Values and Capabilities

Ethics by Design for Vulnerable People

by Naomi Jacobs

Values and Capabilities
Ethics by Design for Vulnerable People

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Eindhoven,
op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op woensdag 29 september 2021 om 16:00 uur

door Naomi Jacobs

geboren te Amsterdam

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof. dr. I. Heynderickx
1e promotor:	prof. dr.ir. A.W.M. Meijers
2e promotor:	prof. dr. W.A. IJsselsteijn
copromotor:	dr. L.E. Frank
leden:	prof. dr.ir. I.R. van de Poel (TU Delft)
	prof. dr. P. Markopoulos
	prof. dr. V. Wild (Augsburg University)

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

The PhD research is executed at Eindhoven University of Technology, Department of Industrial Engineering & Innovation Sciences.

© Naomi Jacobs, 2021
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without prior permission in writing of the publisher.

editors: Philip Brey, Anthonie Meijers, Sabine Roeser and Marcel Verweij

ISBN: 978-90-386-5350-1
ISSN: 1574-941X

Copies of this publication may be ordered from the 4TU.Centre for Ethics and Technology, info@ethicsandtechnology.eu
For more information, see http://www.ethicsandtechnology.eu

Contents

Preface ix

Acknowledgements xi

1 Introduction 1

1.1. Behavior Change Technology for Health and Wellbeing 1
1.2. Note on Method 4
1.3. Vulnerability 5
1.4. Ethics by Design 7
1.5. Value Sensitive Design 9
1.6. Capability Sensitive Design 13
1.7. A Designer's Perspective 15

2 Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People 17

2.1. Persuasive Technology 19
2.2. Dimensions of Vulnerability 21
2.3. Taxonomy of Vulnerability 21
2.4. Persuasive Ethics from a Vulnerability Perspective 22
2.5. The Concern of Taking into Account Users' Interests 25
2.6. The Concern of Manipulation and Coercion 27
2.7. Concluding Remarks 32

Appendix: The Vulnerabilities of Femtech 33

3 Why Value Sensitive Design Needs Ethical Commitments 39

3.1. The Voice of the Researchers 40
3.2. Universal Values and the Naturalistic Fallacy 40
3.3. Ethical Theory 41
3.4. Desiderata for a Mid-Level Ethical Theory 43
3.5. Concluding Remarks 45

4 Capability Sensitive Design for Health and Wellbeing Technologies 47

4.1. Capability Approach and Capability Theory 49
4.2. Capability Sensitive Design 52
4.3. CSD and Health & Wellbeing Technology 59
4.4. Applying Capability Sensitive Design 60
4.5. Challenges for CSD 73
4.6. Concluding Remarks 78

5 Bridging the Theory-Practice Gap: Design-Experts on Capability Sensitive Design 79

5.1. Methodology 84
5.2. Can We Design for Values? 86
5.3. Can we Apply CSD Flexibly? 87
5.4. How to Give Guidance but Avoid (Too Many) Restrictions? 92
5.5. Discussion 94

6 Concluding Remarks 101

Epilogue 107

References 117

Summary 129

Samenvatting 131

About the Author 133

List of Publications 135

Preface

Each chapter of this dissertation has been published as an article in a peer-reviewed journal.

Chapter 2 'Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People' has been published in *Bioethics* as: Jacobs, N. (2020). Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People. *Bioethics.* Vol. 34. No.5. pp.519-526.

The Appendix 'The Vulnerabilities of Femtech' has been published in *Podium voor Bioethiek* as: Jacobs, N. and Evers, J. (2019). De Kwetsbaarheid van Femtech. *Podium voor Bio-Ethiek*. Vol. 26. No.4. pp.13-15.

Chapter 3 'Why Value Sensitive Design Needs Ethical Commitments' has been published in the *Journal of Ethics and Information Technology* as: Jacobs, N. and Huldtgren, A. (2018). Why Value Sensitive Design Needs Ethical Commitments. *Journal of Ethics and Information Technology*. Vol.23. No.1. pp.23-26.

Chapter 4 'Capability Sensitive Design for Health and Wellbeing Technologies' has been published in *Science and Engineering Ethics* as: Jacobs, N. (2020). Capability Sensitive Design for Health and Wellbeing Technologies. *Science and Engineering Ethics*. Vol. 26. No.6. pp.3363-3391.

Chapter 5 'Bridging the Theory-Practice Gap: Design-Experts on Capability Sensitive Design' has been published in the *International Journal of Technoethics* as: Jacobs, N. and IJsselsteijn, W. (2021). Bridging the Theory-Practice Gap: Design-Experts on Capability Sensitive Design. *International Journal of Technoethics*. Vol.12. No.2. pp.1-16.

Acknowledgements

It is often said that a book is the project of many hands, and that is certainly true for this dissertation. I am very grateful for all the support that I received from my supervisors at Eindhoven University of Technology. I want to thank Anthonie Meijers for his careful attention to my thoughts and ideas throughout the years and for his continuous encouragement to always challenge myself. I am indebted to Lily Frank for all the hours she spent attentively listening to my ideas and to the many questions she posed that were of such great value in bringing my initial ideas to fruition in this dissertation. I am ever appreciative of Wijnand IJsselsteijn's unwavering enthusiasm and encouragement and I greatly enjoyed writing the last chapter of this dissertation together. I want to thank Alina Huldtgren for introducing me to the Value Sensitive Design community at the start of my PhD project and for encouraging me to voice my thoughts on VSD and put them to paper. I am grateful to Els-Kersten van Dijk for all the support and encouragement she has given me at the start of this project and for bringing so much wit and warmth to the work floor.

I want to express my sincere gratitude to the independent members of my doctoral committee, Ibo van de Poel, Verina Wild and Panos Markopoulos. I am deeply grateful for their willingness to engage with my work and participate in my PhD defence.

I am truly privileged by the friendship and support I have received from my paranymphs Marjolein Lanzing and Jenneke Evers who, together with Eveline Groot, are my philosopher queens. Thank you for being such good friends and allies, you truly inspire me.

A heartfelt thank you to all the friends I have met along the way: Iris Loosman for being such a warm and witty colleague all those years, Matthé Scholten for the endless strolls through the park while having great chats that were peppered with invaluable career advice, Nienke de Graeff for the lively conversations on philosophy and life that always leave me energized, Heleen Rutjes for being such a great and helpful colleague at the HTI department, and all the members of the 4TU.Ethics PhD community with whom I got to share ideas during our writing retreats.

I want to thank all of my colleagues at the Philosophy & Ethics group in Eindhoven. A special shout out to Sven Nyholm who has been such a great mentor to me at the 4TU.Ethics PhD writing retreats and who has always been so generous with sharing his feedback and advice.

I am grateful to Jesper Ahlin Marceta for inviting me to KTH Royal Institute of Technology in Stockholm after meeting me at a conference in Norway where we talked about Value Sensitive Design during a fancy conference dinner. At KTH, I met Barbro Fröding who has given me the warmest welcome and has become a friend and a true source of inspiration to me.

A debt of gratitude is also owed to Lisa Doeland and Elize de Mul with whom I had the exciting and enriching experience of writing the philosophy book *Onszelf voorbij, kijken naar wat we liever niet zien*. I have learned so much from the both of you and writing the book together was, funnily enough, the most productive distraction from my PhD project that I could wish for.

Lastly, I want to give an ever warm and heartfelt thank you to my family and friends for their unwavering trust in me. I want to thank Marion, my mother, for always being curious and engaging in conversation about my research as well as any other aspect of life. Gerard, my father, for passing on the joy of writing to me, and Ezra, for being my bright brother. Thank you, Laura and Iris, for being such glorious friends.

Thomas, I want to thank you for everything. You are my joy in life.

1 Introduction

Many know the pleasure of unhealthy habits. Perhaps you enjoy drinking a glass of wine or two, smoking a cigarette, or eating a piece of well-glazed cake late at night. Unfortunately, many of us also know that these enjoyable habits are bad for our health. We are continuously informed of how smoking, overeating, and physical inactivity contribute significantly to the development of various chronic diseases, and how these chronic diseases in turn affect our personal well-being as well as indirectly add to high health care costs (WHO 2015). '*But oh well*', I hear myself- and many others with me- think, '*what could possibly be the harm of having just one more piece of this and just one more glass of that*', right?

1.1. Behavior Change Technology for Health and Wellbeing

To encourage people like me and many others to stick to a healthier lifestyle, a variety of technologies have been developed in recent years that aim at health-related behavior change. Such health-related behavior change technologies (BCTs), or persuasive technologies for health and wellbeing as they are also often called[1], are a class of technologies that are designed to change a person's attitude, behavior, or both in relation to health and wellbeing (IJsselsteijn et al 2006). These technologies are aimed at health promotion or disease management (Orji and Moffatt 2016) and are part of the bigger trend of personalized medicine and healthcare which entails a move away from a *one size fits all* model, to healthcare and medicine applications that are tailored to the individual (Sharon 2017). Two essential aspects form the core of these personalized healthcare applications: they are (1) data-driven and (2) focused on participation of the individual. In other words, the applications require the active participation of individuals as generators and contributors of vast amounts of personal data, which is subsequently used to target, tailor, and personalize care to each individual's unique profile and experience. These two aspects of data-centeredness and participation of the individual lie at the core of health-related BCTs that stimulate users to collect, measure, and display data concerning almost any bodily function, mental state, and behavioral activity (Sharon 2017).

---

[1] I will be using the terms 'behavior change technology' and 'persuasive technology' interchangeably throughout this dissertation.

Examples of health-related BCTs include e.g., Headspace, a meditation and mindfulness app that offers a range of meditation exercises to improve users' mental health and happiness; Fitbit, an activity tracker that aims to stimulate and persuade users to exercise more; HAPIfork, a smart fork that monitors and tracks the user's eating habits and aims to persuade users to eat more slowly and thereby eat less; MyFitnessPal, a calorie counter and diet tracker app, which persuades users to eat healthier and exercise more; Sobriety Counter, an app that persuades users to stop drinking alcohol via motivational features such as visualizing the money a user saves by not drinking or providing scientific health statistics about the user's body and how it improves without alcohol; the GlowCap, a persuasive medication adherence technology that consists of a smart pill bottle and cap that flashes an orange light when a user should take in her medication; or MySugr, a diabetes management app, that tracks users' blood sugar and provides them with personal diabetes coaching. These are just a few examples of the many health-related BCTs that are currently available. The possible benefits of these BCTs are rather straight-forward: they help users to exercise more, achieve healthier eating habits, be more mindful, drink less alcohol, adhere to their medication schedule, and so forth. The overall hope of these technologies is that they will enable people to achieve and maintain healthier habits and lifestyles and to better manage disease or chronic conditions.

Despite the various advantages that these health-related BCTs bring about, various authors have drawn attention to multiple ethical challenges to which this type of technology gives rise. Deborah Lupton (2013), for example, has argued that these technologies promote a discourse of 'healthism'. The term healthism refers to "a discourse embraced by the socio-economically privileged who are able to position 'health' as a priority in their lives and have the economic and educational resources to do so" (2013, p.397). The concern with the discourse of healthism is that the social and economic determinants of health are neglected in favor of a focus on 'empowerment' and 'taking charge' of one's own health. This could have the unwanted outcome that people who are *not* able to partake in the healthism discourse because they lack the socio-economic privilege or because they are ill, are viewed as lacking self-responsibility and get positioned as inferior and morally deficient (Lupton 1995; Lupton 2013; Crawford 2006; Buse 2010).

A related ethical challenge that Lupton (2013) identifies is that these technologies might contribute to what she calls a 'politics of measurement'. By that, Lupton refers

to the fact that these technologies are based on quantification, often using complex algorithms to process and display the collected data. The data and numbers that these technologies produce and subsequently visually display to the users are presented as objective and neutral measurements of bodily functions and behavior activity of the users. However, as Lupton points out, "the ways in which phenomena are quantified and interpreted, and the purposes to which these measurements are put, are always implicated in social relationships, power dynamics and ways of seeing" (2013, p. 399). The numbers and data visualization produced by these health-related BCTs are thus not neutral but instead represent certain value judgements, assumptions and meanings. The idea that the data that are produced by these technologies are not objective or neutral but are instead value-laden, is a topic that I will return to shortly in more detail.

Another ethical concern, voiced most prominently by Soshana Zuboff (2015; 2019), is that the digital environments of many of these technologies commodify and attempt to modify our behavior to serve the financial interests of the creators, instead of the interests of the users.

Furthermore, multiple authors have emphasized privacy concerns: Marjolein Lanzing (2019), for example, has pointed out that users surrender "a treasure trove of highly sensitive information to these technologies that have nestled themselves comfortably within our most intimate spheres" (p.3), while many of these technologies are produced by commercial enterprises that sell user data to third parties for financial interests (Sax 2020).

Fortunately, over the years multiple authors have worked on ethical guidelines for persuasive (health) technologies in order to respond to ethical challenges. Daniel Berdichevsky and Erik Neuenschwander (1999), for example, were the first authors to write an account that is entirely dedicated to the ethics of persuasive technology (PT). In their account, they distinguish seven ethical principles for PT design and one golden rule that they consider most important. Their golden rule states that "the creators of a PT should never seek to persuade a person or persons of something they themselves would not consent to be persuaded to do" (1999, p.52). This golden rule, according to Berdichevsky and Neuenschwander, should be leading in any PT design. Other scholars, such as Andreas Spahn (2011) and Fahri Yetim (2011), have argued for a discourse ethics perspective on the development and use of PT, while Janet Davis (2009) has argued for the methodological framework of Participatory Design or Value Sensitive

Design to be applied to PT. Like Davis, I am in favor of applying the methodological framework of Value Sensitive Design (VSD) to persuasive health technologies for behavior change. I will elaborate on VSD in the next section of this introduction and argue why VSD is such an illuminating approach to ethically assess the design and use of these technologies. Right now, however, I want to draw attention to a striking aspect that has been absent in the discussion so far.

What struck me most when examining the debate on the ethics of (health-related) persuasive technologies, was that as of yet no study had explicitly focused on the ethical concerns that arise with the design and use of these technologies for *vulnerable* people. I thought this to be striking since these technologies are designed to help people change their attitudes or behaviors; something that is often particularly valuable for vulnerable people in order to better cope with their vulnerabilities. For people to cope with their vulnerabilities could either mean that they avoid the consequent harm that is likely because of their vulnerability or, alternatively, some means of lessening (the degree of) their vulnerability itself.

Because the ethical significance of vulnerability seemed so relevant in this context, and since it was largely missing from the literature, I set out to examine the ethical significance of vulnerability in relation to health-related behavior change technologies.

1.2. Note on Method

Before going deeper into the ethical significance of vulnerability, I want to make a note on method. This dissertation consists of a philosophical inquiry into moral concepts and ethical values that are of importance in relation to the design and use of health-related behavior change technologies. Philosophical inquiry consists of systematically examining ideas, concepts and argumentations. In this dissertation, I systematically examine moral concepts such as vulnerability, justice and diversity and normatively evaluate their significance in relation to the design of emerging health and wellbeing technologies. I critically evaluate the limits and strengths of the ethical design framework Value Sensitive Design (VSD). And subsequently, I reflect upon the normative conditions that are needed for an ethical design framework for health and wellbeing technologies based on a systematic reflection of Martha Nussbaum's capability theory. In chapter 5 of this dissertation, I take on a "philosopher-investigator" approach (Claassen 2011) and I cross the boundary from conceptual theorizing to conducting an empirical study wherein I thematically analyze data that I gathered

from interviews with various design-experts. This empirical investigation in chapter 5 informs and enriches the further philosophical theorizing in the concluding remarks.

