (2000). I argued that neither of these succeed.

[10] Pascal famously argued that one should choose to believe in God, no matter how low one believes the chance of God actually existing is (provided that it is believed to be greater than zero). This is because the potential benefits of belief, if God exists, are enormous (heaven) and these are only available to those who believe. Furthermore the costs of belief are very low. The 'Many Gods' objection is an influential objection to Pascal's reasoning. Suppose that the deity Thor exists and Thor will only allow those who believe in him and in no other deity into heaven. If this is the case then belief in God will lead one to be denied entry into heaven. But it seems possible that Thor exists. So, Pascal's reasoning (applied to many deities) leads both to the conclusion that one should believe in God and the conclusion that one should not believe in God.

that the US Government should act in response to climate change and that it should not act in response to climate change.[11]

Sunstein sets up the PP paradox somewhat differently from Manson. According to him, regulation guided by sPP will itself fall afoul of sPP because '... it might well deprive society of significant benefits and hence produce serious harms that would otherwise not occur' (Sunstein 2005: 29). So, for example, taking a precautionary approach to the regulation of a new drug, by insisting on stringent testing before the drug is made available to the pubic, will deprive society of a benefit, the benefit of having an experimental drug available. This denial of benefit is itself potentially harmful because, if the experimental drug is in fact efficacious in curing or preventing a disease, then lives that could have been saved by use of the drug will be lost.[12]

Given that it is well known that sPP leads to paradoxical outcomes, if applied consistently, why do people continue to attempt to apply sPP? Part of the answer to this question is that the PP lends itself to inconsistent application because, unlike CBA, it is not a comparative concept. When we apply standard formulations of the PP we apply them to particular proposed actions and not to the comparison of alternative courses of action. So, provided that it does not occur to us to apply sPP to the course of action that an application of sPP recommends, we do not find ourselves in paradox. A second part of the answer, supplied by Sunstein (2005: 35–63) is that our thinking is often governed by heuristics that result in a variety of cognitive biases that can cause inconsistent application. Particularly important in understanding lay precautionary reasoning is the availability heuristic, first identified by Tversky and Kahneman (1973).

The availability heuristic is a rule of thumb that people use to provide intuitive assessments of the magnitude of a risk. So when people are asked, for example, how serious is the risk of an accident at a nuclear power plant, a major terrorist attack on their own country, or the outbreak of a new disease, they will tend to make such assessments by equating the magnitude of risk with their ability to bring instances of the particular threat in question to mind. Cognitive availability will be affected, *inter alia*, by familiarity and by salience. If an instance of a class of dangerous events has occurred recently, people will intuitively increase their assessment of the likelihood that it will be repeated, particularly if their awareness of its occurrence has been reinforced by the media. Use of the availability heuristic can lead to serious overestimations of the magnitude of risk, especially in circumstances where risks are associated with excessive public fear. Because there is a limited number of topics that are highly 'available' to us at any given time, when one risk is highly available to us, we are likely to fixate on it and neglect other relevant risks which are less

[11] The Kyoto example is from Manson (2002: 273).

[12] Sunstein (2005) does not appear to differentiate this form of the PP paradox from the Manson formulation. He goes on to describe another example of the PP paradox in which the relocation of 270,000 people in response to the risks of radiation exposure following the Chernobyl incident – an instance of precautionary reasoning – leads to direct harms (psychological harms) rather than to the deprivation of benefits.

available to us. For example, if we are focussed on threats to the environment posed by climate change, when these are highly available, we are liable to neglect other risks, including risks to the economy and social stability that are posed by alterations to our lifestyle that would be required to minimise carbon emissions and thereby reduce the threat to the environment posed by climate change.

11.3 Common Sense and Precaution

It is sometimes thought that even if sPP does lead to contradictory paradoxical policy recommendations, we should be optimistic about finding some way out of the paradox, because the precautionary reasoning that seems to underpin all forms of the PP seems also to be embodied in common sense reasoning (Sandin, 2007). Common sense reasoning generally appears to enable us to make definite decisions and does not often appear to lead us to agonies of indecision brought on by paradox. And it appears that common sense endorses something like the PP. Common sense proverbs assure us that we are 'better safe than sorry' and that 'an ounce of prevention is worth a pound of cure' (Sandin, 2007: 105). I will argue, though, that the PP probably has less in common with common sense reasoning than appearances suggest and that it is plausible to think that CBA has much in common with common sense reasoning.

Sandin provided us with the following example of common sense decision making that appears to him to be based on something like the PP:

> I am hiking in the mountains, and I am thinking of having a drink of water from a small stream. I do not know that the water is safe to drink. There just might be a reindeer carcass a short distance upstream, poisoning the water. Thus, as the saying goes, 'When in doubt, don't'. I abstain from drinking the water, or at least boil it before drinking it. (Sandin 2007: 99).

In this example of common sense decision making, a risk that might result from an action is explicitly considered, while the possible benefits of action are not explicitly considered and the risks of inaction are not explicitly considered. If Sandin was also to consider the risks of inaction as well as the potential benefits of action, then his decision making would seem to be based on something more like CBA than the PP. But these factors do not appear to be considered, so it appears that Sandin's common sense reasoning has much more in common with the PP than CBA. However, the appearance that this example of common sense reasoning, as well as others, has more in common with the PP than CBA may be nothing more than apearance. It is very plausible to think that much of our ordinary decision making occurs 'offline' and that conscious explicit reasoning is only part of the overall process that leads to a decision being made (Bargh and Chartrand, 1999; Schneider and Shiffrin, 1977). Deliberative conscious reasoning and intuitive, non-conscious processing combine to create overall decisions, or so say 'dual processing theorists', an increasingly influential school of thought in social and cognitive psychology (Chaiken and Trope, 1999; Kahneman and Frederick, 2002).

Dual Processing theory is a relatively new trend in psychology and there is no agreed upon model of how deliberative conscious reasoning and offline, intuitive processing are integrated in ordinary decision making. One possible means of integration is that a decision maker consciously focuses on an important aspect of a decision, while non-conscious parts of her brain attend to other aspects of a decision and draw these to the attention of consciousness if and when they become important. Sandin consciously attends to the risk of drinking poisoned water and at the conscious level he ignores the risks involved in failing to take a drink. However, it is plausible to think that at the non-conscious level these considerations are being monitored. There are risks involved in not drinking. Perhaps by failing to take a drink Sandin will become dehydrated and suffer from dizziness, fainting or even death. Sandin ignores these factors at the conscious level but not at the non-conscious level. If he is in significant danger of dehydration a non-conscious part of his brain will send a message to his conscious mind – typically in the form of a strong desire to drink – which can override the decision to avoid the risk of drinking. Also, there may be possible benefits of action. Perhaps Sandin's attractive hiking partner will be impressed by his willingness to risk drinking from the stream and fall in love with him – an outcome which may please him greatly. If social cues indicate that is indeed a significant possibility, then a non-conscious part of his brain may also send a message to his conscious mind, urging him to risk taking a drink so as to show off to his hiking partner.

Once we take account of the non-conscious aspects of ordinary decision making, the common sense reasoning that takes place in this example starts to look a lot less like the PP than it did when we only considered the conscious aspects of decision making. If information about the potential benefits of an action as well as information about the potential risks of a failure to take the action in question are processed non-consciously and if this information contributes to an overall decision, then that overall decision begins to look a lot more like CBA than the PP. I am not arguing that common sense thinking about risk is definitely more like CBA than the PP. The relevant science is not yet advanced enough to establish such a conclusion. However, the case for thinking that common sense thinking about risk is similar to the PP has definitely not been proven and indeed it looks to be difficult to square with recent work in social and cognitive psychology.

11.4 Weckert and Moor

John Weckert and Jim Moor claim that we can resolve the PP paradox by invoking the distinction between positive, negative and intermediate duties (2006: 200). They set up the paradox as follows:

1. Action A1 might cause bad effect Eb1 (harm eventuates because of A1)
2. Remedy R1 (don't do A1) stops Eb1 (PP applied)
3. But suppose that A1 causes good effect Eg1 (Eg1 eliminates some harm)
4. Then R1 stops Eg1 (harm eventuates because of R1)

5. So, if PP should be applied to A1 (because A1 causes harm) it should also be applied to R1 (because R1 prevents an action that would eliminate some harm) (Weckert and Moor, 2006: 197).

Weckert and Moor go on to argue that there is an in principle way of choosing between applying the PP to A1 and R1 (2006: 199). We have positive duties to do good and negative duties not to harm. As they note, it is widely held in moral philosophy that negative duties are generally more compelling than positive duties. Failing to save the life of a drowning child in our immediate vicinity, when we could easily do so, is a reprehensible omission. But causing a child to drown is worse. As well as positive and negative duties there are intermediate duties. These are duties to 'avert harms that one's past conduct may cause in the future . . .They are positive insofar as they require the agent to do something and negative insofar as this requirement is continuous with the duty to avoid causing harm to others' (Pogge, 2005: 34). If we were instrumental in causing a child to be in a position where she was drowning – for example, by removing a sign that she might have seen, warning of the presence of deep water – then we have a more compelling duty to save her life than we would if we had simply stumbled across her while she was drowning.

A ranked order of positive, negative and intermediate duties can be used as a way of avoiding paradoxical conclusions issued by the PP in circumstances where the application of the PP leads to conflicting policy recommendations, and in which the implementation of those conflicting policy recommendations can be understood as involving different types of duties.[13] So Weckert and Moor are right that their approach can lead to a resolution of at least some paradoxes that are thrown up by the application of the PP. But does their approach provide solutions to all instances of the PP paradox, as they appear to claim? And for the cases that it does provide solutions, are these satisfactory solutions? I will argue that the answer to both of these questions is no.

The Weckert-Moor solution to the PP paradox is of no help in cases where conflicting duties are of a similar type. In cases where we have a negative duty to perform action x and a negative duty to perform action ∼x or an intermediate duty to do x and an intermediate duty to do ∼x then the Weckert-Moor solution does not relieve us from paradox. Are there such cases? Yes there are. Manson's (2002) aforementioned example in which the US Government is faced with the dilemma of whether or not to comply to the Kyoto protocol appears to be one such case. It seems that the US Government has an intermediate duty to comply with the Kyoto protocol, as Americans significantly contribute to the fossil fuel emissions that are driving climate change. However, if they do comply with the protocol, then they may risk causing massive civil unrest, as a response to the resulting economic hardships

[13] Weckert and Moor (2006: 199) attribute the view that intermediate duties are stronger than positive duties and weaker than negative duties to Pogge (2005). However, in the passage that they quote Pogge (2005: 34) only claims that intermediate duties are more stringent than positive duties. He makes no claim there about their strength relative to negative duties.

that civilians will be required to endure. As the US Government is significantly responsible for the social circumstances that dispose its citizens towards civil unrest in such circumstances, it seems that they have an intermediate duty not to adhere to the Kyoto protocol.

Perhaps the reason that Weckert and Moor fail to consider Manson's well known example is that the way that they set up the PP paradox is consistent with Sunstein's (2005) formulation rather than Manson's (2002) formulation. They stipulate that the paradox occurs in cases where taking a precautionary action causes harm in so far as it prevents a 'good effect'. Typically, the avoidance of harm, being a negative duty is more important than the failure to cause benefit, which is usually the failure to perform a positive duty. So in standard cases, under this formulation of the paradox, we have a means of resolution. However, in cases that do not fit Weckert and Moor (2006) and Sunstein's (2005) formulation of the paradox, where the harms in question go beyond the mere failure to provide benefits, we are liable to be faced with a comparison of instances of the same type of duty.

Weckert and Moor (2006) are right about this much; it is commonly accepted in moral philosophy that, all things being equal, negative duties and intermediate duties are more important than positive duties. But not many moral philosophers are willing to allow that negative duties are always more important than intermediate duties and that these are always more important than positive duties.[14] Consider a case in which causing an innocent person to be harmed will save many innocent lives. Suppose, for example, that an innocent but curious civilian has wandered into a situation where a bomb is being defused, and in a well meaning but misguided attempt to help, is about to set off the bomb inadvertently. The only way I can prevent this from happening is by throwing a nearby rock at the curious civilian, knocking him unconscious. It seems very plausible to think that, in this case, my positive duty to save the lives of many outweighs my negative duty to prevent harm to the curious civilian. Negative and intermediate duties may generally be more compelling than positive duties, but they do not always outweigh them. In general, the weighing of negative, intermediate and positive duties is a matter of judgment and a matter about which there are no simple rules. Weckert and Moor's solution to the PP paradox seems attractive in large part because it introduces a simple test to resolve the PP paradox. However, this simple test is not realistic. The weighing of different types of duties requires judgment and invites disagreement. So, on their solution, the PP would not provide clear guidance to decision makers. Instead, decision makers would have to make a delicate and possibly controversial moral judgment, in order to implement the PP.

[14] Indeed Pogge, whom they cite as the source of the distinction between positive, intermediate and negative duties is at pains to insist, against earlier critics, that he does not intend that negative and intermediate duties are always more important than positive duties (2005: 34–35).