1.3. Vulnerability

Now, what exactly is vulnerability? In the bioethics literature, there are broadly two responses to this question. Either vulnerability is conceptualized as (1) an ontological condition of humanity, or (2) as a marker for context-specific needs. The first view that vulnerability is an ontological condition of humanity links vulnerability to the capacity to suffer that is inherent in our human embodiment (Mackenzie et al 2013). This view that vulnerability is a universal, inevitable, and enduring aspect of our human condition to be fragile and susceptible to wounding and suffering is put forward by various scholars among which are Martha Fineman (2008), Judith Butler (2004, 2009), Alasdair MacIntyre (1999) and Bryan Turner (2006). Other scholars, in turn, have advocated the second view that conceptualizes vulnerability as a marker to identify specific persons or groups that require extra attention and care. These scholars, among which are Robert Goodin (1985), Samia Hurst (2008), and Florencia Luna (2009), focus on the contingent susceptibility of particular persons or groups to specific kinds of harm or threats by others, instead of focusing on vulnerability as an ontological human condition. The difference between the two views is clearly captured by Catriona Mackenzie, Wendy Rogers and Susan Dodds, who write: "whereas the ontological response to the question 'What is vulnerability?' stresses our common embodied humanity and equal susceptibility to suffering, this second response stresses the ways that inequalities of power, dependency, capacity, or need render some agents vulnerable to harm or exploitation by others" (2013, p.6).

Both conceptualizations of vulnerability have their own shortcomings. The ontological conception of vulnerability faces the challenge that it obscures rather than enables the identification of the context-specific needs of particular groups or individuals within populations at risk. Such a conceptualization of vulnerability would thus not be very useful when you want to identify the specific ethical concerns that may arise with health-related BCTs for particular groups or individuals. The second conceptualization of vulnerability would be more useful when you aim to identify the special needs of vulnerable people, e.g., with regard to the design and use of BCTs. However, the labeling of specific people or groups as vulnerable can create the problem that it might lead to unwarranted and unjust paternalistic responses, stereotyping, disqualification, or discrimination (Dodds 2008; Luna 2009; Fineman 2008).

In order to overcome these shortcomings while at the same time safeguarding the important features of vulnerability that both conceptualizations capture, Mackenzie, Rogers and Dodds (2013) have developed a taxonomy of vulnerability. Their taxonomy reconciles both views on vulnerability in a new, overarching conceptualization of vulnerability that distinguishes distinct but overlapping kinds of vulnerability (Mackenzie et al 2013). With their taxonomy, Mackenzie, Rogers, and Dodds (2013) reconcile the ontological conception of vulnerability with the conception of vulnerability as a marker to identify specific persons and groups that require extra attention and care in a taxonomy of vulnerability. Their taxonomy distinguishes different sources and states of vulnerability and enables a fine-grained analysis of the sense in which vulnerability is both an ontological condition of our humanity and a context-specific phenomenon. Furthermore, the taxonomy clarifies how vulnerability affects a comprehensive range of people in some way or another, and identifies who of them have specific needs. The taxonomy is powerful because it helps to clearly identify (1) the wrong or harm a person is vulnerable to, (2) the source of this vulnerability, and (3) the suitable safeguards that are needed in response.

The taxonomy distinguishes two *states* of vulnerability: vulnerability can be *dispositional* or *occurent*. In addition, there are different *sources* of vulnerability: *inherent* and *situational*. Inherent sources of vulnerability are intrinsic to the human condition; they arise from our corporeality and our affective and social natures. Situational sources of vulnerability, on the other hand, are context specific. These sources are caused, or exacerbated by, the social, political, economic, or environmental context that a person or social group is in. There is also a particular subset of situational sources, namely *pathogenic* vulnerabilities. Pathogenic vulnerabilities can arise when a response intended to ameliorate vulnerability has the paradoxical effect of exacerbating existing vulnerabilities or creating new ones. For example, people with cognitive disabilities who are vulnerable due to their care needs are susceptible to pathogenic forms of vulnerability such as emotional or physical abuse by their caregivers.

Now, sources of vulnerability are problematic when they cause a person or a group of persons to experience a diminished capacity to meet or protect their needs or safeguard their interests, and are therefore at increased risk of suffering harm or wrong.

I discuss the ethical significance of vulnerability in relation to the design and use of health-related BCTs in detail in chapter 2 of this dissertation. Making use of

Mackenzie et al.'s (2013) taxonomy, I discuss various harms and wrongs that vulnerable users[2] of BCTs can be susceptible to. Subsequently, I argue what safeguards are needed in order to assure that the design and use of these technologies take the vulnerabilities of users sufficiently into account, and how users' interests and autonomy can be respected.

In the appendix to chapter 2, I examine together with Jenneke Evers the significance of taking into account vulnerability for a specific type of health-related persuasive technology, namely *femtech* applications. Femtech entails technologies that assist women with their reproductive health. By looking through the lens of vulnerability at three hypothetical personas making use of femtech applications, Evers and I argue that these technologies can exacerbate existing vulnerabilities or even create new vulnerabilities when insufficient attention is being paid to the ethics of vulnerability in the design and implementation of these apps.

In chapter 4, which elaborates on the ethical design framework 'Capability Sensitive Design' (CSD), the concept of vulnerability plays a key role as well. How exactly vulnerability is a key feature in CSD, and how vulnerability relates to the concepts of autonomy and capability, is discussed in more detail at a later stage of this introduction. Now, I first turn to the topic of ethics by design.

1.4. Ethics by Design

Let's go back for a moment to what Deborah Lupton (2012) earlier described as 'politics of measurement'. With the term politics of measurement, Lupton referred to the idea that the numbers and data visualizations that personalized health technologies such as BCTs present to their users, are to a high extent the result of choices made about *what* data is worthwhile to collect, as well as *how* that collected data is quantified, and in *what ways* it is interpreted, and subsequently displayed to the users. All of these choices represent decisions about what is important to measure, quantify, and present,

[2] Although the labeling of specific people or groups as vulnerable can create the problem of unwarranted and unjust paternalistic responses, stereotyping, disqualification, or discrimination, as pointed out by Dodds 2008; Luna 2009; Fineman 2008, I do make use of the terms 'vulnerable people' and 'vulnerable users' throughout this dissertation. The primary reason I have chosen to still use these terms despite the criticism that they might have a stigmatizing effect is because of readability. I would like to invite the reader, however, to keep in mind that with 'vulnerable people' or 'vulnerable users' I refer to people who are (susceptible to) experiencing vulnerabilities or (susceptible to) encountering vulnerabilizing factors due to either internal, situational or pathogenic sources of vulnerability.

and what not. In other words: the ways in which these technologies measure, quantify, interpret and display data, always represents certain value judgements. Therefore, Lupton argues that the numbers produced by these technologies should not be understood as to be neutral or objective, but instead they represent a politics of measurement. That is to say: they are value-laden. A health-related BCT consists of the solidification of thousands of design decisions that reflect judgements, goals, priorities, and values of the creators. All of these design decisions, whether they concern the functionality, usability or aesthetics of the technology, affect what the technology is able to do and what *not*, how that will be available and how not, and for who that will be available and for who not.

The awareness that designers make value judgements, and that values can be actively supported, or undermined, in technology designs has increased rapidly in recent years (Van den Hoven, Vermaas, Van de Poel 2015). In response to this increasing awareness, multiple 'ethics by design' approaches emerged (see e.g., Flanagan and Nissenbaum 2014; Friedman, Kahn and Borning 2013; Friedman and Hendry 2019). These ethics by design approaches all aim to provide practical advice on how to consciously take into account values during the design process. Despite the diversity of these approaches, there are three characteristics that all approaches have in common:

(1) Every ethics by design approach shares the claim that "values can be expressed and embedded in technology" (Van den Hoven, Vermaas, Van de Poel 2015). That is to say: design decisions shape and affect the set of interactions and constraints of a technology to users, and these decisions can support or undermine various values.

(2) They all share the claim that "conscious and explicit thinking about the values that are imparted to our inventions is morally significant" (Van den Hoven, Vermaas, Van de Poel 2015). Since technology shapes and alters our lives and societies, it is of moral significance to think about the ways in which we *want* technology to shape human lives and societies, as well as reflect upon what we do *not want* for human-technology interactions.

(3) The claim is shared by all ethics by design approaches that "value considerations need to be articulated early on in the process at the moment of the design and development when they can still make a difference" (Van den Hoven, Vermaas, Van de Poel 2015). This aspect of integrating ethics early on in the design process is important

in order to stimulate designers to be explicit about their value commitments and have them proactively reflect upon, and account for, the values that they built into their design.

1.5. Value Sensitive Design

The most prominent and influential ethics by design approach is Value Sensitive Design (VSD) (Friedman et al 2013; Friedman and Hendry 2019). VSD emerged in the early to mid-1990s out of the landscape of computer science and human-technology interaction studies (see e.g., Friedman 1997; Friedman and Kahn 2003; Friedman and Nissenbaum 1996), with the aim to respond to "the perceived need for a broad-based design approach to account for human values and social context" (Friedman and Hendry 2019, p. 14). These early VSD scholars, of whom Batya Friedman and David Hendry are the most prominent ones, recognized that the predominant and traditional focus of engineers and designers had been on functionality, efficiency, usability, and affordability. Yet, these VSD scholars emphasized that there is more to technology than that (Manders-Huits 2010). As Friedman and Hendry eloquently articulate in their recent book on VSD:

> "Technology is the result of human imagination- of human beings envisioning alternatives to the status quo and acting upon the environment with the materials at hand to change the conditions of human and non-human life. As a result of this human activity, all technologies to some degree reflect, and reciprocally affect, human values. It is because of this deep-seated relationship that ignoring values in the design process is not a responsible option" (2019, p.1).

Therefore, besides paying attention to functionality, efficiency, usability, and affordability, VSD emphasizes the importance of taking into account values in the technological design process. Consequently, the objective of VSD is to address and account for values in a "principled and systematic manner throughout the technical design process" (Friedman and Hendry 2019, p.4). Now, how does it work?

VSD consists of a tripartite and iterative methodology, which entails a conceptual, empirical and technical investigation (Friedman and Hendry 2019). Each of these investigations is carried out iteratively, mutually informing and being informed by the other investigation.

The conceptual investigation involves a philosophically informed conceptual analysis, with the goal to identify and articulate the central values at stake in a particular

design context, and the stakeholders that are affected by the technology design (Manders-Huits 2010; Friedman and Hendry 2019). The empirical investigation sets out to inform the conceptual investigation on the human context in which the technology is situated. Both qualitative and quantitative research methods can be used to examine how stakeholders assess the technology in question (Friedman and Hendry 2019). The technical investigation explores how the values that are identified and assessed in the conceptual and empirical phase, can subsequently be integrated and supported by the technical design. This is where VSD has a pre-emptive stance with regard to ethics and technology: it purports to incorporate the results of the conceptual and empirical phases into the technology design in a proactive way (Friedman and Hendry 2019; Manders-Huits 2010). Through these three modes of investigation, which are alle three needed to "inform and shape and reshape each other" (Friedman and Hendry 2019, p.35), VSD provides a method to proactively account for values in a design process. One the one hand, VSD "serves as an analytical tool to open up valuation processes within technology design and development that are usually black-boxed or neglected. On the other hand, it provides a constructive tool that enables and supports the realization of specific desired values in the design and development of new technologies" (Simon et al 2020, p.5).

What I signaled when looking closely into the VSD approach, however, is that although VSD is a highly promising approach to ethics of technology design, it faces various challenges. Together with Alina Huldtgren, I identified three major challenges that VSD faces. In chapter 3 I discuss these challenges in detail, let me now introduce them briefly. The first challenge that VSD faces is that the approach obscures the voice of its practitioners and thereby risks claiming unfounded moral authority. Practitioners of VSD nowhere in the process need to make explicit their own voice, including their own values and normative commitments. This causes the risk that unfounded and unjustified moral authority is being claimed by the practitioners in the VSD process (Borning and Muller 2012; Jacobs and Huldtgren 2018). Although in the latest literature on VSD, Friedman and Hendry now do state that "designers are encouraged to make their own values, as well as the project vales, explicit and transparent throughout the design process" (2019, p.38), it remains unclear how exactly designers should do this.

The second challenge that VSD faces is that it takes stakeholder values as leading values in the design process, without questioning whether what *is* valued by stakeholders also *ought* to be valued. VSD thereby risks conflating facts and values (Manders-

Huits 2010; Jacobs and Huldtgren 2018). VSD practitioners assume to know what to do in a *normative sense* when knowing *empirically* what values stakeholders hold, thereby risking the naturalistic fallacy (Jacobs and Huldtgren 2018).

Thirdly, VSD faces the challenge of not being able to provide normative justification for making value trade-offs in the design process (Manders-Huits 2010; Jacobs and Huldtgren 2018). In cases where two or more values cannot be realized at the same time in a technology design, VSD is not able to provide direction on how designers should prioritize one value over the other.

In chapter 3 of this dissertation, I discuss these challenges in more detail and I set out to answer the question how VSD can overcome these challenges. In short, what I signaled was that all three challenges arise from the fact that VSD lacks a solid normative foundation. I therefore argue in chapter 3 that VSD practitioners need to complement VSD with an ethical theory in order to overcome these challenges. Because an ethical theory can provide sources of justification and explicit argumentation for moral claims and considerations, which are needed to make principled judgments, to avoid conflating facts with values, and to legitimize value trade-offs during the design process.

After having argued that VSD needs to be complemented with an ethical theory in chapter 3, I turn to the question *what kind* of ethical theory would be best suited to do so? Roughly, a distinction can be made between top-down, bottom-up, and mid-level ethical theories. In chapter 3, I argue in favor of a mid-level ethical theory to complement VSD. Mid-level theories often consist of a cluster of pivotal moral principles. These principles function as an analytical framework, forming the starting point for applied ethics in a specific domain, e.g., biomedical ethics, environmental ethics, or ethics of technology. Mid-level principles function as general guidelines for the formulation of more specific rules and are usually regarded as generating "prima facie" obligations, i.e., as an obligation that must be fulfilled unless it conflicts with an equal or stronger obligation (W. D. Ross 1930). In order to come to concrete action-guidance, principles must be specified. The process of specification consists of reducing the indeterminacy of abstract norms by narrowing the scope, which comes down to "spelling out where, when, why, how, by what means, to whom, or by whom the action is to be done or avoided" (Richardson 1990, p.289). When two rival, but valid, specifications of principles conflict with each other, then these conflicting

specifications need to be balanced and weighed against each other (Beauchamp and Childress 2013).

Although mid-level approaches are not free from criticism, there are two main arguments in favor of why a mid-level theory is the right kind of ethical theory to accompany VSD practitioners. Firstly, mid-level approaches are convergent; differences on the highest level of moral theory most often converge at the level of mid-level action guiding principles. As James Sterba puts it: "traditional ethical theories, be they Aristotelian, Kantian, Millian, or whatever, have come to be revised and reformed in such a way that, at least in their most morally defensible formulations, they no longer differ in the practical requirements they endorse" (2005, p.1). For example, a utilitarian and a Kantian deontologist might differ at the foundational level of moral theory, these differences ultimately can recede at the level of mid-level moral principles where the utilitarian and the Kantian deontologist could agree, for instance, on the importance of the principle of respect for autonomy in a specific practice such as research ethics (Arras 2016). In a context where people with various disciplinary backgrounds, interests, and priorities have to work together, which often is the case in design-contexts, convergence on the practical level is crucial to come to joint decisions.

Secondly, mid-level principles can provide action-guidance in concrete cases, in contrast to general precepts from classical moral theories that are often too indeterminate to do so. Mid-level principles can provide action-guidance in practice because they are a) domain specific, that is, the selection of a cluster of principles often takes place after examining considered moral judgments and the way moral beliefs cohere in a certain context (Beauchamp and Childress 2013). And b) because through the methodological tool of specification, the indeterminacy of abstract norms is reduced and content is added to abstract principles, ridding them of their indeterminateness and providing action-guiding content for the purpose of coping with complex cases (Beauchamp and Childress 2013).