11.5 Gardiner and the Rawlsian Core Precautionary Principle

Recently Stephen Gardiner (2006) has proposed a new form of the precautionary principle, which he refers to as the 'Rawlsian Core Precautionary Principle' (RCPP).[15] The RCPP is based on John Rawls (1999) maximin principle. To apply maximin one compares the worst potential outcomes of a range of possible policies and selects the policy which has the least bad worst potential outcome (Gardiner, 2006: 45). RCPP is significantly more precautious than wPP. It is closer in spirit to sPP than wPP because decisions under RCPP turn on a consideration of risks and not on a balancing of risks and benefits.

Applications of maximin appear to proceed quite differently from applications of the PP, conventionally understood, and some critics have complained that a principle based on maximin would be too different from standard instances of the PP to count as a genuine instance of the PP.[16] Whether or not RCPP is best understood as a form of the PP or an alternative to the PP, it has a clear advantage over all other forms of the precautionary principle because it is a comparative principle. Under RCPP we compare a range of policy alternatives and select one of them.[17] So we are able to avoid the paradox that arises when we employ a strong version of the PP.

Maximin is often objected to on the grounds that it seems to produce the wrong result if applied in circumstances in which benefits matter. Consider the following example due to Harsanyi (1975).

> Suppose you live in New York City and are offered two jobs at the same time. One is a tedious and badly paid job in New York City itself, while the other is a very interesting and well-paid job in Chicago. But the catch is that, if you want the Chicago job you would have to take a plane from New York to Chicago (e.g., because this job would have to be taken up the very next day). Therefore there is a very small but positive probability that you might be killed in a plane accident. (Harsanyi 1975: 39)

An application of maximin in this example would lead to the counterintuitive recommendation that you should remain in the tedious and badly paid job in New York. Rawls (1999) and Gardiner (2006) do not dispute that maximin would deliver the wrong result if applied in the above scenario. Instead they argue that maximin should only be applied in a limited range of circumstances that does not include the above example.

---

[15] An earlier proposal to base the PP on maximin is due to Hansson (1997).

[16] See Sandin (2007: 102) and Sunstein (2005: 61).

[17] It may not always be possible to make clear comparisons, because the least bad worse potential outcomes of some of the competing policy options may be incommensurable. One policy option in climate management might have extreme global warming as its wost outcome, while another may have extreme famine as its worst outcome. We may be unable to agree as to whether extreme global warming is a worse outcome than extreme famine and there may be no rational way of comparing these two worst potential outcomes. Thanks to the editors for pointing out the importance of this issue.

Gardiner paraphrases Rawls, who identifies three conditions of applicability as follows:

[1] . . . decision makers either lack, or have reason to sharply discount, information about the probabilities of the possible outcomes of their actions.
[2] . . . the decision makers care relatively little for potential gains that might be made above the minimum that can be guaranteed by the maximin approach.
[3] . . . the decision makers face unacceptable alternatives (rejected alternatives have outcomes that one can hardly accept) (Gardiner, 2006: 47).

Gardiner then adds a fourth criterion of acceptability to the RCPP, which is that we should only consider uncertain but realistic outcomes, and not unknown theoretical outcomes (2006: 51–52). This criterion is brought in to head off the objection that like sPP, RCPP can counsel us to act to avoid very unlikely but possibly catastrophic, purely theoretical outcomes. Gardiner does not make much headway in the tricky task of finding a principled way of distinguishing between realistic and unrealistic outcomes, as he himself admits (2006: 52).[18]

The main problem with RCPP is that it appears only to be applicable in a range of circumstances that is much narrower than the range of circumstances in which proponents of the PP usually attempt to apply the PP. The most limiting of the four conditions of applicability is [2], the stipulation that decision makers must happen to 'care relatively little for potential gains'. Often times the PP is invoked when decision makers do care about gains, but nevertheless argue that we should set consideration of these aside and focus on potential losses. Consider, for example, debates about GM foods, a context in which the PP is frequently invoked. Proponents of GM foods sometimes argue that GM crops offer the potential to allow humanity to obtain a state of global food security and that this benefit outweighs any potential risks that the production of GM crops may lead to. Opponents of GM crops do not argue that we should care relatively little for global food security. Instead they argue that the chance of this occurring is low and that the risks of allowing GM crops to be grown are more significant and more likely to be realised than proponents of GM crops allow.

Gardiner is aware that this example does not appear to meet his conditions of applicability, but he suggests that appearances have deceived in this case. According to Gardiner, arguments by opponents of GM crops, for the claim that GM technology is unlikely to lead to global food security, as well as arguments for the claim that global food security could be obtained without the use of GM crops can be understood as arguments about whether condition [2] holds or not (Gardiner, 2006: 56). In other words, opponents of GM crops who appear to be arguing that GM crops are unlikely to bring about the desired aim, and unnecessary for the achievement of that desired aim are actually arguing that the desired aim is one that we should

[18] He is far from alone in this regard. For criticism of other attempts to make this distinction see Clarke (2005: 123–124) and Majone (2002: 104).

'care relatively little for'. This strikes me as a highly implausible way to understand these debates. The fact that opponents of GM crops go to some lengths to try to persuade us that GM crops will not provide global food security and that we could achieve it by other means suggests to me that they think this aim is one that we should care about a great deal, and they wish to rebut arguments due to proponents of GM crops who hold out the possibility of achieving an aim that almost everybody cares a great deal about. Contra Gardiner, RCPP is not applicable in this case, where the PP is often appealed to, and it is probably not applicable in many other such cases.

11.6 Conclusion

We have examined different forms of the PP, with a view to seeing if any of these would be suitable, *inter alia*, for the regulation of new technologies. Weak versions of the PP may be suitable, but they are not importantly different from CBA. Strong versions of the PP lead us to paradoxical outcomes if applied consistently. We examined Weckert and Moor's (2006) attempt to resolve the PP paradox and found it to be unsuccessful in important cases, and difficult to implement. Gardiner's (2006) attempt to rejig the PP on the basis of Rawlsian maximin was somewhat more successful. However, it appears that this solution is only applicable in a range of circumstances that is much narrower than the range of circumstances in which appeals to the PP have typically been made.

Advocates of wPP are often motivated by a concern that CBA is sometimes applied too narrowly. They are probably right to be so concerned. There are those who have wanted to insist that evidence of risk of potential harm must be conclusively established before it can be utilised in CBA and if they were to have their way then CBA would be applied extremely narrowly.[19] Despite their use of the term 'the precautionary principle', advocates of wPP are not establishing a new principle of risk management. Instead they are proposing ways to ensure that current tools of risk management are used properly. Advocates of versions of sPP really are proposing a genuine alternative to CBA, but it is a flawed alternative as it leads to paradox, if applied consistently. I do not want to rule out the bare possibility of this problem being overcome, but current efforts do not appear promising. The most successful attempt is that due to Gardiner (2006). However, Gardiner has had to resort to radical surgery to save sPP – turning it into a variant of maximin – and even this is only applicable in a very narrow range of cases, much narrower than the range of cases in which advocates of sPP have attempted to apply sPP.[20]

---

[19] According to Cranor (2004), Frank Cross takes this view.

[20] Thanks to Marcüs Duwell, Jim Moor, Rebecca Roache, Paul Sollie and John Weckert for helpful comments.

References

Adler, Matthew. "Incommensurability and Cost-Benefit Analysis", *University of Pennsylvania Law Review* 1998, 146, 1371–1418.

Anderson, Elizabeth. "Values, Risks and Market Norms." *Philosophy and Public Affairs* 1988, 17, 54–65.

Bargh, John A. and Chartrand, Tanya L. "The Unbearable Automaticity of Being." *American Psychologist* 1999, 54, 462–479.

Bromley, Daniel, W. and Paavola, Jouni (eds). *Economics, Ethics and Environmental Policy: Contested Choices*. London: Wiley-Blackwell 2002.

Chaiken, Shelly and Trope, Yaacov. *Dual-Process Theories in Social Psychology.* New York: Guilford Press 1999.

Clarke, Steve. "Future Technologies, Dystopic Futures and the Precautionary Principle." *Ethics and Information Technology* 2005, 7, 121–126.

Cranor, Carl. "Toward Understanding Aspects of the Precautionary Principle", *Journal of Medicine and Philosophy* 2004, 29, 3, 259–279.

Friedman, Milton. *Price Theory: A Provisional Text.* Chicago: Aldine, 1962.

Gardiner, Stephen M. "A Core Precautionary Principle." *The Journal of Political Philosophy* 2006, 14, 1, 33–60.

Hansson, Sven Ove. "The Limits of Precaution." *Foundations of Science* 1997, 2, 293–306.

Hansson, Sven Ove. "Philosophical Problems in Cost-Benefit Analysis." *Economics and Philosophy* 2007, 23, 163–183.

Jordan, Andrew and O'Riordan, Timothy. "*Protecting Public Health and the Environment: Implementing the Precautionary Principle*." In *The Precautionary Principle in Contemporary Environmental Policy and Politics*. Carolyn Raffensperger, and Joel. A. Tickner (eds). Washington: Island Press, 1999, 15–35.

Kahneman, Daniel. and Frederick, Shane. "Representativeness Revisited: Attribute Substitution in Intuitive Judgement." In *Heuristics and Biases: the Psychology of Intuitive Judgement*. Thomas Gilovich, Dale Griffin and Daniel Kahneman (eds). Cambridge: Cambridge University Press, 2002, 49–81.

Knight, Frank. *Risk, Uncertainty, and Profit.* Boston, MA: Hart, Schaffner and Marx, 1921.

LeRoy, Stephen, F. and Singell, Larry D. Jr. "Knight on Risk and Uncertainty." *Journal of Political Economy* 1987, 95, 2, 394–406.

Majone, Giandomenico. "What Price Safety? The Precautionary Principle and its Policy Implications." *Journal of Common Market Studies* 2002, 40, 1, 89–109.

Manson, Neil, A. "Formulating the Precautionary Principle." *Environmental Ethics* 2002, 24, 263–274.

Pogge, Thomas. "Real World Justice." *The Journal of Ethics* 2005, 9, 29–53.

Rawls, John. *A Theory of Justice*. Cambridge MA: Harvard University Press, 1999.

Sandin, Per. "Dimensions of the Precautionary Principle." *Human and Ecological Risk Assessment* 1999, 5, 5, 889–907.

Sandin, Per. "Common Sense Precaution and Varieties of the Precautionary Principle." In *Risk: Philosophical Perspectives*. Tim Lewens (ed). London: Routledge, 2007, 99–112.

Sandin, Per, Peterson, Martin, Hannson, Sven Ove, Rudén, Christina, and Juthe, Andre. "Five Charges Against the Precautionary Principle." *Journal of Risk Research* 2002, 5, 287–299.

Saunders, Peter and Ho, Mae Wan. "The Precautionary Principle is Coherent." *ISIS Paper* 2000: http://www.biotech-info.net//PP_coherent.html. Last visit: 3 August 2007.

Schmidtz, David. "A Place for Cost-Benefit Analysis." *Philosophical Issues (A Supplement to Nous)* 2001, 11, 148–171.

Schneider, Walter and Shiffrin, Richard M. "Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention." *Psychological Review* 1977, 84, 1, 1–66.

Stern, Jessica and Wiener, Jonathan B. "Precaution against Terrorism." *Journal of Risk Research* 2006, 9, 4, 393–447.

Sunstein, Cass. *Laws of Fear: Beyond the Precautionary Principle*. Cambridge: Cambridge University Press, 2005.

Tversky, Amos and Kahneman, Daniel. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 1973, 5, 207–232.

United Nations Environment Programme. *Rio Declaration on Environment and Development*, 1992. www.unep.org/Documents.Multilingual/Default.asp?ArticleID=1163&DocumentID=78&l=en. Last visit: 3 August 2007.

Weckert, John and Moor, James. "The Precautionary principle in Nanotechnology." *International Journal of Applied Philosophy* 2006, 20, 2, 191–204.

Wingspread Statement on the Precautionary Principle, 1998. www.gdrc.org/u-gov/precaution-3.html. Last visit: 3 August 2007.

Chapter 12
Complex Technology, Complex Calculations: Uses and Abuses of Precautionary Reasoning in Law

Deryck Beyleveld and Roger Brownsword

Abstract After a honeymoon period in environmental law, the so-called "precautionary principle" has received sustained criticism. This paper does not try to rescue the precautionary principle as such. However, it is argued, using Pascal's Wager, that there are conditions under which precautionary reasoning is valid, which provides a general principle for the limiting case. Although the limiting principle does not apply straightforwardly to the principle that those accused of criminal offences should not be convicted unless found guilty beyond all reasonable doubt, this case suggests an alternative principle that employs precautionary reasoning in a proportionate manner. To apply the limiting principle involves difficult judgments about the relative undesirability of options presented and the proportionality of the precautionary response. Nevertheless, it is argued that the limiting principle provides a strong argument against the death penalty, and that precautionary reasoning is more widely involved in legal reasoning than generally appreciated; e.g., wherever the burden of proof is placed more strongly on one party, where the threshold in relation to a particular option is raised, and, perhaps, in slippery slope and floodgates arguments.