After having argued that VSD should be complemented with a mid-level ethical theory I, of course, had to bite the bullet. In other words; I could not only argue that VSD needs a mid-level ethical theory, I now had to explore how exactly that complementation could be done. I took on this endeavor and explored how VSD could be complemented by the mid-level ethical theory of Martha Nussbaum's capability

theory (2000; 2006; 2011).[3] This resulted in 'Capability Sensitive Design' (CSD), a design framework that merges VSD with Nussbaum's capability theory and overcomes most of VSD's challenges. CSD, as I will argue in chapter 4 in more detail, is able to normatively assess technology design in general, and technology design for health and wellbeing for vulnerable people in particular.

1.6. Capability Sensitive Design

Nussbaum's capability theory (2000; 2006; 2011) has a substantive normative foundation. The normative foundation of Nussbaum's capability theory is the claim that to assess people's well-being, one should look at what people are able to do and be -their capabilities and functionings- and thus the kind of life that they are effectively able to lead. Furthermore, it includes the claim that all people are morally equal and deserve a life worth living, which entails that everyone is entitled to ten basic capabilities in order to live a worthy life.

The specific aim of CSD is to proactively pay attention to the expansion of one or more of these ten capabilities from the start of a design process. The added value for designers adopting CSD, is that CSD overcomes most of VSD's challenges and provides a clear normative foundation for design practices, with practical ethical guidance on what capabilities matter and why, and how these capabilities can be expanded in concrete technology design.

Now, one can question why Nussbaum's capability theory is chosen to complement VSD and not another mid-level ethical theory such as e.g., Beauchamp and Childress' principlism (2013) or Martin Peterson's geometric account to moral principles (2017)? That is because this thesis focuses on the context of technology design for health, well-being and behavior change, and it is Nussbaum's capability theory that seems particularly well-equipped to ethically assess technology design for health and wellbeing. In short, that is because the capability theory has as its prime aim to expand

[3] Strictly speaking, Nussbaum herself never identified her capability theory as a mid-level ethical theory. However, as will be explained in more detail in chapter 4 of this dissertation, at the core of Nussbaum's theory is the moral principle that all people are morally equal and deserve a life worth living, and that all people are morally entitled to at least ten capabilities in order to be able to life a worthy life. These ten capabilities as identified by Nussbaum can be understood as a cluster of moral principles that function as general guidelines for the formulation of more specific rules, and therefore Nussbaum's capability theory can be interpreted as a mid-level ethical theory.

people's capabilities, which corresponds particularly well with Lennart Nordenfelt's influential conceptualization of health as a person's ability to realize one's vital goals (Nordenfelt, 1986), as well as with the later refinement of that definition by Sridhar Venkatapuram (2013) of health as a person's ability to achieve or exercise a cluster of basic human activities or capabilities. Given these influential conceptualizations of health, Nussbaum's capability theory seems well-suited to normatively assess technology design for health and wellbeing.

Furthermore, the capability theory is of relevance for an ethics of vulnerability, which forms another important focus point of this thesis. The capability theory is attentive to vulnerability and human diversity because of its emphasis on people's distinct abilities to *convert* the resources available to them into achieved functionings. That is: on peoples varying personal, social and environmental conversion factors. As Catriona Mackenzie argues, it is because of this attention to personal circumstance and social and environmental context, that the capability theory is able to provide "a fine-grained analysis of the meaning of equality as well as different sources of social injustice within a certain context, and the impact that has on people's opportunities" (2013, p.49). It is precisely this fine-grained analysis that fits so well with the taxonomy of vulnerability by Mackenzie, Rogers and Dodds (2013) that I discussed earlier. That is because, as Mackenzie points out:

> "specific capability deficits can signal sources of occurent or dispositional vulnerability and vice versa. The notion of vulnerability also signals the actual or potential harm that may result from particular capability deficits and highlights the obligation to address those deficits in order to remediate vulnerability" (2013, p.50).

My motivation to choose Nussbaum's capability theory to complement VSD is because the theory seems particularly well-equipped to ethically assess technology design for health and wellbeing, as well as that the theory is of particular relevance for an ethics of vulnerability. In the context of this dissertation, which looks at an ethics of technology design for health and wellbeing particularly for *vulnerable* people, this seems most fitting.

This does not mean, however, that no other mid-level ethical theory could complement VSD or that Nussbaum's capability theory is always the *best* mid-level ethical theory to complement VSD. What I argue is that Nussbaum's capability theory is *particularly well-suited* to complement VSD in the particular context of ethical technology design for health and wellbeing for vulnerable people. It is however certainly not the case that Nussbaum's capability theory is the *only* ethical theory able to

complement VSD and I therefore do not deem it required in this dissertation to present an overview of arguments why Nussbaum's capability would be "better" suited to complement VSD than any other mid-level ethical theory.

1.7. A Designer's Perspective

What remains, then, is the ultimate question whether CSD is able to bridge the 'theory-practice gap'. In other words, whether CSD is practically applicable and of value in actual design practice. To find out, Wijnand IJsselsteijn and I entered into dialogue with various design-experts on ethics by design in general, and CSD in particular. The main objective of our study was to let the design-experts reflect upon the question whether the CSD framework could be of practical value in their design (research) practice. The study consisted of thematic interviews with nine design-experts, in order to explore design-experts' experiences with designing for values, what they regard as the strengths and weaknesses of CSD, and if CSD could be of practical use to their design (research) practice. We thematically analyzed these interviews; we identified, analyzed and reported themes within the gathered data. The semi-structured interviews were recorded, transcribed, and then coded, after which the codes were collated into potential themes. IJsselsteijn and I then extensively discussed and refined the potential themes in order to specify the overall story that the analysis tells. The final analysis is presented in chapter 5. The main result of our study is that CSD is less relevant for *steering* the design process, but very well *accompanies* designers in the messy, iterative and interactive process that design is.

In the concluding remarks, I reflect upon the challenges that I have raised in this thesis and propose ways forward for future research. The concluding remarks are followed by an epilogue, in which I freely envision a fictional day in the lives of a design team that applies CSD for the first time to a new technology design.

2 Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People[4]

Behaviors that pose health risks such as smoking, overeating, and physical inactivity contribute significantly to the development of various serious chronic diseases (WHO 2015), affecting personal well-being as well as indirectly adding to high health care costs. Persuasive technologies for health-related behavior change have been proposed as a way to help people reduce or eliminate these behaviors.

Persuasive technology (hereafter PT) is a class of technologies that are designed to change a person's attitude, behavior, or both (IJsselsteijn et al 2006). Important is that the standard definition of PT by Fogg prescribes that PTs always bring about *voluntary* change of behavior or attitude (Fogg 2003). Although force and misleading information or dishonest communication may persuade as well, the definition of PT by Fogg that is endorsed in this chapter explicitly excludes instances of influence by coercion or manipulation from the realm of PT. What exactly the difference is between persuasion, manipulation and coercion is elaborated on in the section "Persuasive Technology", as well as how to assure that persuasion by PT excludes any instances of coercion or manipulation.

Furthermore, this chapter focuses on cases wherein PT is used to persuade persons to act in their best interests and in alignment with their personal goals (e.g., to eat more healthy). Cases wherein persons are persuaded by PT to act in a manner not in line with their own interest (e.g., to spend money on products they do not need) are explicitly excluded from the scope of this paper. The paper focuses on instances of persuasion by PT in which a user would reasonably consent to both the *ends* of the persuasion (i.e., the target behavior of the PT) as well as the *means* of the persuasion (i.e., the persuasive tools used by the PT to persuade the user to perform the target behavior). This is discussed in more detail in the sections "Persuasive Technology" and "The Concern of Manipulation and Coercion".

---

[4] This chapter has originally been published as: Jacobs, N. (2020). Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People. *Bioethics.* Vol. 34. No.5. pp.519-526.

PTs for health and well-being are aimed at health promotion or disease management Orji and Moffatt 2016). Examples include HAPIfork, a smart fork that monitors and tracks the user's eating habits and aims to persuade users to eat more slowly; MyFitnessPal, a calorie counter and diet tracker app, which persuades users to eat healthier and exercise more; Sobriety Counter, an app that persuades users to stop drinking alcohol via motivational features such as visualizing the money a user saves by not drinking or providing scientific health statistics about the user's body and how it improves without alcohol; the GlowCap, a persuasive medication adherence technology that consists of a smart pill bottle and cap that flashes an orange light when a user should take in her medication; or MySugr, a diabetes management app, that tracks users' blood sugar and provides them with personal diabetes coaching.

The use of such PTs can give rise to multiple ethical concerns. As of yet, however, no study in the field of persuasive ethics has explicitly focused on the ethical concerns that arise with the design and use of PTs for *vulnerable* people.[5] This is striking because PTs are designed to help people change their attitudes or behaviors; something that is often particularly valuable for vulnerable people in order to better cope with their vulnerabilities. For people to cope with their vulnerabilities could either mean that they avoid the consequent harm that is likely because of their vulnerability or, alternatively some means of lessening (the degree of) their vulnerability itself. With regard to PT, an example of the first instance is the GlowCap: a PT that helps people who are vulnerable to forgetfulness to adhere to their medication regimen. An example of the second instance is an app that aims to reduce people's vulnerability to sexual assault by providing them exercises to become better at self-defense. Ideally, the degree of the vulnerability to sexual assault itself will be lessened over time with the help of the PT.

---

5 See for example: Berdichevsky, D. and Neuenschwander, E. (1999). 'Towards and Ethics of Persuasive Technology'. *Communications of the ACM*. Vol. 42. No: 5. pp. 51-58; Davis, J. (2009). 'Design Methods for Ethical Persuasive Computing'. *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*.; Yetim, F. (2011). 'A Set of Critical Heuristics for Value Sensitive Designers and Users of Persuasive Systems'. *ECIS 2011 Proceedings, Helsinki*.; Spahn, A. (2011). 'And Lead Us (Not) into Persuasion...? Persuasive Technology and the Ethics of Communication'. *Science and Engineering Ethics*. Vol. 18. No. 4. pp. 633-650.; Smids, J. (2012). 'The Voluntariness of Persuasive Technology'. In: Bang, M., Ragnemalm, E.L. (Ed.). *PERSUASIVE 2012. LNCS*, vol. 7284, pp. 123–132. Springer, Heidelberg.; Burri Gram-Hansen, S. (2013). 'Towards an Approach to Ethics and HCI Development Based on Løgstrup's Ideas'. *INTERACT '09 Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I*. pp. 200-203.

In this chapter, I will argue for the need to address the ethical concerns that arise from the design and implementation of PTs for vulnerable people, highlighting specific ethical concerns and developing suggestions for ways to deal with them. The chapter proceeds as follows: first, the notion of PT is delineated. Second, the concept of vulnerability is clarified and I discuss its ethical significance in relation to PT for health and well-being. Lastly, I discuss two ethical concerns that arise from the design and use of PTs from a vulnerability perspective.

2.1. Persuasive Technology

PTs make use of persuasion, but what exactly is *persuasion*? In the predominant typology of influence used in the bioethics literature, persuasion is often understood as rational persuasion, meaning influence by reason and argument (Blumenthal-Barby 2012). Beauchamp and Childress (2013, p.139), for instance, state that in persuasion "a person must come to believe in something through the merit of reasons another person advances." Rational persuasion is distinguished in this dominant typology from coercion (influence by force, depriving the coerced person of choice entirely) and from manipulation (meaning everything in between rational persuasion and coercion) (Blumenthal-Barby 2012).

Especially 'manipulation' is a broad and underdeveloped category in this typology and Blumenthal-Barby has rightfully pointed out that this traditional tripartite categorization of influence used in bioethics literature is "in desperate need of conceptual refining and ethical analysis" (2012, p.345).

In order to conceptually refine the categories of persuasion, coercion, and manipulation, I propose the following definitions of each category: I suggest letting go of the narrow understanding of persuasion as being solely rational and instead understand persuasion in a much broader sense that includes both rational means as well as non-rational means of influence. Non-rational means of influence includes e.g., framing, setting up defaults, changes in choice architecture, playing on emotions, peer pressure, or appeals to authority (petty and Cacioppo 1986; Thaler and Sunstein 2008; Blumenthal-Barby 2012). Of crucial importance is that persuasion, both by rational as well as by non-rational means, never (1) significantly blocks or burdens options, that (2) a person is aware of the fact that she is being intentionally influenced, and aware

of the mechanisms of that influence, and (3) that the influence is in the best interests and in alignment with the personal goals of the person being influenced.[6]

Manipulation is best understood as defined by Susser, Roesser and Nissenbaum (2018, p.22) as: "imposing a hidden or covert influence on another person's decision-making." The most important distinction between persuasion and manipulation is that the latter "disrupts the target's capacity for self-authorship", i.e.; someone is being influenced in a way that thwarts their capacity to become aware of this.

Coercion, lastly, is to be understood as influencing someone by irresistible treats, depriving someone of choice entirely (Susser, Roesser and Nissenbaum 2018).

In standard ethical analysis, coercion is almost always impermissible, except when someone forms a threat to self or others (Blumenthal-Barby 2012). Manipulation, in which a person's decision-making power is covertly subverted and their autonomy is undermined, is always ethically impermissible. Whether persuasion is ethically permissible or not depends on the following two issues: (1) whether the persuasion happens in alignment with a person's own goals and whether the PT sufficiently takes into account a person's interests and needs (this issue is further elaborated on in the section "The Concern of Taking Into Account Users' Interests"). And (2) whether the instance of persuasion thwarts a person's autonomy (this issue is discussed in detail in the section "The Concern of Manipulation and Coercion").

Lastly, one must note that persuasion by technology is different from persuasion by humans in several important ways. Firstly, technology does not inherently possess the ability to respond to a human being in the same way another human being would. This necessitates an explicit consideration of what kinds of responses are desirable. In addition, technology is inherently persistent: a computer does not get tired, discouraged or frustrated like humans do. Furthermore, technology is ubiquitous and may have access to people's most private locations, like the bed- or bathroom, where a

---

6 These three criteria are based upon criteria identified by Thaler, R.H. (2015) 'The Power of Nudges, for good and bad'. *The New York Times*. Accessed on June 8th 2019 at: https://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html on what constitutes a morally permissible nudge. And on the criteria identified by Blumenthal-Barby, J.S. (2012). 'Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts'. *Kennedy Institute of Ethics Journal*. Vol.22, No.4. pp.345-366. on when autonomy and nonargumentative influence are compatible.

human persuader would not be allowed to come (Fogg 2003). Technology thus changes the game of persuasion, giving rise to new ethical challenges.

2.2. Dimensions of Vulnerability

In the bioethics literature, vulnerability has often been conceptualized in two distinct ways: either as (1) an ontological condition of humanity, or (2) as a marker for context-specific needs.

The first view that vulnerability is an ontological condition of humanity entails that vulnerability is a universal, inevitable, and enduring aspect of the human condition to be fragile and susceptible to wounding and suffering (Fineman 2008). However, the problem with such an ontological conception of vulnerability is that it obscures rather than enables the identification of the context-specific needs of particular groups or individuals within populations at risk. Such a conceptualization of vulnerability is not useful when we aim to identify the specific ethical concerns that may arise with PT for particular groups or individuals.

In response to this ontological view of vulnerability, others have conceptualized vulnerability on the contrary as a marker to identify specific persons or groups that require extra attention and care (Levine et al 2004; Luna 2009). This approach to vulnerability is more useful when we aim to identify the special needs of vulnerable people with regard to the design and use of PT. However, the labeling of specific people or groups as vulnerable can create another problem, since it might lead to unwarranted and unjust paternalistic responses, stereotyping, disqualification, or discrimination (Mackenzie et al 2013; Meek Lange et al 2013).

Instead of endorsing one of the two abovementioned conceptualizations of vulnerability, this chapter endorses a third view to vulnerability that reconciles both previous views on vulnerability in a new, overarching conceptualization of vulnerability. This third view on vulnerability that is endorsed in this chapter is the view by Mackenzie, Rogers and Dodds (2013) who have developed a taxonomy of vulnerability.