Keywords Precaution · Precautionary principle · Precautionary reasoning · Pascal · Gewirth · Burden of proof · Slippery slope · Floodgates · Proportionality

12.1 Introduction

According to the philosopher, James Moor, "[t]echnological revolutions do not arrive fully mature."[1] Rather, they take time, unfolding in stages and gathering pace as knowledge, understanding, and use of the technology spreads. If this is true

D. Beyleveld (✉)
Durham Law School, Durham University, Durham, United Kingdom
e-mail: deryck.beyleveld@durham.ac.uk

[1] James H. Moor, "Why We Need Better Ethics for Emerging Technologies" in Jeroen van den Hoven and John Weckert (eds), *Information Technology and Moral Philosophy* (Cambridge: Cambridge University Press, 2008) 26, at 27.

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_12,

of technological revolutions, it is also true of, so to speak, subrevolutions—that is, of technological revolutions that are embedded in another larger technological revolution—of the kind associated with, for example (Moor's own examples), mobile phone technology and the Web (in each case, representing a subrevolution within the larger revolution in computer technology)[2]; and, no doubt, much the same might be said about the development of ultra-fast telecommunication technologies. Crucially, as Moor puts it, the futures of such technologies, "like the futures of small children, are difficult to predict."[3] In this context, and especially so in relation to modern technologies (information, bio, nano, and neurotechnologies) that have enormous manipulative, transformative and disruptive potential,[4] it is entirely understandable that regulators should be urged to adopt a precautionary approach—indeed, such exhortation is entirely understandable whether precaution is understood broadly as the taking of steps to assess, manage or reduce risk or, more narrowly, as the avoidance of risk-taking in the face of uncertainty about the existence of conceivable risk.

Whilst a precautionary approach in a broad sense commands widespread support, the so-called "precautionary principle", after enjoying something of a honeymoon period in the hands of environmental lawyers, has become a target for sustained criticism. The principle, as formulated by the Nuffield Council on Bioethics, holds that regulators may "impose restrictions on otherwise legitimate commercial activities, if there is a risk, even if not yet a scientifically demonstrated risk, of environmental damage."[5] However, the precautionary principle is formulated in many different ways[6]—indeed, as its critics would have it, in far too many different ways[7]—such that many now doubt that it offers a sound basis for regulation.

The purpose of this paper is not to rescue the precautionary principle as such. Along with its critics, we accept that the most controversial uses of precautionary reasoning are those that conclude that, because an activity might possibly have consequences that are catastrophic, the activity should not be engaged in at all. While

---

[2] Ibid., at 31–32.

[3] Ibid at 27.

[4] Compare Mathias Klang, *Disruptive Technology* (Göteborg: Göteborg University, 2006).

[5] Nuffield Council on Bioethics, *Genetically Modified Crops: The Ethical and Social Issues* (London, May 1999) at 162. For a more specific elaboration, see *Pfizer* [2002] ECR II-3305, at para 143: "a preventive measure cannot properly be based on a purely hypothetical approach to risk, founded on mere conjecture which has not been scientifically verified." So, mere conjecture and hypothesis will not suffice. But, a precautionary measure may apply where the risk "has not yet been fully demonstrated" (para 146). The underlying science must be consistent with principles of "excellence, transparency and independence" (para 172).

[6] For helpful recent overviews of the EC jurisprudence, see Veerle Heyvaert, "Guidance Without Constraint: Assessing the Impact of the Precautionary Principle on the European Community's Chemicals Policy" (2006) 6 *The Yearbook of European Environmental Law* 27, esp. 29–37, and "Facing the Consequences of the Precautionary Principle in European Community Law" (2006) 31 *European Law Review* 185.

[7] For a recent assault on the principle, see, e.g., Gary E. Marchant and Douglas J. Sylvester, "Transnational Models for Regulation of Nanotechnology" (2006) 34 *Journal of Law, Medicine and Ethics* 714.

general application of such a principle undoubtedly has absurd consequences, we will argue, using Pascal's Wager[8] as a starting point, that there are conditions under which such reasoning is valid, which we will state by formulating a general principle for the limiting case. We will then examine the possible application of this limiting principle to the well known principle that those accused of criminal offences should not be convicted unless found guilty beyond all reasonable doubt. We will argue that the principle does not apply straightforwardly in this case. However, an alternative principle, which employs precautionary reasoning in a proportionate manner is suggested by this case. In its application, this principle calls for some difficult judgments concerning both the relative undesirability of the options with which one is presented and the proportionality of the precautionary response that is taken. Nevertheless, we will argue that the limiting principle does at least provide a powerful argument against the death penalty.

If our reflections to this point are sound then they suggest that precautionary reasoning is involved in all instances where the law places the burden of proof more strongly on one party, or where the threshold in relation to a particular option is raised—for example, after the (Butler-Sloss) Cleveland Inquiry, the regulatory response was to make it more difficult for child protection units to justify removing a child (for its own safety) from the family home.[9] It also seems to be the case—and we are mindful of Mark Twain's warning that, once we find a hammer, everything starts to look like a nail—that precaution is much more widely implicated in legal reasoning than we perhaps might have appreciated. Is it not the case, for example that precaution, if not in so many words, is also implicated in slippery slope (and, possibly, floodgates) arguments?[10]

The paper is in five principal parts. First, we sketch the reasons why the precautionary principle is open to a multitude of interpretations as well as being unbalanced in its application. If a principle of precaution could be expressed in a much tighter way and if the lack of balance could be overcome, then we might have a principle that regulators should employ. Secondly, we examine Pascal's Wager as a possible model for precautionary reasoning from which we derive a suggested principle of precaution. Thirdly, we examine the application of this principle to the paradigmatic presumption of innocence, the so-called "golden thread" of criminal justice systems, and argue that it requires modification if it is to be widely applicable.[11] Fourthly, we consider the relationship between our precautionary model and slippery slope arguments. And, finally, by focusing on the recent *Biotech Products* (GM crops)[12] dispute at the WTO, we uncover a further complication about the

---

[8] See Blaise Pascal, *Pensées,* trans A., J. Krailsheimer (London, Penguin 1966) p. 151.

[9] The standard is set by the " significant harm" test in s. 31(2) Children Act, 1989.

[10] See Shaun D. Pattinson, *Medical Law and Ethics* (London: Sweet and Maxwell, 2006) pp.15–17.

[11] *Woolmington v DPP* [1935] AC 462, 480 (per Viscount Sankey).

[12] *European Communities—Measures Affecting the Approval and Marketing of Biotech Products* (WT/DS291/23 (United States), WT/DS292/17 (Canada), and WT/DS293/17 (Argentina), August 8, 2003), decided by the WTO in Autumn 2006.

relationship between precautionary policies, scientific uncertainty, and cultural and moral differences.

Our conclusions are twofold. Our first conclusion is that any model of precautionary reasoning, even if not open to the objections ranged against the precautionary principle, will prove controversial in its application because, in the final analysis, we are trying to make moral judgments that are contested, not only as a matter of moral principle, but also in relation to the evidence on which they draw. Our second conclusion is that it is worthwhile to undertake a systematic analysis of the role that precautionary reasoning plays (and should play) in law generally.[13] Even if precaution is not quite everywhere in the law, it is a pervasive fact of regulatory life. To this extent, this paper is the beginning of a prospectus for a full-scale inquiry into the role of precaution and proportionality in the law, the challenge being to identify the basis on which regulators can make good use, not abuse, of precautionary policies.

12.2 The Precautionary Principle and the Standard Critique

If environmental sustainability, biodiversity, and a green ecosphere are all that concern us, then we will probably embrace the precautionary principle. Any activity that might be damaging to the environment should simply not be pursued. Period. However, once the precautionary principle is offered up for general application, its weaknesses are all too apparent. In particular, it invites any number of interpretations and it seems to ignore the sacrifice that is made for the sake of precaution.

The invitation to read the precautionary principle in more than one way arises because its particular formulation hinges on the meaning that we give to the variables that are implicated in the articulation of the principle. Characteristically, we find the following four key variables:

(i) a degree of scientific uncertainty;
(ii) concerning some class, kind, or type of hazard;
(iii) where the damage or hazard is of a certain degree of severity/probability;
(iv) as a result of which some steps to avoid having to risk the hazard are advocated.

---

[13] Nicolaas de Sadeleer has written an excellent treatise on the role and function of, amongst other principles, "the precautionary principle" in law in his *Environmental Principles: From Political Slogans to Legal Rules* (Oxford: Oxford University Press, 2002). The approach taken by de Sadeleer differs from our own in focussing on more or less explicit references to the precautionary principle in, for the most part, though not exclusively, environmental law from an historical perspective, rather than trying to develop a purely *a priori* analysis of features that would render a precautionary approach rational, which can then be used to identify uses of precautionary reasoning and their rationality in any area of law whether or not this is explicitly recognised. De Sadeleer's approach leads him to an important thesis about the function that a shift to a precautionary approach has in maintaining rationality in a system of "post-modern" law. Such an ideological focus is not necessarily incompatible with our own, but it is beyond the scope of this paper to consider his analysis in the light of our own.

We can speak briefly to each of these variables.

The degree of scientific uncertainty

If we use the language of lawyers, we can say that the scientific uncertainty is such that we cannot be sure *beyond all reasonable doubt* (or perhaps, even, all possible doubt) that risk of the relevant hazard either does or does not exist. It follows that those who argue for precaution must maintain either (a) that the evidence of relevant hazard being risked is made out on *a balance of probabilities* or (b) that the evidence falls short of that standard. Where advocates of precaution can make out the case on a balance of probabilities, so that the risk is more likely than not, then even if this does not yet qualify as a scientifically demonstrated risk, it is still a plausible basis for precaution. Characteristically, though, the argument for precaution does not meet this threshold and advocates are making out their case from a much less compelling platform—either from the very weak position that it has not yet been shown beyond all reasonable doubt that there is no risk, or that there is some evidence of risk albeit short of the balance of probabilities threshold. However, whether the case for precaution assumes that the evidence lies above or below the balance of probabilities threshold, there is scope for endless argument about just how strong the evidence needs to be before precaution kicks in.

The class, kind, or type of hazard or risk

In relation to the second variable, whereas in some communities, especially in scientific communities, precaution tends to be focused on risks to health and environment, in other communities, precaution is thought to be appropriately applied additionally to a broad class of economic, social, and cultural risks.[14]

The degree or character of the perceived hazard

Then, there are questions about the degree or character of the damage that is risked. With regard to severity, in relation to environmental hazards, it is often stipulated that the damage threatened must be serious and irreversible,[15] as well as large-scale.[16] But, the first and the third of these stipulations, in particular, leave plenty of room for interpretation. Moreover, these leeways in the interpretation of the precautionary principle carry over if it is applied to non-environmental risks—for example, to the risks associated with stem cell-based therapies or germ-line gene therapy. As regards probability, the issues concern what degree of probability associated with what severity of hazard justifies risk avoidance.

---

[14] Compare, e.g., Ronald Sandler and W.D. Kay, "The National Nanotechnology Initiative and the Social Good" (2006) 34 *Journal of Law, Medicine and Ethics*, 675, at 679.

[15] As in the famous Principle 15 of the Rio Declaration (June 1992) which requires parties to "take precautionary measures to anticipate, prevent or minimize the causes of climate change and mitigate its adverse effects." In this context, the Declaration continues: "Where there are threats of serious or irreversible damage, lack of full scientific certainty should not be used as a reason for postponing such measures."

[16] See *EU Communication on the Precautionary Principle* (IP/OO.96) (February 2000).

The measure of precaution to be taken

Finally, once the principle is engaged, it remains to specify the required response. If there is a relevant risk, how is it to be tackled? Does precaution require outright prohibition or cessation, or is it sufficient to limit or adjust the activity or simply to monitor it?[17] Again, the opportunities for finessing the principle are endless.[18]

However, even if the variables within the precautionary principle could be stabilised, we should not forget the potential perversity of precaution, surely nowhere better exemplified than in the price paid by Samuel Butler's fictitious Erewhonians when they decided to destroy their machines for fear that they would supplant humans.[19] Modernising this old lesson, Cass Sunstein has subjected the precautionary principle to a damning critique. Sunstein's point is that the taking of precautionary measures—or, at any rate, the taking of precautionary measures that involve giving up some activity—itself involves risk (sacrifice) and that this must be brought into what is otherwise a one-sided narrow screen calculation.[20]

To appreciate the force of this point, we need only reflect on the choice between a safety-first (ex ante) regulatory strategy that delays the market authorisation for what might prove to be life-saving medical products and a strategy that facilitates bringing products to market subject to ex post safety monitoring. More generally, as Adam Burgess concludes in his study of precautionary responses to the perceived hazards presented by mobile phones and cell phone towers, there is a worrying tendency to create risk shadows and then commit large resources to chasing them—thus, "In the elusive quest to establish a risk-free existence, our autonomy, intelligence, and capacity for change and enlightenment stand in danger of being compromised and diminished."[21]

We conclude, therefore, that, with its present bias and deficiencies, the precautionary principle has no future. However, it would be silly to abandon a precautionary approach in the broad sense; and it would be a missed opportunity if, in our eagerness to ditch the precautionary principle, we overlooked a model of precautionary reasoning that regulators could, and should, adopt.