2.3. Taxonomy of Vulnerability

Mackenzie, Rogers, and Dodds (2013) have reconciled the ontological conception of vulnerability with the conception of vulnerability as a marker to identify specific

persons and groups that require extra attention and care in a taxonomy of vulnerability. This taxonomy distinguishes different sources and states of vulnerability and enables a fine-grained analysis of the sense in which vulnerability is both an ontological condition of our humanity and a context-specific phenomenon. Furthermore, the taxonomy clarifies how vulnerability affects a comprehensive range of people in some way or another, and identifies who of them have specific needs. The taxonomy is powerful because it helps to clearly identify (1) the wrong or harm a person is vulnerable to, (2) the source of this vulnerability, and (3) the suitable safeguards that are needed in response.

The taxonomy distinguishes two *states* of vulnerability: vulnerability can be *dispositional* or *occurent*. In addition, there are different *sources* of vulnerability: *inherent* and *situational*. Inherent sources of vulnerability are intrinsic to the human condition; they arise from our corporeality and our affective and social natures (Mackenzie et al 2013). Situational sources of vulnerability, on the other hand, are context specific. These sources are caused, or exacerbated by, the social, political, economic, or environmental context that a person or social group is in (Mackenzie et al 2013). There is also a particular subset of situational sources, namely *pathogenic* vulnerabilities. Pathogenic vulnerabilities can arise when a response intended to ameliorate vulnerability has the paradoxical effect of exacerbating existing vulnerabilities or creating new ones (Mackenzie et al 2013). For example, people with cognitive disabilities who are vulnerable due to their care needs are susceptible to pathogenic forms of vulnerability such as emotional or physical abuse by their caregivers.

Sources of vulnerability are problematic when they cause a person or a group of persons to experience a diminished capacity to meet or protect their needs or safeguard their interests, and are therefore at increased risk of suffering harm or wrong.

2.4. Persuasive Ethics from a Vulnerability Perspective

To indicate the various ethical concerns that could arise for a vulnerable person from the use of PT for health-related behavior change, I discuss two examples of such technologies used by persons that are vulnerable in distinct ways. The first example concerns a persuasive medication adherence technology that is used by an older woman. The second example concerns a calorie counter app that aims to persuade users to lose weight that is used by a young woman with an eating disorder.

Miriam and The GlowCap

The Vitality GlowCap is a persuasive "medication adherence system comprised of a smart cap and bottle".[7] "Automated visual and audible alerts during scheduled dosage windows signal that it is time for the user to take his or her medications"[8], the GlowCap thus intervenes at the right time, making use of the persuasive tool of suggestion. Furthermore, the GlowCap makes use of the persuasive tools of self-tracking and surveillance: the user has access to an online portal where she can view her own progress report on how well she adhered to her medication regimen. The company behind the PT can also send this online progress report to clinicians, care managers, and family or friends at the primary user's behest, who can subsequently act on the information provided by the report.

Imagine Miriam, an 88-year-old woman who lives independently at her home. Miriam needs to take various medicines each day in order to lower her bad cholesterol, manage her high blood pressure, and deal with osteoporosis, amongst others. However, Miriam has trouble remembering when she needs to take her medication. The GlowCap can assist Miriam to adhere to her medication schedule, which ultimately helps Miriam to live independently at home. However, beside this advantage there are also serious concerns attached to the use of the GlowCap for a person such as Miriam that might cause her to become more vulnerable than she initially was. As mentioned, the GlowCap makes use of the persuasive tool of surveillance by creating an online progress report on how well Miriam adheres to her medication schedule and the company behind the GlowCap can share this online report with clinicians, care managers, and family or friends at the primary user's behest. They can subsequently act on the information provided by the report. Due to her age, Miriam relies on the care and support of family members and caregivers and this reliance could make Miriam reluctant to refuse granting those caregivers and family member's access to her online data that is gathered by the GlowCap. But being surveilled by family and caregivers diminishes Miriam's privacy, and the realization that her family and caregivers might act upon the information that is gathered in her online progress report (e.g. decide that Miriam is no longer able to live independently at home but needs to go to a nursing home, while this is against Miriam's own will) could render Miriam to feel more powerless than she felt before. Thus, although the GlowCap aims to ameliorate Miriam's medication adherence, it might have the negative effect of rendering her to

7 Accessed on 20-01-2019 at: https://nanthealth.com/vitality/.

8 Ibid.

feel more powerless than she felt before. Without proper attention to the complexities of vulnerability, a technology designed with the intent to ameliorate certain vulnerabilities may instead become an additional pathogenic source of vulnerability.

Elisa and MyFitnessPal

Imagine Elisa, a woman in her early twenties who struggles with exercise bulimia.[9] Elisa uses MyFitnessPal, a calorie-counting app that allows users to track and input their daily food intake, provides feedback on the number of calories and nutrients needed per day, allows users to set weight and nutrient goals and advises on how to reach these goals (Levinson et al 2017). Elisa derives her self-worth to a great extent from what others think of her and believes she has to meet certain beauty ideals to be accepted by others. This is a situational source of vulnerability that Elisa suffers from, which makes her susceptible to exercise bulimia. This susceptibility to exercise bulimia is an occurent and inherent vulnerability to Elisa.

The app MyFitnessPal is designed to help users obtain and maintain healthy weight goals; however, research shows that it "is widely used in an eating disorder population and is perceived as contributing to eating disorder symptoms" (Levinson et al. 2017). MyFitnessPal makes use of the persuasive tool of conditioning: it rewards users with trophies or badges for receiving certain health-related goals, and enables competition with other users by comparison of your achievements to the achievements of others using the app. This community component can be especially triggering for people suffering from an eating disorder (ED) since these are competitive illnesses (Sharkey 2018). The persuasive tool of conditioning and its competition component could push driven eating disorder-sufferers to under-eat or fast longer and more often (Sharkey 2018). A real-life ED sufferer explained in an interview that "having an app telling me how far I've gone just spurs me on to want to fast more" (Sharkey 2018). Another real-life ED sufferer testified in the same interview that it could become a compulsion to keep track of your calorie intake, logging in multiple times a day to add every detail of your diet in the app, "changing what I would eat and plan to eat so that it would always be within my 'acceptable' calorie range" (Sharkey 2018). For individuals who do not have a dysfunctional relationship with food, an app like MyFitnessPal can provide the data and structure that is needed to meet health-related target behaviors. For others, like Elisa, it can exacerbate situational and inherent sources of

[9] Exercise bulimia is a subset of the eating disorder bulimia in which a person is compelled to exercise in an effort aimed at burning calories to an excessive level that negatively affects their health. Accessed on 26-06-2019 at: https://en.wikipedia.org/wiki/Exercise_bulimia.

vulnerability and support, or even encourage, unhealthy and harmful behavior with potential long-lasting physical and mental consequences.

The examples of Miriam and Elisa highlight various ethical concerns that might arise with the use of PT by people who are vulnerable. There are two concerns in particular that need close examination and I discuss them below.

2.5. The Concern of Taking into Account Users' Interests

A concern that clearly comes to the fore in both the example of Miriam's use of the GlowCap, as well as in Elisa's use of MyFitnessPal, is that the design of these PTs is not adequately informed by the experiences, interests or needs of its users. In the case of Miriam, the data sharing function of the GlowCap did not adequately take her autonomy into account but instead made her feel powerless. In the example of Elisa, it became clear that the design of a diet app like MyFitnessPal is not informed by the experiences of users who struggle with body image problems and eating disorders.

In order to avoid that a technology takes the needs and interests of users insufficiently or inadequately into account, technology design should be informed by the experiences, interests, and needs of prospective users. The aim is to ensure that technology design is inclusive and avoids mistaken assumptions about how the technology will be used that does not take into account the likely diversity of users. To assure that the interests and needs of (vulnerable) users are taken into account, designers have to elicit the needs of their prospective users during the design process while keeping two crucial aspects in mind, as put forward by Pommeranz et al: "(1) taking real life contexts into account and (2) supporting communication between stakeholders and designers" (Pommeranz et al 2012).

What goes wrong if designers fail to take into account these two crucial aspects becomes clear when we take a critical look at the 'golden rule' for PT design that was developed by Berdichevsky and Neuenschwander (1999) in their influential article on the ethics of persuasive technology. Their golden rule prescribes that as a designer, you should never do to others what you don't want to be done to you. However, for this golden rule to be meaningful a designer needs to assume that everyone is more or less like her, while they are not. When you design for people who, due to their inherent or situational vulnerabilities, do not have the same needs or interests as you do, how

does the golden rule then guarantee that you as a designer create what is best not only for yourself but also for *others*?

The pitfall is that designers compile an idealized person that is based on their own experiences, needs and preferences and take that as their vantage point. Thereby not adequately taking real life contexts into account and subsequently providing vulnerable users who differ from this idealized vantage point with less effective attention and provisions in the technology design. This pitfall clearly occurs with the design of MyFitnessPal, which is aimed at 'ideal' individuals who have a functional relationship with food and fails to provide effective attention and provisions to vulnerable users who have a dysfunctional relationship with food.

Another related pitfall is that designers create interventions aimed to ameliorate the inherent or situational vulnerabilities of their intended prospective users, but that these interventions have instead the contradictory effect of creating or exacerbating vulnerability, for instance by (unintentionally) undermining the autonomy of the user or by increasing their sense of powerlessness, as we have seen occur with Miriam and the data sharing function of the GlowCap. In order to avoid this pitfall, it is crucial that there are adequate tools that support communication between designers and stakeholders on the values, needs and interests important to these stakeholders.

Taking real life contexts into account and supporting communication between designers and stakeholders can, to a large extent, avoid these pitfalls. When the design of PT is informed by the experiences, needs and interests of the (vulnerable) people the technology is designed for, the pitfalls of unintentionally creating or exacerbating vulnerability by design, or the risk of providing vulnerable users with ineffective attention and provisions by design are minimized. To make sure that a design is indeed informed by the experiences, needs and interests of the (vulnerable) people the technology is designed for, we need adequate elicitation tools. The problem is, however, that there is a lack of elicitation tools that support a shared understanding of interests between stakeholders and designers, and that support self-reflection by stakeholders (Pommeranz et al. 2012). Furthermore, there exist great personal differences in which method for expressing needs and interests works best (Pommeranz et al. 2012). This is a problem in general, but it is an even bigger problem for vulnerable people in particular who are often already less able to indicate their needs or protect their interests. Thus, what we are in urgent need of is further research on adequate elicitation tools that specifically work best for vulnerable people.

2.6. The Concern of Manipulation and Coercion

To respect a person's autonomy is to acknowledge a person's right to hold views, make choices and take actions that are based on a person's own values and beliefs (Beauchamp and Childress 2013). Respect for personal autonomy is an important value, especially in responses to vulnerability. That is, as Mackenzie (2013, p.45) points out, because respect for autonomy can "counter the sense of powerlessness and loss of agency that is often associated with vulnerability" and because respect for autonomy can counter the risks of coercion and manipulation.

The problem with manipulation (i.e., imposing a hidden influence on another person's decision-making and thereby infringing upon their autonomy) and coercion (i.e., influencing someone by irresistible treats and depriving someone of choice entirely) is that it has corrosive effects on a person's autonomy. In order to protect people's autonomy, instances of manipulation and coercion should be explicitly excluded from the realm of PT.

However, the boundary lines between coercion, manipulation and persuasion are not always as clear-cut. As already discussed in the section "Persuasive Technology", for PT to exclude instances of manipulation or coercion and to respect a person's autonomy, PT must meet the following three criteria. A PT may never (1) significantly block or burden a user's options, a PT must (2) make its user aware of the fact that she is being intentionally persuaded and make her aware of the tools of persuasion, and (3) a PT must persuade a person in alignment to her own personal goals, i.e. a person must share the targeted behavioral outcome of the PT.[10] Thus for example, in short, with a PT that aims to persuade users to exercise more, a user must share this target behavior and should be aware that the PT will make use of persuasive tools such as sending encouraging messages at the right time and giving rewards for achievements in order to persuade the user to exercise more and the user should always have options to choose otherwise than the PT suggests or to stop using the PT altogether.

---

[10] These three criteria are based upon criteria identified by Thaler, R.H. (2015) 'The Power of Nudges, for good and bad'. *The New York Times*. Accessed on June 8th 2019 at: https://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html on what constitutes a morally permissible nudge. And on the criteria identified by Blumenthal-Barby, J.S. (2012). 'Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts'. *Kennedy Institute of Ethics Journal*. Vol.22, No.4. pp.345-366. on when autonomy and nonargumentative influence are compatible.

For vulnerable people who may, due to various inherent, situational or pathogenic sources of vulnerability at play, have a diminished capacity to protect their needs and safeguard their interests, it is especially important that a PT meets these three criteria and that the technology does not thwart their autonomy. But how can we assure that a PT meets these three criteria and that users' autonomy is respected, and instances of manipulation and coercion are excluded? This can be assured by a valid consent procedure for PT.

What exactly constitutes a valid consent procedure for PT is a question that requires an extensive analysis that exceeds the limits of this chapter. What I provide here is an outline of how a consent procedure for PT should look like.

To start with, it must be clear *to what* exactly users of a PT are consenting to. I distinguish four aspects of PTs to which users consent. First, there are the goals and intended behavioral outcomes of a PT, e.g., better medication adherence, losing weight, exercise more or drink less alcohol. Secondly, there are the persuasive tools that a PT utilizes. Persuasive tools are the strategies that a PT applies to change a user's attitude or behavior. Examples are the persuasive tool of reduction; reducing complex behavior to simple tasks can influence users to perform the desired behavior. Or the persuasive tool of self-monitoring; by letting users track their own performance they know how well they are performing the target behavior, which increases the likelihood that they will continue (Fogg 2003). Thirdly, there are the types of individual interactions of the PT with the user throughout the use of the technology. This can, for example, be a tailored text message or a blinking light that indicates that this is the opportune moment to perform the target behavior. The fourth aspect that users need to consent to is the use and storage of their data by the company behind the PT. In order to assure valid consent, a user should give consent to *all four* aspects of a PT.[11]

We can account for that via design requirements that make it easy for the user to give consent on all four aspects. What I propose is that design requirements for PTs are informed by Onora O'Neill's (2003) conception of consent, since O'Neill has come up with a relatively simple conception of consent that appears to be well suited to the context of PT. O'Neill has argued that the practice of consent should be understood

[11] This is similar to what is sometimes referred to as 'separate granular consent options' in relation to the conditions for consent in the European General Data Protection Regulation (GDPR).

as a way to prevent that someone is manipulated or coerced. This is assured when someone is given a limited amount of accurate and relevant information (thereby assuring that someone is not overwhelmed by information) and the person is provided with user-friendly ways to extend this amount, thereby checking that they are not manipulated. As well as easy ways of rescinding consent once given, thereby checking that they are not coerced.

O'Neill's conception of consent can be translated into design requirements for PT. Consider the following, somewhat simplified, example: an app that aims to persuade users to drink less alcohol. For such a persuasive app to assure valid consent from its users, the design requirements for consent should look something like this: a simple user interface where the four aspects of the PT are clearly distinguished. When the user clicks on 'intended behavioral outcome' she sees the target behavior clearly explained, in this case alcohol intake reduction. The user than sets a specific goal, e.g., a maximum of 2 units of alcohol per week. Subsequently, the user needs to answer 'do you consent to this intended outcome?' with the options 'yes/no/more info'. After ticking on the option 'more info' the user is provided with more detailed information on the intended behavioral outcome and is again presented with the options yes/no. The same mechanism holds for aspect two, where the user is e.g., asked 'do you consent to the use of the persuasive tool of self-monitoring, which consist of self-tracking the amount of alcohol you take in per day?', again with the options yes/no/more info. For the third aspect, i.e., the sort of individual interactions, the user has to answer the question 'do you consent to the app sending you a maximum of five text messages throughout the day? Yes/no/more info. Lastly, for the fourth aspect the user has to answer the question 'do you consent to the company behind the app to use and store your data?' Yes/no/more info.

Of course, if a user answers 'no' to one or several of the aspects, the user should still be able to make full use of the PT. Furthermore, the design should be such that it is easy for the user to change consent settings at any time to enable the user to rescind consent once given[12], which is necessary in order to assure O'Neill's requirement that a person is never coerced.