---

[17] For an excellent discussion, see Richard B. Stewart, "Environmental Regulatory Decision Making Under Uncertainty" in Timothy Swanson (ed), *An Introduction to the Law and Economics of Environmental Policy: Issues in Institutional Design* (Amsterdam: Elsevier Science, 2002) 71.

[18] Compare Council for Science and Technology, *Nanosciences and Nanotechnologies: A Review of Government's Progress on its Policy Commitments* (London, March 2007) para 52 et seq (for the view that, while precaution does not entail a moratorium on the development of nanotechnology, the basic research should not be neglected). For similar emphasis on the importance of conducting the basic research, see the French National Consultative Ethics Committee for Health and Life Sciences, Opinion No 96 ("Ethical Issues Raised by Nanosciences, Nanotechnologies and Health", 2007) p. 10.

[19] Samual Butler, *Erewhon* (London: Penguin Books, 1985: first published 1872).

[20] Cass Sunstein, *Laws of Fear* (Cambridge: Cambridge University Press, 2005).

[21] Adam Burgess, *Cellular Phones, Public Fears, and a Culture of Precaution* (Cambridge: Cambridge University Press, 2004) 281. But, the elusive quest continues: see, e.g., Geoffrey Lean, "Wi-Fi: the Backlash" *The Independent on Sunday*, July 15, 2007, p. 14 (Haringey council resolving to adopt a precautionary approach to wi-fi technology in schools).

12.3 Pascal's Wager: A Model for Precautionary Reasoning?

One of the most famous arguments that rely on precautionary reasoning is contained in what is known as Pascal's Wager.[22] If God is conceived of as an omnipotent being (as is the case in the Bible), then it is impossible to know whether or not God exists. However, (according to Pascal), the biblical God requires us to believe in Him and live the life prescribed in the Bible or suffer eternal damnation. So, reasons Pascal, you can believe that God exists or not believe that God exists. But if you are wrong in not believing that God exists you lose everything, while if you are wrong in believing that God exists, you lose nothing. Ergo, you rationally must believe that God exists.

There are fatal problems with this argument. One of these is that belief (as against conforming behaviour) is not something that can be chosen. At best, the argument can provide good reason for acting in accordance with God's requirements for action. However, even so limited, there is another problem, which is that the existence or non-existence of the biblical God is not the only issue of fact that is relevant to the fear of eternal damnation that drives the argument. If it is unknowable whether or not the biblical God exists, it is equally unknowable whether or not a different "God" exists, whose requirements regarding belief/action might conflict with those of the God of the Bible. In other words, the omnipotent being that exists might damn you for believing in the biblical God. So a parallel argument constructed in relation to the requirements of such an alternative "God" would require non-belief in the biblical God/non-conformity with the requirements set by the biblical God.

The problem here is that the hazard to be avoided depends on the existence of the biblical God, and precaution, driven by the severity of that hazard, is directed to make it rational to presume the existence of the biblical God. So, if we disentangle these things, the question is whether we can find a valid application of the idea, central to Pascal's Wager, that actions are to be avoided simply because they *might possibly* threaten wholly unacceptable outcomes. We think that we can.

Some moral philosophers (e.g., Immanuel Kant[23] and Alan Gewirth[24]) claim to have demonstrated that there is a moral principle that is categorically binding on all agents. In other words, reason categorically requires agents to act in accordance with this principle and not to violate it. In the case of Kant, this principle is the Categorical Imperative, which in one of its formulations requires all agents to treat all other agents never solely as means but always at the same time as ends in themselves. In the case of Gewirth, this principle is the Principle of Generic Consistency, which requires all agents to act in accordance with a structure of so-called generic rights possessed by all agents. Suppose then that someone "A" agrees that either

---

[22] See Blaise Pascal, *Pensées,* trans A., J. Krailsheimer (London, Penguin 1966) p. 151.

[23] See *Groundwork of the Metaphyisics of Morals* (1785) translated with a commentary by H. J. Paton under the title *The Moral Law* (London Hutchinson 1948).

[24] See *Reason and Morality,* (Chicago: Chicago University Press 1978).

Kant or Gewirth (it does not matter which) is correct on this point. It is still possible for A to try to evade the practical consequences of this acceptance by denying that there are any agents other than A. This is because to be an agent is to possess certain mental and emotional capacities. While A knows directly that A has these capacities, A cannot know that any other being has the same capacities. At best A can know that other beings behave as though they have these capacities. Nevertheless it is possible that beings that act as though they have these capacities do not. Hence it is possible, as far as A can know, that all other beings that act as though they are agents are not agents, only apparent agents.[25]

In this example the only issue relevant to how A must act is whether or not B is an agent. If B is an agent, then A must treat B in accordance with the principle that A believes to be categorically binding otherwise the principle will be violated. If B is not an agent then A need not so treat B. Although A cannot know whether or not B is an agent, to part paraphrase Pascal, "If A wagers that B is an agent and loses then he loses little; but if he wagers that B is not an agent and loses, then he loses everything. Hence, A must wager that B is an agent" The only proviso is that A must be able to treat B as an agent, and this A will be able to do if (as we have supposed) B acts as though B is an agent.

We submit that this is sound reasoning. While the precautionary reasoning here is formally identical to that involved in Pascal's Wager, the structure of its application is different in that the requirement that sets up the value to be served (the avoidance of violating the rights of an agent/ of failing to treat an agent as an end in itself) does not depend on B's existence as an agent. In Pascal's Wager, uncertainty about the existence of the biblical God carries with it uncertainty about the need to obey the prescriptions attributed to this God, hence uncertainty about whether the hazard to be avoided actually exists. On the other hand, when arguing, as above, that the Categorical Imperative/the PGC requires agents to treat apparent agents as agents, the hazard to be avoided is given by the Categorical Imperative/the PGC, which is axiomatic for the precautionary argument and not a conclusion derived from it. Of course, if Kant's/Gewirth's arguments for a categorical imperative are not sound[26] *and* belief that there is a categorical imperative is merely a matter of subjective commitment, then the conclusion of precautionary reasoning guided by the values enshrined in the categorical imperative will not be something that those who do not accept that there is a categorical imperative will be required to accept. Commitment to a categorical imperative will be no more rational than fear of eternal damnation by the biblical God. Consequently, the conclusion of the precautionary

---

[25] For more on this precautionary argument see Deryck Beyleveld and Roger Brownsword *Human Dignity in Bioethics and Biolaw* (Oxford: Oxford University Press 2001) pp. 119–134.

[26] It is no secret that we consider Gewirth's argument to be sound. For a defence of Gewirth's argument see, in particular, Deryck Beyleveld, *The Dialectical Necessity of Morality: An Analysis and Defense of Alan Gewirth's Argument to the Principle of Generic Consistency*. (Chicago: Chicago University Press 1991). For an analysis of the relationship between Gewirth's and Kant's arguments, see Deryck Beyleveld and Roger Brownsword, *Human Dignity in Bioethics and Biolaw* (Oxford: Oxford University Press 2001) pp. 86–110.

argument will be no more rationally compelled than commitment to the value that guides it. Nevertheless, the precautionary argument guided by commitment to a categorical imperative is not viciously circular in the way that Pascal's Wager itself appears to be.

Be this as it may, the general principle involved in the above argument may be stated as follows.

> P1: If doing X (e.g., not treating B as an agent) risks something that is categorically prohibited (not granting the generic rights to an agent) whereas not doing X (treating B as an agent) risks something not categorically prohibited, then doing X must be prohibited even if it cannot be known whether or not the risk is real (whether or not B is an agent, thus, whether or not B has generic rights to violate), provided only that not doing X is possible

P1 has special features. In particular, it concerns prohibition of speculative risk on the basis of a hazard being categorically prohibited. Were the hazard to be less than categorical either prohibition could not be justified/or the risk would have to be less than speculative.

We will later comment on some of these issues, which are key to the question about the legitimacy of wider applications of precautionary reasoning in a narrow sense. But first it is worth seeing if P1 itself has any application in legal reasoning.

12.4 Innocent Until Proven Guilty and Precaution

It is a general principle of criminal law that persons accused of crimes are not to be convicted unless proven to be guilty beyond a reasonable doubt. Being guilty on the balance of the evidence is not enough to secure a conviction. Any reasonable doubt about guilt must yield an acquittal.

Can this policy be justified by P1? The reason for this policy is clearly to avoid convicting an innocent person of a criminal offence. The policy, however, increases the risk that a guilty person will be acquitted. Now, for P1 to apply, convicting an innocent person must be something that is categorically prohibited, whilst acquitting a guilty person must be something not so heinous as to be categorically prohibited. If we suppose this to be the case, however, and whether or not we consider it to be the case is clearly a value-judgment, P1 does not seem to apply because in such a case it is surely necessary that persons convicted of criminal offences be shown to be guilty beyond any possible doubt. It is *possible* to acquit persons accused of crimes (whether they are guilty or not). If convicting them when they are innocent is categorically prohibited when it is possible not to convict them, they must be acquitted unless it is certain that they are guilty.

These observations, however, suggests another principle

> P2: If doing X risks (or constitutes) something that is undesirable (convicting an innocent person) whereas not doing X (acquitting a guilty person) risks (or constitutes) something less undesirable, then doing X must be prohibited when there is a degree of suspicion that X might be done, even if it cannot be known whether or not the risk is real (whether or not X is certainly guilty), provided only that not doing X is possible

P2 is, however, too vague to instantiate the reasoning involved in the requirement to acquit those not found guilty beyond a reasonable doubt. This is because the "beyond a reasonable doubt" condition prescribes a specific level of suspicion that X might be done. Furthermore, it is surely reasonable that this level of suspicion must be proportional to the degree of undesirability of doing X as compared to the undesirability of not doing X. This suggests the more precise principle.

> P2*: If doing X risks (or constitutes) something that is undesirable (convicting an innocent person) whereas not doing X (acquitting a guilty person) risks (or constitutes) something less undesirable, then doing X must be prohibited when there is a degree of suspicion that X might be done that is proportionate to the undesirability of doing X in relation to the undesirability of not doing X, even if it cannot be known whether or not the risk is real (whether or not X is certainly guilty), provided only that not doing X is possible

The requirement to acquit those not found guilty beyond a reasonable doubt is justified by P2* if it is judged that the requirement of guilt beyond a reasonable doubt is proportionate to the importance (relative to that of not acquitting a guilty person) of not convicting an innocent one.

If P1 does not apply to the requirement to acquit those not found guilty beyond a reasonable doubt, then as we have already indicated this is only because the prohibition on convicting the innocent is not (at least universally) categorically prohibited and/or P1 applies prohibition regardless of the degree of evidence beyond mere possibility for the prohibited outcome occurring. Those who attach a deontological value to not convicting the innocent might well argue that this is categorically prohibited; but, if they do, they should require acquittal on the mere possibility of innocence. On the other hand, if they are of a more utilitarian persuasion, they might hold that whether or not conviction of the innocent is to be categorically prohibited depends on the consequences of conviction. Thus they might contend that where conviction requires the death penalty to be imposed conviction of an innocent person is more serious than when conviction carries a "lesser" penalty, and they might be prepared to accept that execution of an innocent person is categorically prohibited. If so, P1 applies and requires that the death penalty should not be applied to those convicted of crimes unless they are shown to be guilty beyond a possible doubt.

If it is, furthermore, argued, that guilt in relation to a crime always involves an element of *mens rea,* then it surely follows that there must always be a possibility of innocence. How can we possibly know with certainty another person's intentions? Indeed, as work in the new brain sciences casts doubt on the extent to which defendants are "in control",[27] how can we confidently maintain the penal character of criminal justice? If we entertain any such doubts, the death penalty should never be imposed. While there are many ifs here, the general form of this reasoning is surely both clear and sound.

When we review the other end of the criminal justice system, we find a different tilt to the precautionary approach and a different placing of the burden. So long as

---

[27] See, e.g., Patricia Smith Churchland, "Moral Decision-Making and the Brain" in Judy Illes (ed), *Neuroethics* (Oxford: Oxford University Press, 2006) p. 3.

Smith is merely a suspect or a defendant in the criminal process, as we have said, the burden is on the prosecution. However, once Smith is convicted, everything changes. Crucially, even when Smith has served a tariff custodial punishment, we cannot assume that he will be released back into the community. For, if Smith is one of those persons now indefinitely imprisoned (for the sake of public protection) under Section 225 of the Criminal Justice Act 2003, the burden lies on Smith to persuade the relevant parties that it is safe to release him. Is such an approach defensible under the precautionary model that we have outlined? On the one hand, by doing X (detaining Smith indefinitely) we seriously interfere with Smith's freedom and do him wrong. It is perfectly possible to release Smith; but we do not do so if we believe that there is some risk that Smith might violate the legitimate interests of third-parties. On the face of it, this looks like an abuse of precaution and a disproportionate burden on Smith.