---

12 An anonymous reviewer has rightly pointed out that the possibility for users to rescind their consent once given could interfere with the goal of keeping users persistent in order for users to achieve changes in their behavior or attitude. This is indeed a valid point, but I want to emphasize that it is of primary importance to assure that users voluntarily consent to their use of a PT (and if they do no longer, they should be able to rescind their consent once given). Only when consent is reasonably assured, the goal of keeping users persistent should be obtained.

However, unintended outcomes of a PT can complicate consent. Often, it is difficult to predict in advance whether a PT will have the intended effect on a user's behavior, since the outcome might differ depending on the individual characteristics of each user. Think e.g., of the unintended outcome that the GlowCap had on Miriam's sense of powerlessness, creating a pathogenic source of vulnerability for her. The taxonomy of vulnerability calls attention to the fact that various sources of vulnerability affect the overall result of an intervention. In addition, the effects of especially dispositional vulnerabilities and situational sources of vulnerability, although potentially influential, may not be immediately visible. To forestall this problem of unintended outcomes as best as possible, it is important to provide a user with easy ways of rescinding consent once given.

Furthermore, a person can change over time. In the context of PTs this is especially plausible since that is exactly what the technology is after: a change in attitude and behavior. This could lead to the situation that a person does not consent anymore to one of the four aspects of the PT, like she did at the beginning of her use of the technology. Again, in order to forestall this problem, it is crucial to provide the user with easy ways of rescinding consent once given.

Also, it is important to keep in mind that a consent procedure itself may compound (a) source(s) of vulnerability. For example, when the information provided in the consent process is formulated in a way that is hard to understand for people with little education. They might find the information that is provided to them hard to understand but at the same time feel the need to make a decision, thereby consenting (or not consenting) to something that they do not quite understand. Having to respond to something that you do not quite understand could cause the uncomfortable feeling of falling short. In such a case, the consent procedure compounds a new pathogenic vulnerability. In order to forestall this, the information provided in a consent procedure should not only be accurate and relevant as O'Neill has pointed out, but also easy to understand. In general, it is important that designers pay attention to the possibility that a consent procedure itself may compound new vulnerabilities for (vulnerable) people and designers should reflect upon how to forestall this.

Moreover, one could question at a more fundamental level whether we need consent requirements with regard to PTs at all? Could someone not just stop using a PT the

minute she senses she is no longer voluntarily using it? After all, unlike the decision to undergo surgery, the decision to use a PT seems easily reversible.

This assumption is based on a misunderstanding, especially with regard to vulnerable people who might become more easily dependent, due to either psychological factors or external pressure, on the workings of a PT. Being dependent on a PT could have the result that it is difficult to 'just stop' using the PT when you are no longer voluntarily using it.

Consider the example of a vulnerable person who suffered from a severe burnout and is using a PT that aims to reintegrate her in the workplace. For this person it might be difficult to stop using the technology since that might complicate her reintegration trajectory, although she no longer voluntarily ascribes to the PT. Or consider the fact that employers more frequently prescribe the use of PTs to employees. If an employee would decide to stop using the technology, it might occur that she herself than has to repay the costs made to her employer, leaving her with a hard penalty for opting out.

Lastly, a serious concern with any form of consent procedure (to a larger or smaller extent depending on what the procedure looks like) is that users do not really read the terms and conditions that a technology or service provides. Instead, Bernal has pointed out that users generally scroll down a long page of writing and click 'OK' at the end to confirm they agree to the terms and conditions, without actually reading them. Bernal (2014) calls this "click-wrap consent", which is close to meaningless. A similar point is made by Custers (2016), who points out that people generally don't read privacy policies. And *if* people do read them, they often don't understand them, or they lack the required knowledge to make an informed decision, or they are not offered the choice that reflects their preferences. This is a serious complication with consent procedures that is not easy to solve. However, the consent procedure inspired by O'Neill that has been put forward in this chapter is a relatively easy and user-friendly consent procedure that lowers the threshold for users to really read, understand, provide, and possibly rescind consent when they want to.

To conclude this section, we need design requirements for PTs that assure genuine consent by accounting for all four aspects of PTs, i.e. the intended behavioral outcomes, the persuasive tools, the type of individual interactions between PT and user, and the use and storage of user data by the company behind the PT, by giving users a limited amount of accurate and relevant information and provide them with user-friendly ways to extend this amount, thereby checking that they are not manipulated.

As well as easy ways of rescinding consent once given, thereby checking that they are not coerced.

2.7. Concluding Remarks

I began this chapter by arguing that while PT poses ethical challenges in general, it is especially important to pay attention to the challenges it poses for vulnerable people. Vulnerability is a complicated concept that is both an ontological condition of our humanity and highly context-specific. To better understand vulnerability and how to take it into account in the design of PTs, I used the taxonomy of Mackenzie, Rogers, and Dodds. This taxonomy enables us to identify (1) the wrong or harm a person is vulnerable to, (2) the source of this vulnerability, and (3) the suitable safeguards that are needed in response. I subsequently highlighted two ethical concerns that arise when designing PTs for vulnerable people: taking into account users' needs and interests, and securing user autonomy. I argued that taking into account users' interests is crucial in technology design, but it poses difficult challenges with regard to vulnerable users. Future research is needed on interest elicitation tools specifically for people with vulnerabilities. With regard to autonomy, I argued that it is crucial to have a notion of consent that is suitable to the domain of PT in order to assure that instances of manipulation and coercion are excluded from PTs. Such a notion of consent needs to assure respect for user autonomy with regard to all four aspects of PT, as well as easy ways of rescinding consent once given. Only if these two ethical concerns are carefully taken into account in the design and implementation of PTs for health and well-being, are these technologies suitable to be used by vulnerable people.

Appendix – The Vulnerabilities of Femtech

Naomi Jacobs and Jenneke Evers[13]

Femtech is the collective name for apps focused on women's reproductive health, including apps to track your period, to help manage your fertility, or to track and support your pregnancy. Femtech is popular: it is the fourth most popular type of app amongst adults and the expectation is that by 2023 the femtech market will be worth 50 billion US dollar (Frost and Sullivan 2018).[14]

Femtech promises to empower women by providing them self-knowledge and control over their bodies. The period- and fertility tracker *Glow*, for example, advertises with the slogan that users can "simply be in control" of their fertility when they use the app. The period- and ovulation tracker *Clue* proclaims that it helps women to "demystify" their menstrual cycle. However, despite these promises of self-knowledge and control, these apps raise serious concerns. These concerns become most clear when we look at them through the lens of vulnerability. Vulnerability indicates a person's diminished capacity to meet or protect her needs or safeguard her interests, putting her at increased risk of suffering harm or wrong (Jacobs 2020). Explicitly paying attention to vulnerability in relation to femtech, enables us to signal users' increased risks of suffering harm, which, as we argue, is ethically significant.

Mackenzie, Rogers and Dodds (2013) have created a taxonomy of vulnerability that provides insight into the various sources of vulnerability; i.e., inherent, situational or

---

13 This appendix has originally been published as: Jacobs, N. and Evers, J. (2019). De Kwetsbaarheid van Femtech. *Podium voor Bio-Ethiek*. Vol. 26. No.4. pp.13-15. The idea for this article has been a joint idea by Jenneke Evers and me that arose at the workshop "Digital Behavioral Technology, Vulnerability and Justice" at LMU and TUM. The elaboration of the original idea and the writing of the article has been my contribution. Evers, with her expertise in law and digital technologies, particularly contributed to the parts of the article on data usage by femtech applications. Together we revised and rewrote parts of the article, which resulted into this appendix.

14 Accessed on April 28th 2021 at: https://ww2.frost.com/event/calendar/2018-frost-sullivan-predictions-and-outlook-digital-health/

pathogenic sources of vulnerability. Inherent sources of vulnerability are intrinsic to our human condition and have to do with our corporeality and our affective and social nature. Situational sources of vulnerability are context-specific; these sources arise, or are exacerbated by social, political, economic or environmental conditions. Pathogenic vulnerabilities, which are a sub-set of situational sources of vulnerability, can arise when a response intended to ameliorate vulnerability has the paradoxical effect of exacerbating existing vulnerabilities or creating new ones. Although all sources of vulnerability as distinguished by Mackenzie, Rogers and Dodds (2013) can be problematic, we will focus specifically on pathogenic sources of vulnerability. That is because pathogenic sources of vulnerability can particularly undermine a person's autonomy and increase a sense of powerlessness. By applying the taxonomy of vulnerability to three hypothetical users of femtech apps, we will argue that these apps can form a pathogenic source of vulnerability to its users. By being a pathogenic source of vulnerability, these apps can put users at an increased risk of suffering harm or wrong and undermine their autonomy, which is ethically problematic.

Hannah, 18-years old

Hannah is an 18-year-old woman who grew up in a strict religious Christian Reformed family. Hannah has been using the app *Eve by Glow* to track her irregular menstrual cycle. She trusts the information that the app is providing her about her fertile window and about the days it is "safe" to have unprotected sex without becoming pregnant. Despite this, Hannah has become pregnant unintendedly after having sex with her boyfriend. She then decides to undergo an abortion at seven weeks of pregnancy.

Hannah is vulnerable in various ways. First of all, her pregnancy was an inherent source of vulnerability because pregnancies involve physical risks and Hannah's irregular menstrual cycle is an inherent source of vulnerability because it makes (knowledge about) her fertile window irregular. Furthermore, Hannah faces situational sources of vulnerability because her social circle, including her family and friends, might be (highly) critical of abortion because of their religious beliefs. This potential lack of social support makes Hannah situationally vulnerable. Then there is the app itself, that forms a pathogenic source of vulnerability for Hannah since the app does not seem able to deal with Hannah's situation and keeps sending her notifications that her period is overdue and that she might be pregnant, although she just had an abortion and waits for her menstrual cycle to start again. Hannah may experience these app

notifications as painful and saddening, which may cause her to feel *even more* vulnerable than she already was. When Hannah's cycle starts again, the app automatically adjusts all her data averages, with the result that the app's future algorithmic predictions about Hannah's cycle are no longer accurate (Tiffany 2018). Although promising to provide women with knowledge and control, the app fails precisely on that part: it is unable to provide Hannah with accurate information because her situation cannot be captured by the data-reduced functionalities of the app.

Robin, 28-years old

Robin identifies as a transgender and makes use of an app to track his menstrual cycle in order for him to be both mentally as well as physically prepared for his period. Robin's inherent vulnerability is that his body menstruates although that doesn't "fit" with his experience and perception of his own body. This might make him feel insecure, also in relation to his partner. There are situational vulnerabilities at play for Robin as well; disposing a tampon at the men's toilet causes him stress, and Robin prefers being able to discreetly buy tampons at unobtrusive times, not when he is surprised by the start of his period.

Most femtech apps are specifically aimed at fertility, and portray the menstrual cycle as something "typically feminine". The app *Flo*, for example, has a stereotypical feminine interface of purple and pink colors and addresses its users with 'hey girl'. This forms a pathogenic vulnerability for Robin, who doesn't want to be stereotypically addressed as a girl. Important to keep in mind is that many femtech applications are being developed and designed in a specific socio-cultural context that is being dominated by young white males who are primarily focused on a user group of heterosexual, fertile, cis-women (Delano 2015). In other words, the lack of diversity in the design teams of these apps, influences -and moreover limits- the functionality as well as the intended uses of the design (Hendl, Jansky and Wild 2019). Although someone like Robin can benefit greatly from tracking his period, he is to a large extent excluded from the design of these apps, which may exacerbate his vulnerability.

Claudia, 38-years old

Claudia and her wife Anne want to become parents and therefore Claudia undergoes IVF-treatment to become pregnant with the help of a sperm donor. She makes use of the app *Ovia* to track her cycle in support of her IVF-treatment. Claudia had chosen the app *Ovia* because her employer, a big multinational, reimburses her gym membership when she makes use of the app. Claudia experiences pressure to bring this IVF-treatment to a successful end because it is uncertain whether IVF will remain covered

for lesbians by Dutch health insurance. This is a situational source of vulnerability for Claudia, and partly because of this she makes use of *Ovia* to help track her cycle with the hope of a better chance of success.

The community platforms of *Ovia* cause a pathogenic vulnerability for Claudia; on these platforms she reads about all the women who did become pregnant that month and this gives Claudia a sense of failure. Furthermore, and perhaps most importantly, *Ovia* sells -in violation with European regulations- her data to her US employer and to advertisers (Harwell 2019). Reports in *The Wallstreet Journal* by journalist Sam Schechner and data scientist Mark Secada (2019) show that data is being sold to third parties, like Facebook, as well. Most of these apps do this for advertisement purposes, other apps act as 'data broker' in data ranging from strictly medical data to data on moods and sexual activity. This almost always happens "anonymously" or "aggregated", as the privacy policy of these apps say. What Claudia doesn't know, however, is that when these datasets are combined with other data, they can be traced back to the individual. The fact that her employer now possesses this data, may create a new vulnerability for Claudia: her temporary contract expires in a year, and these data may affect her contract extension (Harwell 2019).[15]

Concluding Remarks

What the stories of Hannah, Robin, and Claudia make clear is that users of femtech are as diverse as their reasons for using these applications. By looking through the lens of vulnerability at the cases of Hannah, Robin, and Claudia, it becomes apparent how femtech apps can be a pathogenic source of vulnerability. That is: although these apps aim to assist users in tracking their cycles and the like, they can easily exacerbate, or create new vulnerabilities for users and thereby undermine their autonomy. The workings of Hannah's app had the result that Hannah could no longer trust the information provided by the app because her real-life experiences could not be captured in the app's design, Robin was being excluded by the design of his app, and the data that Claudia shared with her app might be used against her when her contract renewal comes up.

---

15 Pregnancy discrimination is unfortunately still common. Research by the Netherlands Institute for Human Rights (College voor de Rechten van de Mens) has shown that 43% of pregnant women have to deal with suspected discrimination because of pregnancy or young motherhood (2016). 44% of the women surveyed with an expiring contract, thought their contract was probably not renewed because of their pregnancy.

We understand it to be the responsibility of designers of femtech applications to thoroughly account for diversity in their designs, in order to safeguard that their designs won't exacerbate users' vulnerabilities or create new ones, nor undermine users' autonomy. This responsibility of designers has its source in the moral duty to respect people's autonomy, a duty that currently is not sufficiently respected, unfortunately. Instead, ideally, femtech would alleviate existing vulnerabilities of users and strengthen their autonomy. How designers can give shape to this moral responsibility is discussed in chapter 4, where the ethics by design framework 'Capability Sensitive Design' is presented and discussed in detail.

3 Why Value Sensitive Design Needs Ethical Commitments

Naomi Jacobs and Alina Huldtgren[16]

Value Sensitive Design (VSD) is an approach to the design of technology "that accounts for human values in a principled and comprehensive manner throughout the design process" (Friedman et al 2013, p.55). The unique opportunity that the VSD approach brings to the fore is a proactive integration of ethics in the design of technology (Van den Hoven 2008).

Although VSD draws on ethical theories to identify values that are relevant for technology design (Friedman et al 2013), VSD makes no explicit commitment to particular ethical theories. This gives rise to multiple critiques to VSD.

Manders-Huits argues that VSD cannot provide a "methodological account for distinguishing genuine moral values from mere preferences, wishes, and whims of those involved in the design process" (2010, p.281). And without such a methodological account, VSD practitioners risk attending to an unprincipled or unbounded set of values. Albrechtslund (2007) points out that although VSD draws on ethical theory, it is not clear "*what* theories and *which* values this includes" (p.67). Borning and Muller (2012) argue that VSD practitioners, via the rhetorical move of using a disembodied voice, claim more authority and impartiality than is warranted in areas in which the practitioners' own normative assumptions may limit exactly those qualities of authority and impartiality.

Therefore, we argue, VSD practitioners should make use of an ethical theory complementary to the VSD method. Such an ethical theory can provide sources of justification and argumentation for moral claims and considerations, which are needed to make principled judgments, to attend to a set of bounded and principled values, and to legitimize value trade-offs during the design process.