12.5 The Precautionary Model and Slippery Slopes

Our analysis suggests that precautionary reasoning is involved whenever the law places the burden of proof more heavily on one side of a case than on the other. This is because the central consideration that drives precautionary reasoning is that, given uncertainty about the right answer to a question, it is judged to be more serious to err in one direction than in another.

It also seems to us that precautionary reasoning might be involved in the acceptance of slippery slope arguments. Such arguments are characteristically concerned with holding a regulatory line. For example, a judge might refuse to accept an incremental extension of liability for fear that it will, as a matter of principle or practice or both, weaken the power of regulatory control. In its most pointed application, the slippery slope argument is directed against activities that are conceded to be harmless in themselves on the grounds that acceptance of these activities will, or might, lead to acceptance of harmful activities. So, for example, in the current debate about the use of cytoplasmic hybrid embryos as research tools,[28] we might detect two versions of the slippery slope argument. One version does not take a position on whether the use of cytoplasmic hybrid embryos would be harmless; the view is that the current regulatory line should be held because, quite simply, it is a line that regulators can plausibly stand on. By contrast, the other version accepts that the use of cytoplasmic hybrid embryos would be harmless but worries that, having authorised the use of such hybrids, it would then be difficult to draw a distinction between one kind of hybrid and another, or between hybrids and chimeras.

---

[28] House of Commons Science and Technology Select Committee, *Government Proposals for the Regulation of Hybrid and Chimera Embryos* (Fifth Report of Session 2006–2007) (HC 272-I, 5 April, 2007); the Academy of Medical Sciences, *Inter-Species Embryos* (London, July, 2007); and, most recently, House of Lords and House of Commons Joint Committee on the Human Tissue and Embryos (Draft) Bill, *Human Tissue and Embryos (Draft) Bill*, HL Paper 169-I, HC Paper 630-I (London: The Stationery Office, August 1, 2007).

Such arguments frequently rely on assumptions that require empirical evidence. However, to obtain such evidence requires the intrinsically harmless activities that are held to threaten the intrinsically harmful ones to be performed. At this point, if the slippery slope argument is to work, precautionary reasoning kicks in. If it is not validly deployed then the slippery slope argument will itself be fallacious.

Consider the much-debated cases of physician assisted suicide (PAS) and active euthanasia. The standard reason given by rights-committed legal regimes for their refusal to recognise the *public* lawfulness of assisted suicide or active euthanasia is that such recognition would potentially undermine the right to life of third parties, particularly the rights of vulnerable third-party agents. So, for example, in *Washington v Glucksburg*,[29] we find Chief Justice Rehnquist, delivering the opinion of the Court, saying:

> [t]he State has an interest in protecting vulnerable groups—including the poor, the elderly, and disabled persons—from abuse, neglect, and mistakes. The Court of Appeals dismissed the State's concern that disadvantaged persons might be pressured into physician assisted suicide as "ludicrous on its face."... We have recognized, however, the real risk of subtle coercion and undue influence in end of life situations ... Similarly, the New York Task Force warned that "[l]egalizing physician assisted suicide would pose profound risks to many individuals who are ill and vulnerable ... The risk of harm is greatest for the many individuals in our society whose autonomy and well being are already compromised by poverty, lack of access to good medical care, advanced age, or membership in a stigmatised social group."... If physician assisted suicide were permitted, many might resort to it to spare their families the substantial financial burden of end of life health care costs.
>
> The State's interest here goes beyond protecting the vulnerable from coercion; it extends to protecting disabled and terminally ill people from prejudice, negative and inaccurate stereotypes, and "societal indifference."...The State's assisted suicide ban reflects and reinforces its policy that the lives of terminally ill, disabled, and elderly people must be no less valued than the lives of the young and healthy, and that a seriously disabled person's suicidal impulses should be interpreted and treated the same way as anyone else's...
>
> Finally, the State may fear that permitting assisted suicide will start it down the path to voluntary and perhaps even involuntary euthanasia.[30]

The Chief Justice employs a belt and braces approach to precaution: first, precaution for the sake of vulnerable persons; and, then, precaution in order to prevent a slide towards euthanasia. Yet, how convincing, on either score (belt or braces), is this reasoning?

Notoriously, the problem with the former expression of precaution is that we simply do not know whether the rights of vulnerable agents would be compromised;[31]

---

[29] 117 S.Ct. 2258; 138 L Ed 2d 772.

[30] 117 S.Ct. 2258; 138 L Ed 2d 772, 795–796.

[31] The comparative evidence is very difficult to interpret. Compare Ronald Dworkin, *Sovereign Virtue* (Cambridge, Mass.: Harvard University Press, 2000) Chapter 14, esp. at 470–472, who asks whether we can be confident that such empirical accounts as we have are (i) ethically clean (i.e., untainted by background ideological bias), (ii) methodologically sound, and (iii) applicable from one jurisdiction to another. For extended discussion of both the methodological and the comparability reservations, see Penney Lewis, "The Empirical Slippery Slope from Voluntary to Non-Voluntary Euthanasia" (2007) 35 *Journal of Law, Medicine and Ethics* 197.

and, so long as we refuse to relax the legal restrictions on PAS, we are not going to find out. In other words, precaution is applied to minimise the risk to third-parties but in such a way that we are not able to ascertain whether the risk that we fear is a real one. Similarly, a restrictive approach to PAS prevents any slide towards active euthanasia (with or without consent); but, in the same way, precaution obstructs our knowing whether the risk of a slide is a real one. If those who argue against such blanket criminal restrictions (rather than a regulated procedure for PAS of the kind that has been advocated over the years) were to do so on the basis that the precautionary approach implicated in the policy is applied disproportionately, they would seem to have a point. On the face of it, precaution is abused when it is translated into blanket prohibitions against PAS.

12.6 Precaution and Trade

For a number of years, the European Union operated with a controversial de facto moratorium against the approval of GM crops. Matters came to a head in *Biotech Products*, where the legality of the EU's policy was challenged by the US, Canada and Argentina.[32]

Where, as in *Biotech Products*, the science relating to the safety of GM crops is contested, how is the matter to be resolved? An innocent response is that the question should be determined by reference to the view supported by "sound science", this being taken to be a neutral and reliable arbiter. However, for many commentators on the practice and politics of science, this is a naïve view.[33] Science just is not like that. Scientists reasonably disagree with one another, not just about the bottom-line questions, but about matters of methodology, relevance, and focus, and so on. Science is never going to be theory-neutral (that is the whole point of the enterprise) but some deny, too, that it is "value-neutral".

The case of GM crops is almost a textbook example of such non-neutrality:

> The "products approach" to regulating GMOs assumes that no untoward risk occurs merely from applying this technology to agricultural production. GMOs are subjected to strict rules only when the end products are not substantially equivalent to their conventional counterparts. In contrast, the "process approach" rests on the idea that genetic engineering itself may entail novel and unique risks to human health or the environment. Whereas the United States has embraced the products approach to GM agriculture, the European Union and its member states have tended to adopt the more precautionary process approach.[34]

So, if scientists on one side of the Atlantic make safety judgments by reference to the end product while scientists on the other side of the Atlantic make (different) safety judgments by reference to the process used, and if both practices are regarded

---

[32] Note 12 above.

[33] See, e.g., David Winickoff, Sheila Jasanoff, Lawrence Busch, Robin Grove-White, and Brian Wynne, "Adjudicating the GM Food Wars: Science, Risk, and Democracy in World Trade Law" (2005) 30 *Yale Journal of International Law* 81.

[34] Ibid., at 87

in their own territories as sound science, then "sound science" simply cannot serve as a neutral court of appeal.

In the event, the WTO Disputes Panel in *Biotech Products* made no attempt to settle the question of whether GM crops are safe. The question was not whether the EU position was scientifically vindicated, nor even whether it was a reasonable position as judged by common opinion. Rather, the question was the narrower and more specific one of whether the EU position was consistent with Members' obligations under the Sanitary and Phytosanitary (SPS) Agreement. In particular, the question was whether the EU was entitled to invoke the precautionary position implicit in Article 5.7. This Article provides:

> In cases where relevant scientific evidence is insufficient, a Member may provisionally adopt sanitary or phytosanitary measures on the basis of available pertinent information, including that from relevant international organizations as well as from sanitary or phytosanitary measures applied by other Members. In such circumstances, Members shall seek to obtain the additional information necessary for a more objective assessment of risk and review the sanitary or phytosanitary measures accordingly within a reasonable period.

In other words, under conditions of scientific uncertainty, where the risks are unclear, Members may exercise precaution by derogating from the usual market access principles.

In favour of the EU, it was decided that the de facto moratorium on the approval of biotech products was a legitimate response to the uncertain state of the scientific evidence—in other words, the moratorium was a legitimate provisional and precautionary measure as contemplated by Article 5.7 of the SPS Agreement. However, in favour of the complainants, it was found that the EU was in breach of its obligations by failing to progress individual approval procedures without undue delay and that it was vicariously in breach of its obligations in respect of the safeguard measures taken by the six member states that represent the strongest opposition to GM crops in Europe.[35] Bearing in mind that the Commission had itself defended its own internal market principles by refusing to permit Austria (one of the group of six) to establish a "GMO-free area", one imagines that, at least in some quarters of Brussels, the latter part of the WTO ruling was neither unexpected nor altogether unwelcome.[36]

Would the EU's approach pass muster relative to the precautionary model that we have outlined? The dilemma is whether to do X (open one's markets to GM crops) or not do X (deny market access to GM crops). The argument against the former option is that GM crops, once planted out and commercially exploited, might present a serious risk to human health and the environment. It is possible to avoid this risk by not doing X. The cost of taking this latter option is that it impedes international trade and, possibly, causes some economic hardship to exporting Members.

---

[35] Annex C1(a) of the SPS Agreement requires Members to undertake and complete "without undue delay" procedures for checking and ensuring the fulfilment of sanitary or phytosanitary measures.

[36] Compare Sara Poli, "Restrictions on the Cultivation of Genetically Modified Organisms: Issues of EC Law" on Han Somsen (ed), *The Regulatory Challenge of Biotechnology* (Cheltenham: Edward Elgar, 2007) p. 156.

On balance, the precautionary EU response, as found by the WTO, does not look disproportionate. However, there are two hidden complications.

First, if Members are permitted to rely on a state of scientific uncertainty in order to close their markets to allegedly risky products or services, they might abuse the privilege. There is a risk, in other words, that precaution might be used as a pretext for trade protectionism. If the real reasons behind EU precaution were financial, it would not do, in an international free trade agreement, to prioritise EU financial interests over those of the Americas. Still, this is not a major problem. It means only that the WTO needs to take a hard look at precautionary positions taken up by Members and not simply defer to them.

The second complication is more serious as well as being much more important for our purposes. This is that European resistance to GM crops does not rest purely and simply on uncertain hazard to health, safety and the environment. There is a view that GM crop manipulation is wrong, that it is incompatible with dignity. Here we have an ethic of veneration,[37] following which GM is categorically off limits.[38] This factor, in turn, gives rise to two complications.

One complication is to do with the transparency and honesty of the precautionary policy. The real reasons for market closure need to be declared. If the real reason why Austria and others want GM-free zones is cultural rather than scientific, then this needs to be brought into the open. Irrespective of whether the precautionary argument is a sound one, not to declare it for what it really is certainly is an abuse of precaution. Again, a hard look on the part of decision-makers is probably the best that we can ask for.

The other complication is more profound. This is that a model of precautionary reasoning, as an exercise in the legitimate regulation of risk, necessarily builds on certain values. In the end, there is no neutral application of precaution because precaution is always framed in a value-laden way; precautionary reasoning is necessarily driven by value judgments. If the EU puts its moral cards face up on the table, there is no simple response. For a tribunal to say that European culture and local values cannot be allowed into a precautionary calculation is not to maintain a separation of precaution and value; rather, it is to say that one value-laden paradigm of precaution is to be preferred to another.

---

[37] Jean-Christophe Galloux, Arne Thing Mortensen, Suzanne de Cheveigné, Agnes Allansdottir, Aigli Chatjouli, and George Sakellaris, "The Institutions of Bioethics" in Martin W. Bauer and George Gaskell (eds), *Biotechnology—The Making of a Global Controversy* (Cambridge: Cambridge University Press, 2002), where three ethical frames are identified—one organised around the principle of utility, a second around the principle of democracy, and a third around the principle of veneration.

[38] Compare Nuffield Council on Bioethics, *Genetically Modified Crops: The Ethical and Social Issues* (London, May 1999) paras 1.32–1.50 (discussing the natural/unnatural boundary, taboos, and moral conservatism).