---

[16] This chapter has originally been published as: Jacobs, N. and Huldtgren, A. (2018). Why Value Sensitive Design Needs Ethical Commitments. *Journal of Ethics and Information Technology*. Vol.23. No.1. pp.23-26. Alina Huldtgren contributed to this article by sharing her experiences on working together with designers and applying the Value Sensitive Design method in practice. Her sharing of experience helped me to identify the main challenges of VSD.

The chapter proceeds as follows: first, we argue that VSD practitioners' explicit use of an ethical theory to complement the VSD process can resolve the critique that the voice and values of researchers and designers engaging in VSD is often insufficiently explicit (Borning and Muller 2012).

Secondly, we turn to the much-debated topic of whether or not VSD can rely on a set of universal values (Borning and Muller 2012; Friedman et al. 2013). We argue that by relying too much on empirical studies of universal values, VSD runs the risk of committing the naturalistic fallacy, and that an explicit commitment by VSD practitioners to ethical theory can resolve this.

Thirdly, we turn to the central question of this paper: what kind of ethical theory is best suited to accompany VSD practitioners in the VSD process? We argue for a mid-level ethical theory, and discuss several desiderata that are necessary for such a mid-level ethical theory to successfully accompany VSD.

3.1. The Voice of the Researchers

Borning and Muller (2012) have pointed out that researchers and designers engaging in VSD often fail to make explicit their own part in the design process, including their values and normative commitments, and unintentionally claim (unjustified) moral authority or impartiality.

It is relevant, however, for participants in a VSD project, or readers of the VSD project results, to know what the relevant values of the researchers and designers are, as well as the values that are explicitly supported in the project. Making explicit what values underlie the development and resulting functionality of a certain technology can help to assess the design process, the future impact of a technology on its users, and help potential users to choose a technology matching their own values. The most evident way to do this is by indicating at the start of the VSD process what moral principles the practitioners are making use of and adhere to them throughout the design process.

3.2. Universal Values and the Naturalistic Fallacy

Friedman et al (2013) take the controversial position that certain values are universally held and that this fact can provide normative direction in design. Friedman et al. (2013) contend that this empirical proposition is grounded in a large body of psychological and anthropological data.

Whether or not this empirical claim is correct is not the crucial point here. What is crucial is that versions of VSD that are heavily informed by this notion run the risk of committing the naturalistic fallacy. Advocates of this approach implicitly assume that one will know what to do in a *normative sense*, once one knows empirically whether some values are universally held (Manders-Huits 2010). However, even if empirical data is able to show *that* people hold certain values, it is not able to say anything about whether people *should* hold certain values. Ethical theories can provide arguments why we should care about certain values, why certain values should be prioritized over others, and how to make value trade-offs in case of value conflicts.

We agree with Borning and Muller (2012) that VSD does not need to take a position on whether or not it is empirically true that certain values are universally held. More important is whether VSD can address questions concerning value prioritization in the case of value conflicts. We believe that listening to the voices of stakeholders is important to identify what values matter and their input may also provide directions for value trade-offs. However, as Manders-Huits (2010) points out, when the value stances of stakeholders are the normative input for the VSD analysis, then VSD risks conflating facts and values. To avoid the naturalistic fallacy, the normative input for the VSD analysis should be derived from ethical theory.

3.3. Ethical Theory

What *kind* of ethical theory is best suited to accompany VSD? First, there are ethical theories that work with a top-down approach: general precepts are 'applied' to particular cases. For instance, the principle of utility from classical utilitarian theory, or the categorical imperative from Kantian theory is applied to a practical problem. An often-made critique to such top-down approaches is that the abstract principles in classical ethical theories, like e.g., the utility principle or Kant's categorical imperative, are extensively indeterminate, i.e. the content of these principles are often too abstract to determine specific acts that should (not) be performed (Beauchamp and Childress 2013). Therefore, it's doubtful whether these abstract principles from classical ethical theories "can be fruitfully 'applied' directly, so as to yield univocal answers to complex problems of professional practice and public policy" (Arras 2016).

Secondly, there are bottom-up approaches in applied ethics. Such approaches emphasize that moral certitude (or the best approximation thereof) is to be found on the level of specific cases. Instead of using principles from ethical theory, one needs to look at

all the particularities of a case. One then compares these details with the details of so-called paradigm cases, on which we have settled moral judgments, using analogical reasoning to make a judgment (Arras 2016).

A critique to these bottom-up, casuistic approaches is that they seem to 'black-box' or obscure moral justification, making it "difficult to provide (an account of) public justification of moral judgments" (Van den Hoven 2008, p.54). Another point of critique is that proponents of bottom-up, casuistic approaches write as if paradigm cases speak for themselves or inform moral judgements by their facts alone, however, to move constructively from case to case, a recognized and morally relevant norm must connect the cases. "All analogical reasoning in casuistry requires a connecting norm to indicate that one sequence of events is morally like or unlike another sequence in relevant respects" (Beauchamp and Childress 2013, p.401). So understood, casuistry requires ethical principles after all in order to give analogical reasoning moral direction.

Thirdly, there are mid-level approaches that often consist of a cluster of pivotal moral principles. These principles function as an analytical framework, forming the starting point for applied ethics in a specific domain, e.g., biomedical ethics, environmental ethics, or ethics of technology. Mid-level principles function as general guidelines for the formulation of more specific rules and are usually regarded as generating 'prima facie' obligations, i.e., as an obligation that must be fulfilled unless it conflicts with an equal or stronger obligation (W. D. Ross 1930). In order to come to concrete action-guidance, principles must be specified. The process of specification consists of reducing the indeterminacy of abstract norms by narrowing the scope, which comes down to "spelling out where, when, why, how, by what means, to whom, or by whom the action is to be done or avoided" (Richardson 1990, p.289). When two rival but valid specifications of principles conflict with each other, then these conflicting specifications need to be balanced and weighed against each other (Beauchamp and Childress 2013).

A critique to such mid-level theories is that ranking, specification, and balancing can vary greatly among different people regarding a particular case (Gordon et al 2011), and that it is unclear how a justified decision between opposing principles in such a case is to be made.

Although mid-level approaches are not free from criticism, we believe that mid-level theory is the right kind of ethical theory to accompany VSD practitioners. Let us explicate.

First, mid-level approaches are convergent; i.e., differences on the highest level of moral theory most often converge at the level of mid-level action guiding principles. As James Sterba puts it: "traditional ethical theories, be they Aristotelian, Kantian, Millian, or whatever, have come to be revised and reformed in such a way that, at least in their most morally defensible formulations, they no longer differ in the practical requirements they endorse" (2005, p.1). For example, a utilitarian and a Kantian deontologist might differ at the foundational level of moral theory, these differences ultimately can recede at the level of mid-level moral principles where the utilitarian and the Kantian deontologist could agree, for instance, on the importance of the principle of respect for autonomy in a specific practice such as research ethics (Arras 2016). In a context where people with various disciplinary backgrounds, interests, and priorities have to work together, which often is the case in design-contexts, convergence on the practical level is crucial to come to joint decisions.

Secondly, mid-level principles can provide action-guidance in concrete cases, in contrast to general precepts from classical moral theories that are often too indeterminate to do so. Mid-level principles can provide action-guidance in practice because they are a) domain specific, that is, the selection of a cluster of principles often takes place after examining considered moral judgments and the way moral beliefs cohere in a certain context (Beauchamp and Childress 2013). And b) because through the methodological tool of specification, the indeterminacy of abstract norms is reduced and content is added to abstract principles, ridding them of their indeterminateness and providing action-guiding content for the purpose of coping with complex cases (Beauchamp and Childress 2013).

Now, we explore what conditions are necessary for a mid-level ethical theory to successfully accompany VSD. The list of desiderata we provide here is, however, not meant to be exhaustive.

3.4. Desiderata for a Mid-Level Ethical Theory

(1) Explanatory power: the theory should provide us with insight on the purpose and status of morality, and on how principles, rules and rights are related to obligations, and the like (Beauchamp and Childress 2013).

(2) Justificatory power: the theory should provide grounds of justification and argumentation for moral claims and considerations. These grounds of justification are necessary to make principled judgments and to legitimize value prioritizations.

(3) Simplicity and practicability: the practical requirements of the ethical theory should be simple, i.e.; the practical requirements should not be so demanding that they can only be satisfied by a small number of expert ethicists. We cannot expect everyone on the VSD team to possess (or acquire) adequate knowledge of ethical theory. This raises the question *who* should conduct the ethical analysis during the VSD process: ought the ethical reflection to be conducted by an ethicist who joins the design team, or by the designers themselves? This is a much-debated topic (Manders-Huits and Zimmer 2009; Van Wynsberghe and Robbins 2013) and goes beyond the scope of this paper. However, it is important that the ethical theory is as simple and practical as possible, so that a broad range of people with interdisciplinary backgrounds can understand and apply the theory to practice.

A mid-level ethical theory that meets the abovementioned desiderata is able to solve the various problems that VSD currently faces, as discussed at the beginning of this paper.

First, a mid-level approach can provide a solution to the problem of making explicit the voice of the researcher at least as well as top-down or bottom-up approaches can. Any given mid-level approach influencing the design process can be named and justified for use in this particular design context. Designers engaging in VSD thus acknowledge they are choosing amongst several different ethical approaches and have chosen one on the basis of normative considerations. However, they have not done so randomly or merely as a result of their own personal preferences, but in a way that is grounded in ethical literature and is uniquely appropriate to the domain.

Secondly, mid-level approaches, at least as well as top-down and bottom-up approaches, avoid risking the naturalistic fallacy because they have the potential to critically assess current practices, design choices, and arrangements from a moral point of view.

Thirdly, mid-level approaches are able to tell us why certain values should be prioritized over others and how to make value trade-offs in case of value conflicts, better than top-down or bottom-up approaches are able to. That is because the general precepts of top-down approaches are often too indeterminate to provide us with clear action guidance or decision-making content. And bottom-up approaches lack the ability to publically justify value prioritizations or value trade-offs because they obscure

the moral deliberation process by denying that moral principles play a role in moral deliberation at all.

Examples of mid-level ethical theories that meet the abovementioned desiderata are the theory of principlism by Beauchamp and Childress (2013), the Capability Approach (Robeyns 2017), or Martin Peterson's geometric account to moral principles (2017). Which of these mid-level theories is best suited to accompany designers in the VSD process is up to further debate.

3.5. Concluding Remarks

We believe that VSD practitioners should have the liberty to choose for themselves what ethical theory they make use of to complement the VSD analysis. However, we hope to have convincingly showed that a mid-level ethical theory that meets abovementioned desiderata is most suitable to do so.

4 Capability Sensitive Design for Health and Wellbeing Technologies[17]

In recent years, there has been an increasing awareness of the impact that technology design can have on supporting or undermining values.[18] This awareness that technology design is not value-neutral, but instead embodies moral choices, has led to the development of various design methods that explicitly pay attention to values and ethical considerations. One of the most prominent and influential methods is Value Sensitive Design (VSD): a design methodology that aims to address and account for values in a "principled and systematic manner throughout the technical design process" (Friedman and Hendry 2019, p.4). What is unique about VSD is that it proactively integrates ethics into technology design (Van den Hoven 2008).

Despite being a highly promising approach to ethics of technology design, VSD faces various challenges (for a detailed discussion of these challenges see: Borning and Muller 2012; Davis and Nathan 2015; Jacobs and Huldtgren 2018; Manders-Huits 2010). The three most prominent challenges that VSD faces include: (1) obscuring the voice of its practitioners and thereby claiming unfounded moral authority; (2) taking stakeholder values as leading values in the design process without questioning whether what *is* valued by stakeholders also *ought* to be valued; and (3) not being able to provide normative justification for making value trade-offs in the design process (Jacobs and Huldtgren 2018; Manders-Huits 2010). To overcome these challenges, Noemi Manders-Huits (2010) and myself and Alina Huldtgren (Jacobs and Huldtgren 2018) have argued that VSD practitioners need to complement VSD with an ethical theory.[19]

---

17 This chapter has originally been published as: Jacobs, N. (2020). Capability Sensitive Design for Health and Wellbeing Technologies. *Science and Engineering Ethics*. Vol. 26. No.6. pp.3363-3391.

18 Present-day examples of technology design in which values such as fairness or equality were undermined include racial discrimination in facial recognition software and algorithmic bias in recruiting software tools, while the app I Reveal My Attributes (IRMA) is an example of a software design that explicitly supports the value of privacy.

19 Batya Friedman and David Hendry indicate in their book *Value Sensitive Design, Shaping Technology With Moral Imagination* (2019) that "value sensitive design resists overarching normative directives, for example, about which ethical theory to adhere to" (p.8). The strength of VSD without adherence to an ethical theory, as an anonymous reviewer pointed out, partly lies in its simplicity, versatility and adaptability. However, as this chapter will hopefully convincingly show, there are various crucial advantages for VSD when complemented by an ethical theory. Friedman

The claim made by Manders-Huits (2010) and Jacobs and Huldtgren (2018) that VSD needs to be complemented by an ethical theory is accepted at face value in this paper. However, it is important to note that there may be other solutions to the challenges of VSD not requiring complementation of an ethical theory, like e.g., the 'Sen-procedural-VSD-approach' that Alessandra Cenci and Dylan Cawthorne (2020) argue for. However, this chapter focuses on providing the best possible case for and elaboration of complementing VSD with Martha Nussbaum's substantial capability theory[20] (2000; 2006; 2011), leaving the comparison with the Sen-procedural-VSD-approach (Cenci and Cawthorne 2020) to be dealt with in later work. By complementing VSD with Nussbaum's capability theory the chapter contributes to the further development of VSD in particular, and the domain of ethics of technology design in general.

Various scholars have explored the role of the capability approach (CA) for ethics of technology and for designing for values (e.g.: Cawthorne and Cenci 2019; Cenci and Cawthorne 2020; Coeckelbergh 2009, 2010; Frediani and Boano 2012; Haenssgen and Ariana 2017; Hancke 2016; Johnstone 2007; Mink, Parmar and Kandachar 2014; Murphy and Gardoni 2012a, 2012b; Nichols and Dong 2012; Oosterlaken 2013, 2015a; 2015b; Steen 2016; Zheng 2007, 2009). This chapter contributes to those efforts by providing a systematic investigation of how VSD's tripartite methodology can be combined with the normative foundation of Nussbaum's capability theory -resulting in 'Capability Sensitive Design' (CSD) - and providing insight and tools for designers and engineers to use to operationalize CSD.

Ilse Oosterlaken (2012; 2013; 2015a; 2015b) has taken on a similar endeavor, taking a capability approach to technology design and indicating this as 'Capability Sensitive Design'. However, Oosterlaken's endeavor differs from this present endeavor because Oosterlaken examines various approaches in the domain of 'designing for values' broadly conceived, while this present endeavor focuses on 'Value Sensitive Design' (VSD) (Friedman and Hendry 2019) explicitly.

CSD is intended to be useful for ethicists of technology in general, and for designers and engineers working on health & wellbeing technology in particular. CSD is particularly well suited to ethically evaluate technology design for health and wellbeing for

---

and Hendry seem to recognize this when they write that: "mechanisms exist through which normative elements can be embedded in VSD. For example, a particular ethical theory could be employed in a given project, as Aimee van Wynsberghe (2013) did in her application of care ethics to robotic assistance in health care" (2019, p.8).

20 The distinction between the capability *approach* and capability *theories* is elaborated on in section 4.1.

three reasons. Firstly, the primary focus of CSD on people's capabilities fits very well with the common aim of technology design to enhance and expand what people are able to be and do (Oosterlaken 2012; Van den Hoven 2012). Secondly, by putting focus on conversion factors, i.e., people's abilities to *converse* resources into capabilities, CSD is able to account for human diversity (Oosterlaken 2012; 2015a; Toboso 2010; 2011). Why this is important is elaborated on in more detail in section 4.2. Thirdly, CSD is particularly well suited to the domain of technology design for *health and well-being* because CSD aims to normatively assess technology design based on whether the design expands human capabilities that are identified as valuable. Lennart Nordenfelt (1986) and Sridhar Venkatapuram (2013) have proposed to conceptualize health as a person's ability to realize one's vital goals and to achieve or exercise a cluster of basic human activities. Given that we adhere to this conception of health, then CSD seems to be particularly suited to normatively assess technology designs for health and well-being.