12.7 Conclusion

The frequency with which burden of proof and slippery slope considerations appear in reasoning, policies, and principles, argues for a general examination of the role that precautionary reasoning plays in law. Detailed work will need to be done before we can be in any position to offer a confident comprehensive analysis. However, our analysis in this paper does enable us to draw a number of lessons. First, the application of precautionary reasoning involves value judgments about the desirability/undesirability of doing something X vs the desirability/undesirability of not doing X. Second, except where doing X is judged to be categorically undesirable, this will not be a straightforward or easy matter. Third, when the value attributed to not doing X is not categorical, then questions of the degree/standard of evidence for the relevant risks being instantiated becomes relevant. While evidence must clearly be proportionate to the severity of the hazards involved, such calculations raise difficult issues concerning the commensurability of the variables involved, which, in our opinion, make judgments of this kind not susceptible to non-arbitrary quantification and perhaps not even to non-arbitrary qualification.[39]

The importance of the fact that precautionary reasoning is driven by value judgments must not be underestimated. Precautionary reasoning has been used to justify calls for genetic modification of crops to be banned on the grounds that we do not know that this will not have disastrous consequences to human health or the environment; and, even though the European Group on Ethics in Science and New Technologies has emphasised that precautionary policies are not zero-risk policies, we can be sure that similar objections will be made against nanotechnology.[40] When it is responded that there is no evidence that this is the case, the counter-response is that there is no evidence that there is no proof that this is not the case. To this it will be pointed out that such reasoning could justify a ban on virtually everything, which if the matter can be ended here, appears to reduce the precautionary argument to absurdity. However, the matter cannot be ended here. This is because there is almost certainly the hidden assumption in the precautionary case that genetic modification is an unnecessary activity, one that does not need to be engaged in (because to the extent that it offers meaningful benefits, other activities deemed more desirable for some reason can reap these benefits instead). If a risk, however, speculative, is unnecessary in this sense, then it is not irrational to eschew it on the grounds of its mere possibility. But then it is clear that the issue is at root an evaluative one. This, of course, means that precautionary arguments will be inherently controversial, but if we are right then the controversiality does not lie so much in the form of the arguments but in the details of their substance.

---

[39] See further, Deryck Beyleveld and Roger Brownsword, *Human Dignity in Bioethics and Biolaw* (Oxford: Oxford University Press 2001) pp. 255–258.

[40] European Group on Ethics in Science and New Technologies to the European Commission, *Opinion on the Ethical Aspects of Nanomedicine* (Opinion No 21) (17 January, 2007).

Chapter 13
Ethics of Technology at the Frontier of Uncertainty: A Gewirthian Perspective

Paul Sollie

Abstract Complexity and uncertainty are central to technological innovations. Many emerging complex technologies are developed under conditions of uncertainty. This brings to the fore the question of how we should deal with risky and uncertain technology developments that are potentially detrimental and harmful to human beings and the environment. It is argued that uncertainty gives rise to three fundamental questions with regard to ethical theory, namely its justification, its practicability, and, as a sub-question, how it deals with complex, uncertain cases that thwart practical moral reasoning due to a resultant lack of knowledge. Any ethics of technology should be able to formulate answers to these questions if it is to be regarded as adequate. This article purports to provide an answer to these questions by investigating Gewirth's supreme moral principle—the Principle of Generic Consistency (PGC)—that requires every agent to act in accordance with its own and its recipients' generic rights to freedom and well-being.

Keywords Uncertainty · Ethical technology assessment · Gewirth · Principle of generic consistency · Procedural turn · Substantive and procedural ethics

13.1 Introduction

Complexity and uncertainty are central to technology development. Many new complex technologies are developed under conditions of uncertainty. How should we deal with risky and uncertain technology developments that are potentially detrimental and harmful to human beings and the environment? In Chapter 10 it was argued that uncertainty gives rise to three fundamental questions with regard to ethical theory, namely its justification, its application, and, as a sub-question, how it deals with complex, uncertain cases that thwart practical moral reasoning due to a resultant lack of knowledge. Any ethics of technology should be able to formulate answers to these questions if it is to be regarded as adequate. This chapter

P. Sollie (✉)
Ethics Institute, Utrecht University, Utrecht, The Netherlands
e-mail: paulsollie@gmail.com

P. Sollie, M. Düwell (eds.), *Evaluating New Technologies*, The International Library of Ethics, Law and Technology 3, DOI 10.1007/978-90-481-2229-5_13,

is concerned with providing an answer by investigating Gewirth's supreme moral principle—the Principle of Generic Consistency (PGC)—that requires every agent to act in accordance with its own and its recipients' generic rights to freedom and well-being.

In Section 13.2 I briefly recap the minimal requirements for any adequate future ethics of technology that is able to deal with uncertainty. Section 13.3 is concerned with the first requirement, namely finding and justifying the PGC as the supreme moral principle in Gewirthian ethics. Section 13.4 then turns to the second requirement, i.e. applying the PGC to moral problems. Section 13.5 is an intermezzo and elaborates on the difference between ideal-typical and complex cases. The latter issue relates to the third requirement, namely that any ethical theory should be able to deal with complex cases, which is the concern of Section 13.6. Section 13.7 portraits the consequences of the PGC for technology development and its ethical assessment. Finally, Section 13.8 summarises the article.

13.2 Requirements for an Ethics of Technology

Uncertainty is not a pivotal but a rather underappreciated concept in past and contemporary ethical debates. In Chapter 10 it was however argued that uncertainty plays a fundamental role in complex new technology developments. Walker's typology of uncertainty (Walker et al., 2003), which was originally introduced in policy decision-making on e.g. environmental issues, proved to be an insightful contribution to the field of ethics technology. By introducing this typology to the field of ethics and technology it was possible to map the dimensions of uncertainty that surround and underlie complex technology developments. These dimensions first of all include uncertainty with regard to the subject of evaluation. Besides the question whether this subject is an individual or collective (which was not addressed), the subject is constrained by the fact that it is not omniscient, but lacks information about future states of the world. It does not know how technological applications will materialise and what consequences they will have. The latter observation points at the uncertainty of the object of evaluation, namely the envisaged new technology. The uncertainty with regard to the object and subject of evaluation not only relates to a lack of information, so-called epistemological uncertainty, but also to processes that are inherently uncertain and which cannot be reduced by more or better information. This ontological uncertainty is portrayed in, for instance, the randomness of natural processes or the unpredictability of human behaviour. A third aspect of uncertainty concerns the framework and model that are used to generate answers to complex questions. The uncertainty that comes with frameworks pertains, for instance, to the variables that are used and not used—the *ceteris paribus* clause—or the borders that are set—what aspects of reality are part of the model and what aspects are left out?

These dimensions of uncertainty that are attached to complex technology developments bear directly on the practical aim of ethics to guide correct moral reasoning.

How are we to ethically evaluate new technologies if these are accompanied by uncertainty? The analysis of uncertainty in Chapter 10 resulted in the formulation of three questions that any adequate ethics of technology needs to address in order to be adequate. The questions are:

1. Theoretical aim of ethics: What right-making criteria underlie ethical theory and how are these right-making criteria justified?
2. Practical aim of ethics: How do these right-making criteria guide practical moral reasoning?
3. How does ethical theory deal with complexity and uncertainty, that is, how does it construe a justified relation between a substantive and procedural approach that is able to legitimately deal with uncertainty in a proactive manner?

In the following sections each of the demands and requirements that stem from these questions is investigated from the perspective of Gewirthian ethics. The main focus will be on requirement 3 by arguing that Gewirth and Beyleveld and Brownsword offer a justified construal of the relation between substance and procedure that is required when making legitimate decisions under uncertainty. Hence, this paper argues that Gewirthian ethics is a promising candidate because it has a sound story with regard to the three requirements.

13.3 Requirement 1: Gewirth's Project for a Supreme Moral Principle—Justifying the PGC

One of the focal questions in ethics is whether a supreme moral principle can be justified. A positive answer to this question seems moot, because of a widely held and recognised moral pluralism and a disagreement on the proper justification of such a principle. However, in order to be able to make universal moral claims and judgments, a supreme moral principle is required. In his 1978 *Reason and Morality* Alan Gewirth takes up the fundamental ethical question whether there can be a rational justification for a supreme moral principle. He propounds that such a principle should be grounded in the nature and structure of human agency. The supreme principle he argues for is the Principle of Generic Consistency (PGC): 'Act in accord with the generic rights of your recipients as well as of yourself.' (Gewirth, 1978, 135) Gewirth lucidly summarises the basic premises of his theory as follows:

> . . . an agent is a person who initiates or controls his behavior through his unforced, informed choice with a view to achieving various purposes; since he wants to fulfill his purposes he regards his freedom and well-being, the necessary conditions of his pursuit of purposes, as necessary goods; hence he holds that he has rights to freedom and well-being; to avoid self-contradiction he must hold that he has these generic rights insofar as he is a prospective purposive agent; hence he must admit that all prospective purposive agents have the generic rights; hence he must acknowledge that he ought at least to refrain from interfering with his recipients' freedom and well-being, so that he ought to act in accord with their generic rights as well as his own. (1978, 171)

With regard to the first requirement forwarded in section one, a more detailed portrayal of his argument for the PGC is required. I start off with discussing the method of argumentation and the concept of generic rights before sketching the general the line of argumentation.

Gewirth's argument for the PGC advances by means of a dialectically necessary method. The arguments provided are inferences relative to the agent and the justification of these arguments lies within agents themselves—the method is reflexive. The advantage of the dialectically necessary method is that it does not run the risk of the Münchhausen trilemma that states that grounding a supreme moral principle will finally result in an infinite regress (*regressus ad infinitum*), a dogmatic decision, or a vicious circle (*petitio principii*). The Gewirthian argument proceeds from human agency or, more specifically, from the claim 'I am an agent.' The sequential steps in the argument are logical necessary inferences from the first premise. The Münchhausen trilemma is avoided, because the PGC is argued for neither deductively nor inductively but reflexively (Gewirth, 1978, 7–22). The dialectical method starts off from assumptions, opinions, ideas, or claims of agents about how they view things—agents reason from their own internal perspective and scrutinise the logical entailments of initial premises (Gewirth, 1978, 43). This is to be contrasted with assertoric methods that include assertions about things and their properties independently from the agent's subjective perspective. To illustrate, when I dialectically state that Haruki Murakami's *The Wind-up Bird Chronicle* is a superb novel, this is an assertion about this book relative to my person. It is a statement about the relation between the book and me and about how I regard and assess its quality. When I assertorically argue that *The Wind-up Bird Chronicle* is an outstanding piece of literature, this is a statement about the quality of the book independent from me, because I appeal to objective properties. The method Gewirth employs is not only dialectical, but also necessary. In opposition to a contingent method, which leaves open to agents the initial statements from which to reason, the PGC is secured by a necessary method. A method is considered necessary as from the first premise on an agent cannot deny on pain of contradiction the logical conclusions that follow. To sum up, 'the dialectically necessary method propounds the contents of this relativity as necessary ones, since the statements it presents reflect judgments all agents necessarily make on basis of what is necessarily involved in their actions.' (Gewirth, 1978, 44)

The PGC holds that every agent is required to act in accordance with its own as well as other agents' generic rights to freedom and well-being or contradict that it is an agent. The concept of generic rights requires further explanation. According to Deryck Beyleveld and Roger Brownsword, 'needs are generic if they are prerequisites of an ability to act at all *or* with any *general* chances of success, *regardless of the purposes being pursued*.' (2001, 70) Generic needs relate to the degree of needfulness for action. In a subtle and nuanced manner Gewirth distinguishes hierarchically according to the criterion of degree of needfulness between basic needs, non-subtractive needs, and additive needs. The most fundamental needs are basic needs, which are required for the very possibility of acting at all, such as life, food, shelter, or health. For instance, interferences with a basic need, such as the health of

an agent, can take place on various levels and may subsequently affect the ability to act at all in a variety of ways. There is a difference in interference with basic needs between actions that results in the loss of sight or an arm and those that result in the loss of life. Non-subtractive and additive needs are connected with the successfulness of actions. Non-subtractive needs are needed for successful actions and, more specifically, with the capacity for fulfilling purposive agency, such as accurate information, but are not as basic needs prerequisites for acting at all. Additive needs deal with the needs to improve one's capacities for successful action, such as acquiring new skills (Beyleveld and Brownsword, 2001, 70).

Now we can turn to the argumentation for the PGC. I cannot portray the complete argument here in all its technical details. (For this see Gewirth, 1978, 129–150; Gewirth, 1984, 1–24; Beyleveld, 1991, 13–46; Beyleveld and Brownsword, 2001, 69–86) The argumentation for the PGC consists of three different stages of inferences that should dialectically necessary lead from the claim that 'I am an agent' to the PGC or the claim that any agent ought to act in accordance with its own and other people's generic rights to agency. The starting point for reasoning is the nature and structure of human action, which is conceived of as voluntary purposive behaviour.

Stage one commences with the claim that if I am an agent I need to accept that I do (or have the intention to do) action *A* to achieve purpose *P* (whatever *P* might be) that I have chosen myself. Because *P* is an objective that I have chosen freely and willingly, I therefore value *P* proactively, which motivates me to pursue *P*. Next, there are generic conditions of agency that are prerequisites for the ability to act or the successfulness of acting. Irrespective of what *P* might be, I need the conditions of freedom and well-being in order to be an agent and to act freely to pursue my purpose *P*. Since these are prerequisites for my being a successful agent, I must hold that having these generic conditions are necessary for me to pursue as a means to *P*. 'My having the generic needs is categorically instrumentally good for me' (Beyleveld and Brownsword, 2001, 73). As I value *P* proactively I ought to pursue that I have the generic needs of freedom and well-being. This ought is directed at me as an agent.