The chapter proceeds as follows: first, the distinction between the capability *approach* and capability *theories* is elaborated on together with the core elements that all capability theories share. Secondly, the specifics of CSD are discussed. Thirdly, CSD is applied to a hypothetical design case of technology for health and wellbeing: the design of a therapy chatbot to help improve people's mental health. This basic outline of CSD illustrates (1) the general CSD framework (2) how CSD can be applied to a (hypothetical) health and wellbeing technology design such as a therapy chatbot, and (3) what the merits are of a CSD analysis over the standard VSD approach for health & wellbeing technology design. To conclude, the main challenges that CSD faces are discussed, followed by how CSD could address these.

4.1. Capability Approach and Capability Theory

Since the mid-1940s, the dominant way to measure the overall welfare of a country has been to look at Gross Domestic Product (GDP) or Gross National Product (GNP) per capita (Dickinson 2011). Amartya Sen (1985), however, famously challenged this dominant focus by illustrating empirically how deceiving GDP or GNP per capita can be as measures of wellbeing. Instead of looking at GDP or GNP to measure wellbeing, the economist and pioneer of the capability approach, argued that one should look at what people are able to do and be, and thus the kind of life that they are effectively able to lead, when assessing people's well-being. That is because quantities of or access

to resources or goods don't tell us much about what a person is actually able to be and do in one's life and the real opportunities available to a person (Sen 2009). A visually impaired and a visually non-impaired person, for example, would, in a given context need different amounts and types of resources to enable them to have the same opportunities in life. Thus, although resources are necessary *means* to wellbeing and freedom, we should look at the freedoms and opportunities people have available to them when assessing human wellbeing.

Before elaborating on some of the key characteristics of the capability approach, it is important to briefly clarify the distinction between the capability *approach* and capability *theories* as put forward by capability scholar Ingrid Robeyns (2017). Robeyns explains that the capability approach is an "open-ended and underspecified framework, which can be used for multiple purposes" (2017, p.29). As an open and general idea, the capability approach can be specified or 'closed' in many ways. When one specifies the approach and puts it to a specific use for a particular purpose, Robeyns proposes to speak of a capability *theory*. In short: there is one capability approach and there are many capability theories (Robeyns 2017).

My aim in this chapter is to specify the open-ended and underspecified framework and put it to use for a specific purpose, namely to ethically evaluate the design of health and wellbeing technologies. The capability *theory* that follows is called 'Capability Sensitive Design' (CSD) and consists of a combination of the method of VSD, combined with the core elements from the general capability approach and the normative foundation from the specific capability theory developed by Nussbaum (2000; 2006; 2011).

The purpose of CSD is to normatively evaluate the design of health and wellbeing technology, accounting specifically for human diversity and diminishing (structural) injustices in technology design. CSD is intended to be useful for ethicists of technology in general, and in particular for designers and engineers working on health & wellbeing technology who aim to design with moral sensitivity. Before elaborating further on the specific capability theory of CSD, let me briefly discuss the core elements that *all* capability theories share as put forward by Robeyns (2017).

Functionings and *capabilities* are the core concepts of every capability theory. Capabilities refer to people's freedoms and the valuable opportunities that one can choose from. Functionings refer to what people are *actually* achieving in terms of beings and doings. For example, the opportunity to travel is a capability, while actually travelling is a functioning. Thus, functionings are about the *actually realized*, while capabilities are

about the *effectively possible* (Robeyns 2017). There are functionings and capabilities with positive value, as well as negative value (e.g., committing murder). In themselves, functionings and capabilities are neutral concepts. It is subsequently a normative decision of the *specific* capability theory to decide *which* capabilities are defined as having positive value.

Another core element is the idea that it should not be taken for granted that resource provision leads to increased capabilities or functionings. Instead, the *conversion* of goods and services into what people are actually able to be and do is influenced by personal, social, and environmental *conversion factors*. Conversion factors "determine the degree to which a person can transform a resource into a functioning" (Robeyns 2017, p.45).

There are three different types of conversion factors: (1) *personal conversion factors* are internal to a person, such as metabolism, physical condition, sex, or intelligence. (2) *Social conversion factors* stem from the society in which one lives, such as public policies, social norms, societal hierarchies or power relations related to class, gender, race or caste. (3) *Environmental conversion factors* emerge from the physical or built environment that a person lives in (Robeyns 2017, p.46).

Think for example of a woman who buys a wearable fitnesstracker to help increase her capability of bodily health. This woman could have the personal conversion factor of having a sufficient physical condition to be able to walk and run and in that way use the fitnesstracker. She might also have the right environmental factors needed, such as having broad sidewalks and a park nearby to exercise in. But she might lack the social conversion factor needed if she lives in a neighborhood where it is unsafe for women to go out on their own. Differences in conversion factors form an important source of human diversity, which is a notion of great importance within the capability approach and CSD. The notion of human diversity is discussed in more detail in section 4.2.

Furthermore, it is important to always be clear whether one values something as an end in itself, or as *a means to* a valuable end. Within each capability theory, the capabilities that are defined as valuable are the ultimate ends. Subsequently, normative decisions are made based on the extent to which something -e.g. policies or design choices- promotes people's capability to achieve the functionings they value. However, functionings and capabilities are not necessarily the only dimensions of value, many capability theories may add additional dimensions of value, such as fairness. The fact that there might be other dimensions of value besides capabilities and functionings,

indicates *value pluralism*. Another dimension of value pluralism is the fact that there are multiple capabilities and functionings to be valued, rather than just one. A final core element of every capability theory is that every human being has equal moral worth (Robeyns 2017).

Now that the core elements of all capability theories as put forward by Robeyns (2017) are mentioned, it is important to note that these key insights have often been compared to (rival) ethical theories such as e.g., utilitarian views, the manifold ethical-political theories that pervade contemporary liberal-egalitarian thinking such as John Rawls' (1971) theory of justice, or other multidimensional accounts of wellbeing-related needs such as the account of Len Doyal and Ian Gough (1984). To discuss these comparisons in depth here would exceed the limits of this paper but I would like to refer to the works of e.g., Ian Gough (2014), Ingrid Robeyns (2009), Harry Brighouse and Ingrid Robeyns (2010) in which in depth comparisons are made.

4.2. Capability Sensitive Design

The abovementioned core elements are shared by all capability theories. It is now time to clarify the specifics of the capability theory of CSD. The specific purpose of CSD is to ethically evaluate the design of health and wellbeing technologies. CSD consists of a combination of the method of VSD, combined with the normative foundation from the specific capability theory developed by Nussbaum (2000; 2006; 2011).

The reason why Nussbaum's capability theory is chosen to complement VSD (and not for example Sen's procedural capability account (Sen 1992; 1999; 2009)) is because VSD faces various challenges that have to do with the fact that VSD is a procedural approach that does not make any substantive ethical commitments (Manders-Huits 2010; Jacobs and Huldtgren 2018). VSD needs complementation of a substantive ethical theory that provides grounds of justification and argumentation for moral claims and considerations (Jacobs and Huldtgren 2018) because without such substantive ethical commitments, VSD faces the challenges of (1) obscuring the voice of its practitioners and thereby claiming unfounded moral authority; (2) taking stakeholder values as leading values in the design process without questioning whether what *is* valued by stakeholders also *ought* to be valued; and (3) not being able to provide normative justification for making value trade-offs in the design process (Jacobs and Huldtgren 2018; Manders-Huits 2010). Nussbaum's capability theory provides such

needed substantive normative foundation by defending that all people are morally equal and deserve a life worth living, which entails that every human being should have access to ten central capabilities. By explicitly complementing VSD with this substantive normative foundation of Nussbaum's capability theory, the CSD framework is able to provide sources of justification and argumentation for moral claims and considerations, which are needed to make principled judgments, to attend to a set of bounded and principled values, and to avoid conflating facts with values (Jacobs and Huldtgren 2018). Nussbaum's capability theory helps VSD to overcome the challenge of obscuring the voice of its practitioners as well as the naturalistic fallacy, this will be elaborated on in further detail in the section 'Conceptual Investigation' of this chapter. Important to note, as already indicated in the introduction, is that e.g., Cenci and Cawthorne (2020) have presented a solution to the challenges of VSD not requiring complementation of a substantial ethical theory, arguing instead for a 'Sen-procedural-VSD-approach'. This chapter, however, leaves the comparison with the Sen-procedural-VSD-approach (Cenci and Cawthorne 2020) to be dealt with in later work and instead focuses on providing the best possible case for and elaboration of complementing VSD with Nussbaum's substantial capability theory. Let us now look further into the normative foundation of Nussbaum's capability theory.

Nussbaum has developed a partial theory of justice based on the moral assumption that all people are morally equal and deserve a life worth living. Nussbaum identifies ten capabilities that should be available to all human beings. Nussbaum defends these ten capabilities as being the moral entitlements of every human being. The ten capabilities are:

"(1) being able to live a normal length of lifespan;

(2) having good health;

(3) maintain bodily integrity;

(4) being able to use the senses, imagination, and think;

(5) having emotions and emotional attachments;

(6) possess practical reason to form a conception of the good;

(7) have social affiliations that are meaningful and respectful;

(8) express concern for other species;

(9) able to play;

(10) have control over one's material and political environment" (2000, p.33).

Important to note here is that Nussbaum's list of capabilities is often criticized as paternalistic or illegitimate with the reason that such a list of capabilities should be the outcome of a democratic process or a process of public reasoning (Cenci and Cawthorne 2020; Claassen 2011). Making a capability list as Nussbaum does -and upon which CSD relies- would bypass those people that its theory is to be applied to in practice, one could argue (Claassen 2011). A related critique is that a philosopher such as Nussbaum cannot possibly know which capabilities are important to people because there are epistemological limits to the philosopher's insights, and thus claiming false authority with a list (Claassen 2011). In response to these points of criticism, Nussbaum deliberately formulated the ten capabilities on her list at an abstract level, 'precisely in order to make room for the activities of specifying and deliberating by citizens' (Nussbaum 2000, p.79). She emphasizes that the translation of the capabilities into implementations, policies -or in this context: design requirements- should take into account differences on the specific, local level. The capabilities from Nussbaum's list thus need to be specified in a particular context. Thus, the CSD practitioners together with the stakeholders involved provide context-specific specification of the capabilities, adding content to the abstract capabilities on the open-ended list. Furthermore, Nussbaum emphasizes that her list of ten capabilities is not definite; instead, the list is open for discussion and revision (Nussbaum 2000, p. 77). Thus, Nussbaum's list offers a piece of input into the deliberation process, providing us with a stimulus for public debate, capability-scholar Rutger Claassen (2011, p.501) points out. Now, a preferred strategy for CSD practitioners to make use of Nussbaum's capability list is put forward by Claassen (2011) as the "philosopher-investigator" approach:

> "philosopher-investigators regularly cross the boundaries to gather data in the 'real world' and they may learn from public debates about capabilities, or they may even conduct social scientific research to find out which capabilities people value most. They will retain their right to construct their theories as they deem best; for after all, the fact that a group of people empirically holds the belief that realizing capability x is a moral demand does not mean that these beliefs are

> morally correct. However, with suitable modification, they will allow themselves to let the results of their practical investigations influence their theories and think the latter enriched by these efforts" (Claassen 2011, p.504-505).

A good example of this strategy is the research done by Jonathan Wolff and Avner de-Shalit, who used Nussbaum's capability list as the basis for in-depth interviews and then modified the list to account for the results of these interviews (Wolff and de-Shalit 2007). Thus, the philosopher-investigator's strategy provides an answer to the objection that Nussbaum's list of ten capabilities might be illegitimate, paternalistic or claiming false moral authority because the list is not the mere result of isolated reflection but is also informed by empirical study and stakeholder analysis, thus drawing upon the knowledge of (many) others. A practical tool to assist empirical stakeholder analysis could be the use of a set of "capability cards" as developed by Marc Steen (2016) which provides input to support CSD practitioners and stakeholders to discuss, select and specify the capabilities (from Nussbaum's list) on which to focus specifically in the design context at hand.

Nussbaum's list of ten capabilities forms the normative foundation of CSD together with two moral principles: respecting human diversity and diminishing (structural) injustices in technology design. Let me now briefly elaborate on these principles.

Human Diversity and (structural) injustices in technology design

Each individual has a unique profile of conversion factors, some of which are body-related, while others are shared with all people from one's community, and still others are shared with people with the same social characteristics, e.g., same gender, class, caste, age, or race characteristics (Robeyns 2017, p.113). Taking into account conversion factors -i.e., human diversity- in technology design is important in order to avoid potential unjust risks of harm for some user groups (Oosterlaken 2012).

Some technology is deliberately designed for a specific user group whose members share the same conversion factors, purposefully excluding users that do not share those conversion factors. This does not have to be problematic, think e.g., of the design of a bicycle for children that is designed to fit the size of boys and girls around the age of 4 to 6 years old. Most people younger than 4 years-old or older than 6 years will have body-related conversion factors that will not enable them to make use of the particular bicycle design. This is fairly unproblematic since the design is particularly intended to serve a specific user group (children between 4 and 6-years old), taking into account their specific conversion factors.

However, what I want to focus on here is a much more subtle difficulty concerning human diversity in technology design. That is the difficulty that, even though designers might not purposefully exclude users from their design, by designing technology that is in principle intended for a (large) user group but that not actually serves all members of that user group equally because certain conversion factors of people are not sufficiently taken into account, designers may increase a risk of harm for some users to which designers are (partially) responsible (Oosterlaken 2012). Let me discuss two examples.

Recently, the design of a soap dispenser in a public restroom caused astonishment when it turned out that the machine only worked on white skin (Lazzaro 2017). The designers of the soap dispenser failed to take into account that the infrared sensor of the dispenser could not detect darker skin tones. Although the designers did not purposefully exclude people with darker skin tones, the designers did not sufficiently consider the bodily conversion factor of people with darker skin tones, even though the soap dispenser in principle was designed for people of *all* skin tones. Failing to account for the bodily conversion factor of people with darker skin tones, the design has (unintendedly) put these users at an increased risk of harm.[21]

Another example includes the design of cars; in comparison to men, women tend to sit further forward when driving because women are on average shorter than men. Women's legs -on average- needs to be closer to reach the pedals and they need to sit more upright to see clearly over the dashboard. However, this is not the 'standard seating position' for which cars are designed. Research by Caroline Criado-Perez (2019) has shown that as a result, women are at greater risk of injury when involved in a car crash than men. Both of these examples show that although a technology is designed for a large user group (respectively including people of *all* skin tones and men *as well as* women) the technology design fails to sufficiently take into account the relevant conversion factors of all the people in that user group. With the result that these people (i.e., respectively people with darker skin tones and women) are not able to convert the technology into a mean (as is the case with the soap dispenser for people with darker skin tones) or are at an increased risk of injury when making use of the technology (women drivers). The failure to sufficiently account for the conversion factors of all the users in your intended user group is to a large extend a blame on part of designers, and it is their responsibility to avoid such injustices occurring in technology

[21] Harm is understood here as experiencing racial discrimination, and of course the harm of not being able to properly wash your hands.

design by paying more attention to human diversity, i.e., the relevant conversion factors of all intended users.