The second stage is concerned with the fact that others ought not to interfere with my generic conditions. 'Ought implies can' is invoked to secure that '[i]t cannot be the case that I ought to do *A* if I cannot do *A*.' (Beyleveld, 1991, 40) In order to achieve my freely chosen purpose *P* I need to do *A*. Since the generic needs are prerequisites for *me* to achieve *P*, other agents ought not to interfere with my freedom and well-being *from my point of view, namely my interest in P*. Note that the statements so far are still prudential, i.e. from my internal perspective. Consequently, this requirement of non-interference is, although it is directed at other agents, not other-direct*ing*. From this it is inferred that I must consider that I have the generic rights of agency. Hence, from the initial claim, i.e. 'I am an agent,' it is argued that I must consider that I have the generic rights of agency from *my* perspective.

The third stage is concerned with the logical inference from the fact 'I am an agent entails I must consider that I have the generic rights' to the conclusion that we must grant all agents the generic rights to agency. Gewirth uses the 'Argument

from the Sufficiency of Agency' (ASA) and the principle of universalizability to secure this inference. The reason why I, as a purposive agent, must consider that I have the rights is the fact that I am a purposive agent, which is a sufficient condition for having to consider the generic rights. 'The sufficient reason for which he claims the rights must adduce simply the characteristic of being a prospective purposive agent, so that he must admit that all other prospective purposive agents also have these rights.' (Gewirth, 1978, 119) This in the end secures the PGC, namely that every agent needs to act in accordance with his or her as well as other agents' generic rights to freedom and well-being. If Gewirth's argument and justification is sound, and I believe it to be, then the first adequacy-requirement of ethical theory is satisfied.[1]

13.4 Requirement 2: Practical Moral Reasoning—Applying the PGC

The second requirement entails that right-making criteria should be able to guide correct moral reasoning. Normative ethical theories not only propound right-making criteria that can be held justifiably—the first requirement—, these also need to guide us in dealing with ethical dilemmas and conflicts by demonstrating *how* these apply. Hence, the PGC needs to be authoritative in how to deal with cases in order to be feasible as an ethical theory. The requirement of guidance in practical moral reasoning seems self-evident and a fairly easy and straightforward task. In a sense this is true, but it often turns out to be more difficult to give guidance in correct moral reasoning than it might seem at first sight since there are a number of caveats that should be acknowledged. Before turning to those caveats in Section 13.5 and to more complex moral problems in Section 13.6, this section will briefly discuss the second requirement in relation to ideal-typical, linear-causal cases. Can the PGC give guidance to our correct moral reasoning? Consider the following well-known moral problem of 'the murderer at the door' that was central in a debate between Immanuel Kant en Benjamin Constant. Suppose you are hiding someone in your basement to protect him from a murderer. At some point this would-be murderer knocks on your door and asks you whether you are hiding this person. Should you

[1] It is interesting to note that, although Gewirth has given a powerful and, to my mind, sound justification of the PGC, his theory has hardly received the attention it deserves. Self-evidently his theory has evoked much debate, which ensued after the launch of *Reason and Morality* in 1978. Notwithstanding the fact that a substantial number of scholars have tried to question and attack seemingly controversial aspects of Gewirth's theory, surprisingly within the wider community of ethical theorising Gewirth, and later Beyleveld (1991), Steigleder (1999), and Spence (2006), have largely been ignored. With regard to the criticisms directed at Gewirth's argumentation, I refer to critical essays of, for instance, Singer and Hare that are collected by Regis (1984). However, the best source to explore a critical perspective on Gewirthian ethics is Beyleveld (1991). This is an as nearly as possible *exhaustive* portrayal of the criticisms of Gewirth's ethical theory, in which Beyleveld has gathered, discussed, and refuted most, if not all criticisms of Gewirth's theory that have appeared through 1990.

tell a would-be murderer the truth when he asks so, namely that the intended innocent victim is in your house? What does the PGC require you to do?

According to the PGC agents act morally right if they act in accordance with the generic rights of the recipients of their acts as well as of their own. The PGC requires that we not only respect and refrain from interfering with the capacities for the generic rights of freedom and well-being of the recipients without their consent but also respect their having the capacities. (Gewirth, 1978, 208) The generic needs (basis, non-subtractive, and additive) are necessary conditions for agents to participate in a transaction with other agents, but are also conditions that must be respected in the transaction. For instance, the basic good of 'life' is a both a precondition for being able at all to engage in a transaction, but it is at the same time a basic need of the agent that is to be respected and refrained from with interfering against the agent's consent.

In the case of the murderer at the door, the agent enters a transaction with both the intended victim and the would-be murderer. With regard to both the recipients the PGC requires that the agent refrains from interfering with their freedom and well-being. By telling the truth to the murderer the agent would open up the possibility of an interference with the victim's life, which is a basic good. Moreover, the agent is obligated to assist the intended victim in securing this basic good of life when the victim cannot do so on his own effort and wishes so. However, does the agent not violate the rights of the murderer by not telling the truth? Since lying to the murderer is an interference with his generic rights, the agent is obligated by the PGC not to lie and consequently to tell the truth about the whereabouts of the intended victim. This problems poses a dilemma, because both lying and killing are violation of the PGC. So, *how* is the PGC able to adjudicate in this dilemma? According to the criterion of degrees of necessity for action, Gewirth hierarchically discriminates between basic, non-subtractive, and additive goods (Gewirth, 1978, 343–344). Basic goods trump non-subtractive goods, which trump additive goods. Hence, in the case of 'the murderer at the door,' the basic good of life prevails over the right to the possession of accurate information and not to be lied to in your purpose fulfilment that are non-subtractive goods. Hence, the agent is required by the PGC not to tell the truth about the whereabouts of his intended victim (Gewirth discusses a similar case, 1978, 351–352).

This brief section aimed to show the feasibility of a direct application of the PGC to ideal-typical cases. There are, nonetheless, fundamental differences between ideal-typical cases and more complex, real life cases that frustrate the direct application of the PGC. The next section focuses on the difference between the direct and indirect or ideal-typical and real life cases. The PGC, however, offers a legitimate way of *also* dealing with complex cases of uncertainty.

13.5 Intermezzo: Shift from Direct to Indirect, from Ideal-Typical to Real Life Cases

Problems like 'the murderer at the door' are linear-causal. Such ideal-typical cases generally consist of rational, acting, and evaluating agents that are capable of

overlooking, calculating, and evaluating from a normative framework a limited subset of future states of the world of which one has substantial information regarding the range of possible actions and their consequences. (See Sollie, 2007, 298) These conditions allow for a direct application of the PGC. In this type of moral problems, it is possible to determine the situation, investigate who is involved, who is affected in what way by reference to the generic needs for agency (a relatively substantial harm to additive goods may be preferred to a smaller harm to basic goods), and what are the effects of one's actions, etc. This is a rather uncomplicated task. First, the situation is scrutinised and assessed, then the PGC is applied, and, finally, the PGC yields an answer of how to morally go about in this specific moral problem. This is the way it is often expressed that ethical principles result in definitive answers to moral problems. However, the direct application of the PGC (or, in fact, any principle) does not hold for complex cases under uncertainty such as complex technology developments. Complex cases cannot be solved by applying the PGC directly to the problem, but require an indirect approach (Section 13.6). These complex cases differ from or exceed ideal-typical cases for a variety of reasons of which I will mention a few.

First, it should be noted that these ideal-typical cases are often introduced to explain and illustrate specific ethical theoretical standpoints. They are purposely construed for this goal, for which reason they are rather simple and non-linear of nature. Beyond this more practical reason of showing the efficacy and feasibility of ethical theories, there are other differences. Second, they differ in context and boundaries. Ideal-typical cases are context-free and discussed under *ceteris paribus* assumptions. These cases are isolated or closed cases, because all that matters towards the ethical assessment is the variables that are provided within the case descriptions. No more and no less is assumed; what you see (or read) is what you get. For example, in 'the murderer at the door' we are able to determine who is involved, what aspects are problematic, what are possible future states and what are the ethical implications with regard to the intended action. Real life situations, to the contrary, are hardly ever isolated cases and are, furthermore, characterised by vague and fluid boundaries—they should always be conceived of in context. There are ample factors of the situation and the context that need to be considered in the ethical assessment. To mention a few; all the relevant aspects and variables, unforeseen, and unanticipated consequences for future generations, the randomness of natural processes, the dynamics of technology development (e.g. its pace and the involvement of a large number of agents), the change in scale of application and impact (local technologies that may have global impact), and the unpredictability of human behaviour. These are all factors that should be acknowledged. In practice, however, we often end up—because we need to in order to be able to have at least a minimal grasp of the problem—reducing complexity by framing and developing models that describe complex situations in terms and processes that provide us a means to understanding these situations. A third aspect is moral pluralism. People need not necessarily be convinced by the PGC and its justification. As a matter of fact, the PGC is subject to debate and controversy (See Beyleveld, 1991). What is more, even if all people would agree on and accept the PGC, then still there

will be complex cases in which there is a conflict of duties and consequently a debate on how the PGC applies and what conclusions it yields (See, Gewirth, 1978, 338–350). Fourth, complex cases frequently involve collective agents. The collective agents of complex cases relate to the body of agents that is called upon to make a decision, to the people that are concerned with performing the action, or to the number of stakeholders involved. For instance, with regard to collective decision-making, a group of people is asked to decide on a certain issue, such as in voting in the parliament or in participatory technology assessment. In many such cases it is difficult to determine who counts towards being part of the collective and how the collective decision is to be legitimised to those not involved in the collective decision-making (See, e.g., Skorupinski and Ott, 2002; Grunwald, 2004). The collective nature of complex situations further differentiates it from theoretical cases that normally involve a limited and finite number of agents. Usually a single agent is called upon to make a decision, which involves but a few stakeholders affected by the action. In the imaginary case 'the murderer at the door' the agent has to make a decision between two possible actions that involve two recipients, the intended victim and the would-be murderer. Fifth, decisions whether or not to pursue particular new technology developments are often made under uncertainty, which is also triggered by the previous issues. Uncertainty relates to both epistemological features, such as the collection of relevant information (in new, complex technologies we often lack info about application and consequences) and to ontological features, like the nature of human behaviour and social and environmental processes.

If these observations are sound, one might argue that complex, uncertain cases frustrate the practical aim of ethics in such a way that a direct application of *any* substantive supreme moral principle is not feasible. As argued in Chapter 10, not only pure substantive, but also pure procedural approaches are inadequate. Pure substantive approaches do not suffice under conditions of complexity and uncertainty since we require information to apply their right-making criteria to derive determinate answers with regard to the object of assessment. This information is, however, lacking due to uncertainty. Pure procedural ethics, on the contrary, can only say something about the structure of debate and not about its substance. This type of theory is not able to make moral judgments by guidance of moral principles based on the substance of the subject matter. This is a major deficit in procedural approaches and therefore I contended that we should strive for a substantive theory that is able to justifiably include a procedural approach. As a consequence, what should we expect from a substantive ethical theory when it comes to complex, uncertain cases? The paper argues that any ethics that purports to have a say on complex and uncertain ethical problems also necessitates a concomitant procedural approach. The next section purports to demonstrate that the PGC does not fall prey to the inadequacy of a direct practical application, but entails such a required procedural turn by definition, because it justifiably accounts for both direct and indirect applications.

13.6 Requirement 3: Incorporating Procedure in Substance—the Procedural Turn

In complex or indeterminate cases the PGC cannot be applied directly and consequently this may lead to a disagreement about how agents should act morally. Beyleveld and Brownsword discuss three reasons why people can disagree about what the PGC requires despite its dialectical necessity (2006, 147–148). First, there are cases in which the PGC is indifferent to which solution is to be chosen. Consider, for example, the history of football and rugby union. It is believed to be William Webb Ellis who invented rugby in 1823 by taking the ball in his hands while playing football. From that moment on football and rugby became different sports, either allowing the use of hands or not. The PGC does not entail a preference for either football (no-hands except for the goalkeeper) or rugby (hands), but the PGC will not allow football players to use their hands while playing football, because this would be cheating and pervert the game of football. Second, there are cases in which the generic conditions of different agents are in conflict. In such cases it is not only the difference in generic conditions, such as the importance or the degree, but also the probability of each of the effects on the generic conditions that are to be considered. Often these conflicts are incommensurable as it is impossible to balance condition C^1 of agents A^1 with condition C^2 of agents A^2 and, subsequently, the PGC does not *directly* yield an answer. How should we compare the probability of 0.01 of having A^1's basic good of life diminished by e.g. attracting an illness for a while with the probability of 0.8 of violating A^2's non-subtractive good of not being lied to? These types of conflict are conflicts under risk. Third, there are also situations in which we cannot assign probabilities because of uncertainty. This class of situations refers to those cases which are so complex, i.e. uncertain, that the application of the PGC is rationally moot.

However, the fact that people can disagree about what the PGC requires and that the PGC does not entail direct solutions to complex problems, does not mean that the PGC does not require that we resolve such problems and that the PGC has no story about how to deal with them. As applying the PGC provides no direct answer, we must seek an *indirect* answer. With regard to this, Beyleveld and Brownsword propose a procedural turn, which entails that 'there must be a turn to solutions to the substance of the dispute at hand that are produced by procedures justified by the PGC rather than solutions required by particular substantive interpretations of the PGC.' (2006, 149) To be sound, it must be shown how this procedure is justified from within the PGC and how it guides decision-making in complex and uncertain cases. For this I first turn to Gewirth's argument for the method of consent and then show what Beyleveld and Brownsword's procedural turn, which builds on the method of consent, involves.

Gewirth distinguishes between a direct application (1978, 199–271) and indirect application (1978, 272–365) of the PGC. The former is mainly directed at individual agents and their actions. The latter is concerned with more complex social interactions, social institutions and laws. Through a procedure of consensual and

social legitimisation (1978, 304ff) the PGC allows for a set of social rules and laws that imposes requirements on individuals and their actions. The requirements of the PGC are imposed on social rules, institutions, and laws that serve to protect people's freedom and well-being. For example, if the referee gives a football player a red card, he has to leave the pitch. This coerced sending away of the player is not a violation of his right to freedom since the player has freely accepted the rules of the game. (Cfr. Gewirth, 1979, 1163) These rules and laws must be in accordance with the PGC, for which reason the PGC is still the supreme moral principle. Hence, in complex situations the PGC takes a *detour*, so to speak, by imposing requirements on individuals not directly, but indirectly via rules and laws that are consensually and legitimately justified by it. These rules and laws are not free-standing requirements, but are to support a context of order that serves to protect the generic rights of all agents. So as to prevent harm being done by the state or institutions to the generic rights of agency, the PGC imposes limits on the state. There are certain actions (called civil liberties) that are outside the scope of the state, in order to prevent the state to conflict with the PGC, which requires that each agent is free to engage in any action or endeavour voluntarily and to his unforced choice as long as it does not interfere with, i.e. harm or coerce, other agents. Having and employing these civil liberties constitutes what Gewirth coins the method of consent. The method is itself directly justified by the PGC:

> The *PGC* requires that all persons have equal rights to freedom, and . . . these include equal rights to civil liberties and to participation in the political process. Because of the vital importance of this process for well-being, these liberties are especially important among freedom rights upheld by the *PGC*. [. . .] Since the method of consent consists in institutional arrangements for implementing and exercising these equal rights, it is itself morally justified as a fair procedure. It therefore imposes within its specific sphere the general obligations imposed by the *PGC*. The obligations include not only obedience to the minimal state with its criminal law, but also acceptance of the results of the method of consent when the method of consent is applied to determine officials who shall give effect to the criminal law as well as other officials, laws, and policies of the state. (Gewirth, 1978, 309–310)

Let us return to the procedural turn, which is a specification of the method of consent, and see what it involves and how it guides practical reasoning in complex and uncertain cases. It should be noted that there is not one procedure that is prescribed by the PGC as there are many procedures that might be invoked, as long as they are fair with regard to the nature of the dispute and the complexity of the situation (Beyleveld and Brownsword, 2006, 149). We will generally not consent to a procedure of flipping a coin or drawing straws in electing the next president or in awarding research grants. Different situations may require different procedures. The situation of awarding research grants, for instance, asks for experts who consider and assess research proposals and propound a solution that is evidence-based, balanced, reviewable, transparent, and the like. Alternatively, when choosing a new government in democratic societies, this is carried out by democratic elections and not via expert decision-making. Beyleveld and Brownsword identify four steps in the procedural turn that are to secure the arriving at legitimate, binding outcomes in complex cases.

First, as stated before, the PGC sometime requires that some conflicts cannot be left unsettled and need to be resolved (Beyleveld and Brownsword, 2006, 149). Second, since leaving some conflicts unresolved is acting contrary to the PGC, a decision has to be made in favour of one of the possible solutions to the conflict. If according to the PGC there is only one process of determination applicable to the situation, then this procedure needs to be brought into play. This process will deliver a binding outcome. However, if it turns out to be the case that more than one process of determination is applicable then we enter the third step (Beyleveld and Brownsword, 2006, 150). Consider for example a recent conflict in the Netherlands about whether or not it should approve of a new European law for the European Union. More then one process of determination was applicable in this context, which included, among others, the option that the decision was left to the government or directly to Dutch citizens via a referendum. If all parties involved in the conflict agree on settling this dispute by calling a referendum, then the outcome of this process of determination is binding on all agents.

If the parties in a dispute cannot agree on either one of the available processes, we arrive at the fourth stage. The dispute is no longer solely about the subject matter, but also about the process of determination. Beyleveld and Brownsword are aware of the fact that this procedural turn can turn into an infinite regress. An infinite regress would undermine the aim of guiding practical moral reasoning and, moreover, result in a violation of the PGC, because it leaves a dispute unsettled that is not to be left unsettled. To avoid the infinite regress, they maintain that 'the procedural turn that best serves the PGC is one in which the turn is one to which agents consent. Failing this, at some point, the PGC must call time and agents will be required to resolve their substantive differences. If they do so by consent, they are bound under a procedural justification.' (Beyleveld and Brownsword, 2006, 150–151) It often happens that the parties involved in a debate cannot decide or agree on the procedure to invoke to resolve the issue at hand. It might be argued that we are left with an impasse. However, the quest for finding consensual procedures is not infinite, but is terminated as the PGC requires that we resolve certain issues as they cannot be left unsettled. This termination is not random, but is derived from the PGC and the consent of agents. Even in such intractable conflicts people are committed to the PGC, which requires them to solve the complex issue. Accordingly, they recognize that they will violate the PGC by not settling the dispute, that they have to avoid a deadlock, and that they need to seek compromise on the procedure to follow. The parties will compromise and consent to a procedure if refraining from consent would violate the PGC. Finally, consent to a selected procedure, which is legitimised by the PGC, is the justificatory reason why the outcome is binding on all parties involved.[2]

[2] So far no claim is made about what procedure to invoke when confronted with intractable problems. Without going into detail of *what* procedures we may employ, which is an investigation outside the scope of this paper, some minimal requirements and questions can be formulated that should be observed when invoking a procedure. Without presenting an exhaustive list, there is at least the question of which criteria the procedure should satisfy. Above all, the procedure

In retrospect, when disputes need to be settled and cannot be resolved by applying the PGC directly and straightforwardly, the PGC demands that we take a procedural detour—we do however not succumb to pure proceduralism—that is to guide our moral reasoning by finding the appropriate, consensual, and justified process of decision-making that will result in binding outcomes for all agents. In the last section I will portrait some considerations on technology development that follow from the PGC.

13.7 Considerations for Technology Development

Based on the previous consideration, the PGC is adequate as an ethics of technology under conditions of uncertainty as it gives guidance in resolving complex issues. I will conclude with a few considerations on using the PGC for ethical technology assessment. First, complex technology developments are still fraught with uncertainty and the PGC is not a magical supreme moral principle that provides all the information we need for the ethical assessment. This is a problem for all ethical theories and hence the focus should be on how a theory can justifiably deal with it and not on resolving the fact of uncertainty—at least not on those aspects of uncertainty that are intractable. Gewirthian ethics is such a justified approach.

Second, the fact that we deal with uncertainty might lead to a lack of interest—'we don't know, so we'll find out along the way'—or even negligence of ethical issues of technology developments. Some theories of ethics of technology therefore focus on coping with ethical issues once they have been identified (Moor, 1999). All that retrospective approaches can do is often no more than invoking policy measures to contain the problem and to try to alter problematic technologies. The latter solution is easier said than done, because regularly problematic issues come to the fore only once a technology is established and entrenched in society, which makes it both costly and difficult to change (See, e.g. Collingridge, 1980, 19). Contrary to such theories, the PGC requires that we resolve certain conflicts rather than leaving them unsettled and hereby it prohibits inertia with regard to the ethical assessment of technology development. People are morally obligated to attend to issues concerning technology development, such as energy consumption, waste products, health issues, or privacy. The PGC is not indifferent to such matters since it requires that we should act in accordance with the generic rights of agency of all agents.

Third, the procedural turn provided by the PGC is not an ad-hoc solution in absence of better strategies. The attention to ethical issues of uncertain technology development is not detached from the supreme moral principle. The PGC offers a

---

should be in accordance with what the PGC requires or allows. Procedures that run contrary to the PGC should be avoided. What is more, with regard to choosing a procedure and the actual decision-making, there is the question of who should participate and how they should participate. For instance, one of the major debates concerns whether or not lay-people should be included in decision-making and how they should be included. (See, e.g., Grunwald, 2004; Pontzen, 2006; Skorupinski and Ott, 2002).

legitimate strategy for dealing with complex situations by providing the procedural turn as an '*add-on*.' Neither does complexity require pure proceduralism, nor does it give way to inactivity, since the PGC provides us with a legitimate and justified way of attending to ethical issues of technology development. Whereas uncertainty, once again, frustrates our ability to make a direct, substantive assessment, the procedural detour allows for invoking fair procedures (whatever these procedures are people consent to) to arrive at a legitimate decision that is binding for all.

Fourth, the PGC can accommodate for other approaches in its practical moral reasoning. The PGC can benefit from other approaches that deal with uncertainty and complexity. The PGC may employ other strategies that purport to deal with uncertainty, such as moral imagination, precautionary reasoning (see e.g. Chapter 11 by Steve Clarke or Chapter 12 by Deryck Beyleveld and Roger Brownsword), or scenario analyses (see e.g. Chapter 8 by Nicole Karafyllis or Chapter 9 by Swierstra et al.). For example, via the construction of both positive and negative scenarios, scenario analyses of technology developments may lead to better information of possible future applications and consequences, even if we cannot assign any probability to these scenarios. The results of such strategies can be insightful to applying the PGC to specific complex cases.

13.8 Conclusion

Every ethical theory should be able to deal with three fundamental questions, namely how it is justified, how it guides correct moral reasoning, and how it deals with complexity and uncertainty. This article argued that the PGC is a candidate for meeting all three requirements. First, it provides a sound and convincing justification for the PGC. Second, the PGC also guides practical moral reasoning in both ideal-typical and complex, uncertain cases. With regard to the latter a procedural turn is offered, which is justified by the PGC itself, to arrive at legitimate answers in complex situations, although we should at the same time keep in mind that parts of uncertainty cannot be reduced.

Acknowledgment I would like to thank both Marcus Düwell and Deryck Beyleveld as well as the two anonymous reviewers from Springer for their critical comments and suggestions that helped to improve this article.

References

Beyleveld, D. *The Dialectical Necessity of Morality. An Analysis and Defense of Alen Gewirth's Argument for the Principle of Generic Consistency*. Chicago: The University of Chicago Press, 1991.

Beyleveld, D. and Brownsword, R. *Human Dignity in Bioethics and Biolaw*. Oxford: Oxford University Press, 2001.

Beyleveld, D. and Brownsword, R. "Principle, Proceduralism, and Precaution in a Community of Rights." *Ratio Juris* 2006, 19, 2, 141–168.

Collingridge, D. *The Social Control of Technology*. New York: St. Martin's Press, 1980.
Gewirth, A. *Reason and Morality*. Chicago: The University of Chicago Press, 1978.
Gewirth, A. "The Basis and Content of Human Rights" *Georgia Law Review* 1979, 13, 4, 1143–1170.
Gewirth, A. "The Justificatory Argument for Human Rights" *Social Philosophy & Policy* 1984, 1, 2, 1–24.
Grunwald, A. "Participation as a Means of Enhancing the Legitimacy of Decisions on Technology? A Sceptical Analysis." *Poiesis & Praxis* 2004, 3, 1, 106–122.
Moor, J. "Just Consequentialism and Computing." *Ethics and Information Technology* 1999, 1, 65–69.
Pontzen, H. "Discourse Ethics in TA Procedures: A Game Theory Model." *Poiesis & Praxis* 2006, 4, 3, 219–230.
Regis, E. Jr. (Ed.). *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*. Chicago: Chicago University Press, 1984.
Skorupinski, B, and Ott, K. "Technology Assessment and Ethics." *Poiesis and Praxis* 2002, 1, 2, 95–122.
Sollie, P. "Ethics, Technology Development and Uncertainty: an outline for any future ethics of technology." *Journal of Information, Communication and Ethics in Society* 2007, 5, 4, 293–306.
Spence, E.H. *Ethics Within Reason. A Neo-Gewirthian Approach*. Oxford: Rowman and Littlefield Publishers, 2006.
Steigleder, K. Grundlegung der normatieve Ethik. Der Ansatz von Alan Gewirth. Freiburg/München: Verlag Karl Alber, 1999.
Walker, W.E., Harremoës, P., Rotmans, J., Sluis, J.P. van der, Asselt, M.B.A. van, Janssen, P. and Krayer Von Kraus, M.P. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-based Decision Support."*Integrated Assessment* 2003, 4, 1, 5–17.