Catriona Mackenzie (2013) has drawn a strong connection between human diversity, capability deficits, and vulnerability. Mackenzie argues that specific capability deficits can signal sources of vulnerability, and vice versa, since "people's ability to convert the resources available to them into achieved functionings will vary according to [...] individual differences and external circumstances" (2013, p.50). For example, a woman who is a lesbian, may have the capacity for sexual expression but will not develop the capability to do so if she is a member of a community that strongly condemns same-sex relationships. Furthermore, Mackenzie argues that the notion of vulnerability "signals the actual or potential harm that may result from particular capability deficits and highlights the obligation to address these deficits in order to remediate vulnerability" (2013, p.50). The lesbian woman who does not have the capability to express her sexuality is vulnerable to a range of possible harms, such as social ostracism, a decrease in employment opportunities, or being a victim of homophobic violence. If we think that the capability of sexual expression is important[22], then this capability deficit and the woman's vulnerability to aforementioned harms are unjust and call for action to remediate her vulnerability, Mackenzie rightfully argues (2013, p.51).

Elsewhere I have argued that technology design should explicitly consider the needs and interests of vulnerable people (Jacobs 2020). That is because vulnerability is to be understood as a diminished capacity to meet or protect ones needs or safeguard one's interests, which could lead to an increased likelihood of suffering harm or wrong (Jacobs 2020).

Oosterlaken, who has worked extensively on the relation between technology and capabilities (2013), has pointed out that the key is to realize that if a technology is designed in such a way that for some (vulnerable) user groups the technology will not expand their capabilities, the technology design might be morally unjust (Oosterlaken

[22] Important to keep in mind, as rightfully pointed out by an anonymous reviewer, is that values may change; new values may emerge, the priority or relative importance of a value may change, the way in which a value is conceptualized, as well as how a value is specified may change over time and depending on context. In certain times e.g., homosexuality was (and in certain contexts still is) condemned. The value of sexual expression in those times and contexts has a different meaning, and the corresponding capability and functionings (to sexual expression) may not have not have been considered acceptable.

2013). CSD endorses the ten capabilities identified by Nussbaum (2000) as to have moral value and people should be brought to at least a threshold level of these capabilities to lead a dignified life. Now, if a technology design fails to bring a particular stakeholder group to the threshold level of one or more of these capabilities, then the technology design is not only inadequate but could also be morally unjust. In other words: CSD is able to signal whether there is a (structural) injustice at play in a technology design when a particular stakeholder group for whom the technology is (partly) intended is not being brought up to the threshold level of one or more capabilities that have been identified to have moral value in the particular design context.

Think for example of an architectural design of a town hall, which is a public space that is intended to be public for all, that is designed with a two-step stairway in front of the entry and a 3-inch-high doorsill at the entry door. For people who are in a wheelchair, the stairway and doorsill make the town hall inaccessible to them. For people who have difficulty walking, such as people who walk with crutches or elderly who walk with a walking stick, the town hall is difficult to access. For people who have no difficulty walking, the town hall is easily accessible. CSD endorses the ten capabilities from Nussbaum's list to have moral value to all people in order for them to live a dignified and worthy life. Now, one of the capabilities from Nussbaum's list that is relevant in this context is the capability of having control over one's environment and being able to effectively participate in political and social choices that govern one's life. We now see that people who have no difficulty walking can easily access the town hall, a place where political and social choices are made and where one can effectively enact on controlling one's environment. The design of the town hall thus enhances the capability of having control over one's environment for people who have no difficulty walking. For the people who are in need of walking assistance such as crutches or walking sticks, the town hall is more difficult to access and we see that for these people it is harder to reach the threshold level of being able to have control over one's environment for people. And for the people who are in a wheelchair, the town hall is entirely inaccessible which makes it impossible for these people to effectively govern their environment by ways of political or social choices that are made in the town hall. The design of the town hall thus fails to bring these people to the threshold level of the capability to control one's environment. This makes the design of the town hall not only inadequate, but also indicates there is an injustice at play in the design. That is because the design of the town hall fails to bring a particular vulnerable stakeholder group (i.e., people who experience difficulty with walking), for whom the town hall is also intended, to the threshold level of the capability of having control over one's

environment and being able to effectively participate in political and social choices that govern one's life, which is a capability that is identified to have moral value.

4.3. CSD and Health & Wellbeing Technology

Up to this point, the chapter has reflected on why CSD could be well suited to ethically evaluate technology design in general. In this section, I briefly elaborate on the reason why CSD is well suited to ethically assess technology design for *health and wellbeing*.

Health has been conceptualized in various ways, most prominently as: (1) absence of disease (Boorse 1975); as (2) a state of complete physical, mental, and social well-being (WHO 2006); as (3) the ability to adapt and self-manage in the face of social, physical, and emotional challenges (Huber and colleagues 2011); as (4) the ability to realize one's vital goals (Nordenfelt 1986); and as (5) a person's ability to achieve or exercise a cluster of basic human activities or capabilities (Venkatapuram 2013). Not only do we have various conceptions of health, we also have a variety of health practices. Such as the health practices of biomedical research, care for chronically ill patients, palliative care, reproductive medicine, public policies concerned with health inequalities, or the practice of designing technology to enhance health and wellbeing. Given the variety of the purposes of different health practices, these different practices may need different health concepts. Beatrijs Haverkamp, Bernice Bovenkerk and Marcel Verweij (2018) have argued that "it makes sense to take different health concepts as appropriate for guiding different practices" (p.382), instead of seeking one 'overall' conceptual theory of health that applies to *all* contexts and practices. Haverkamp, Bovenkerk and Verweij (2018) argue that if we want a health concept to be relevant for a certain health practice, such a health concept should guide the practice in "their formulation of and reflection on goals and priorities" (p.382).

The practice of technology design to enhance health and wellbeing can be appropriately guided by the conceptions of health as formulated by Nordenfelt (1986) and Venkatapuram (2013), since these conceptions focus respectively on a person's ability to realize one's vital goals and a person's ability to achieve or exercise a cluster of basic human activities or capabilities, which corresponds well with technology design that aims to enable people's abilities. Consider technology designs ranging from glasses, hearing aids and prostheses that enable and enhance people's abilities, to technology designs such as online applications that help people achieve or maintain a healthy lifestyle (e.g., MyFitnessPal or Fitbit), or manage their diabetes (MySugr) or their

alcohol consumption (Sobriety Counter) and thereby enable them to achieve their goals and abilities.

Now, given that the conceptions of health by Nordenfelt (1986) and Venkatapuram (2013) are most suited to guide the practice of technology design for health and wellbeing, then CSD -based on the core elements of the capability approach and the normative foundation of Nussbaum's capability theory in particular- seems to be particularly well suited to normatively assess such technology designs for health and wellbeing.

4.4. Applying Capability Sensitive Design

In what follows the workings of CSD are demonstrated. I propose that CSD follows the tripartite and iterative methodology of VSD, which entails a conceptual, empirical and technical investigation (Friedman and Hendry 2019). Each of these investigations is carried out iteratively, mutually informing and being informed by the other investigation. I will explicate what each of these investigations entails in CSD by discussing the example of a hypothetical design case of an AI-based therapy chatbot that aims to help people deal with mental health issues such as depression, dysfunctional thinking, or anxiety disorders. The following CSD outline illustrates: (1) the general workings of CSD, (2) what a new technology design for health and wellbeing such as an AI-based therapy chatbot would look like when attention is paid to capabilities right from the start of the design process, and (3) what the advantages are of a CSD analysis compared to the standard VSD approach.[23]

Designing a therapy chatbot for mental health

Data from the World Health Organization shows that "globally, more than 300 million people of all ages suffer from depression" (WHO 2019). While at the same time 33% of countries worldwide "allocate less than 1% of their total health budgets to mental health, with another 33% spending just 1% of their budgets on mental health", furthermore "there is only one psychiatrist per 100 000 people in over half the countries in the world" (WHO 2001). Partly in response to these alarming data, conversational 'therapy chatbots' are being introduced into the domain of clinical

[23] By no means is this a fully comprehensive CSD analysis of such a therapy chatbot, instead it is a basic outline of a CSD analysis in order to show the general workings of CSD and to illustrate the merits that CSD can have over VSD as a design approach, especially in the case of design for health and well-being technology.

psychology (Hoermann, McCabe, Milne and Calvo 2017; Kretzschmar, Tyroll, Pavarini, Manzini, Singh 2019).[24] The aim of these therapy chatbots is to assist people in dealing with mental health issues such as depression, dysfunctional thinking, or anxiety disorders.

One of the first conversational chatbots introduced for mental health was the system called ELIZA developed by Joseph Weizenbaum (1966) that "emulated a Rogerian psychological perspective and used an early form of AI to engage the user in a dialogue of question and answer and asked empathic questions, based on the previous dialogue move" (Calvo, Milne, Hussain and Christensen 2017). Present-day therapy chatbots, such as Woebot, Wysa and Youper, are now making use of natural language processing systems (NLPS). With machine learning algorithms for language processing, these therapy chatbots analyze and process large amounts of natural human language data, replicate patterns of human interactions and respond based on the knowledge database that is available to them at that point in time. Utilizing NLP techniques, these therapy chatbots deliver cognitive behavioral therapy (CBT) interventions to users. CBT is a therapy that modifies dysfunctional emotions, behaviors and thoughts. Online CBT-based interventions have been widely developed and evaluated effectively for the treatment of various mental health issues, such as depression, dysfunctional thinking, and eating or anxiety disorders (Barak and Grohol 2011; Calvo, Milne, Hussain and Christensen 2017; Kaltenthaler, Brazier, De Nigris, Tumur, Ferriter, Beverly and colleagues 2006). Online CBT interventions include for example daily 'check-ins' wherein the chatbot engages users in brief conversations asking how the user is feeling at that moment, or 'mood tracking' that enables users to track their moods with the help of emoticons or keywords over a period.

[24] One could point out -as an anonymous reviewer rightfully did- that this is essentially a policy problem that comes down to the question whether designers should be tasked to make cheaper mental health tools or whether policy makers should increase funding for mental health? The value sensitive-alternative approach developed by Sarah Spiekermann (2015) invites us to think up valuable alternatives to standard technology and calls us to forgo the use of ethically problematic systems and to actively seek ethical alternatives to create value. Although this is indeed a crucial issue to consider at the start of each design process, in the specific case of the design of a mental health chatbot there are significant benefits for users of the technology, such as its 24/7 availability at any place with an internet connection and its ability to avoid stigmatization attached to frequenting a human mental health care professional. Therefore, it can be deemed worthwhile to explore a CSD approach to the design of such a mental health chatbot.

The aim of therapy chatbots that are making use of CBT interventions is to change the attitude and behavior of users for them to better cope with their mental problems such as dysfunctional thinking, depression or anxiety disorders. These therapy chatbots can therefore be understood to fall in the category of behavior change technologies (BCTs). BCTs are technologies intentionally designed to induce behavior change in users, making use of various behavior change techniques. Such behavior change techniques include e.g. suggestion; intervening with a signal or message at the right time (Fogg 2003, p.41). For instance, if it becomes apparent trough the CBT intervention of mood tracking that a user is regularly feeling sad in the afternoons, then by intervening with the CBT intervention of a 'check-in' in the afternoon is the 'right time' to stir a change in behavior of the user. Another behavior change technique is self-monitoring: through self-tracking their behaviors and moods, users learn how well they are performing the target behavior, "increasing the likelihood that they will continue to produce the behavior" (Fogg 2003, p.44). A third behavior change technique that is often used is that of conditioning: by rewarding target behavior the user is enforced "to increase the instances of a behavior or to shape complex behaviors" (Fogg 2003, p.49). Woebot, for instance, sends users funny GIFs when they answer a question correctly.

Conceptual investigation

The aim of the conceptual analysis is to: (1) select the capabilities and corresponding functionings that are relevant in the particular design context, (2) get clear who the stakeholders are that are affected by the technology design, and subsequently (3) identify what the relevant conversion factors at play are for these stakeholders.

Nussbaum's list of ten capabilities forms the starting point for the selection of relevant capabilities in the particular design context at hand. A practical tool to assist CSD practitioners in selecting relevant capabilities from the list could be a set of "capability cards" as developed by Steen (2016) which provides input to support CSD practitioners and stakeholders to discuss, select and specify the capabilities from Nussbaum's list and help them decide upon which to focus specifically in the design context at hand.[25] The precise specification of the selected capabilities in the technology design will continue to evolve in the empirical and technical investigations, since CSD follows the

[25] Please recall that Nussbaum's list is open-ended; it is not definite but open for discussion and revision (Nussbaum 2000, p.77) thus it is possible that CSD practitioners and stakeholders agree to design for a capability that is not on the list.

integrative and iterative methodology of VSD wherein the three phases of investigation -conceptual, empirical and technical- mutually inform each other.[26]

In the conceptual investigation, CSD practitioners explore the central issues as well as big questions related to the design context at hand. In the case of designing a chatbot for mental health, one could think of big challenges such as whether chatbots might degrade human friendships since people might confide in technology rather than in the people around them once the chatbots are widely available, or whether the chatbots might facilitate a future situation in which only the rich will still have access to human health care professionals. With tools such as 'Envisioning Cards'[27] CSD practitioners together with stakeholders can reflect on the challenges attached to the design context and subsequently select the relevant capabilities to guide them in the design process. For example, the capability of having control over one's material and political environment together with the capability of having good health could guide CSD practitioners in designing a chatbot that presents us an alternative future situation than one in which only the rich have access to human health care professionals. Or the capability of having social affiliations with others could guide CSD practitioners in designing a chatbot that diminishes the risk that human friendships ultimately degrade.

Important to note here is that "design typically concerns products that do not yet exist" (Van de Poel 2012, p.296). This causes an epistemological challenge, as Ibo van de Poel has pointed out (2012). This entails that CSD practitioners need knowledge both of "what constitutes well-being for users and how that well-being might be affected by new technologies", as well as awareness "that such knowledge needs to be translated into, for example, design requirements, criteria or technical parameters that can guide the design process" (Van de Poel 2012, p.296). Knowledge of the first aspect, on what

---

[26] A risk rightfully pointed out by an anonymous reviewer is that once a selection of capabilities is made, CSD practitioners and stakeholders might overlook the importance of one or more capabilities that later on in the design process turn out to be of importance as well. This risk can be minimized, however, by following the iterative and integrative methodology of CSD that emphasizes the importance of moving back and forth between the conceptual, empirical and technical investigations and have findings from a later phase in the design process influence earlier findings and vice versa.

[27] The Envisioning Cards are a toolkit developed by (Friedman, Nathan, Kane and Lin 2011) that can be used for ideation, co-design, heuristic critique, evaluation, and other purposes (Friedman and Hendry 2019, p.85).

constitutes wellbeing for users, can be explored with the "philosopher-investigator" approach (Claassen 2011) as discussed in section 4.2 of this chapter, a strategy on how this knowledge can subsequently be translated into design requirements is presented in the 'technical investigation' when the concept of the 'capability hierarchy' (Van de Poel 2013) is discussed.

Another important challenge put forward by Ibo van de Poel (2012) is the problem of aggregation, which entails that if wellbeing constitutes plural and incommensurable capabilities or values then the challenge presents itself how to aggregate these capabilities into an overall measure of wellbeing. A possible solution to this challenge is recently put forward by Alessandra Cenci and Dylan Cawthorne (2020), who point out that what is necessary to circumvent the aggregation problem is "qualitative deliberative workshops based on a focus groups research design, since incommensurable values-capabilities as well as "aggregated" social/public value *naturally* emerge from small-sized and localized procedures based on face-to-face interpersonal critical discussion and deliberation" (Cenci and Cawthorne 2020, p.27). A study by Alessandra Cenci and Azhar Hussain (2019) points out, however, that "Nussbaum's rigid and over specified list" is less suitable to facilitate this than Sen's version of the capability approach with its "underlying procedural ethics ideal" (Cenci and Cawthorne 2020, p.28). However, with the "philosopher-investigator" approach (Claassen 2011) that CSD practitioners are recommended to follow, as discussed earlier in section 4.2 of this chapter, there is sufficient room for the necessary critical discussion and deliberation among practitioners and stakeholders.

Capabilities

Now, in the specific context of the design of an AI-based therapy chatbot to help improve people's mental health, the following capabilities from Nussbaum's list could be identified as relevant:

(1) The capability of being able to live in health, both physically as well as mentally. Health, as argued for in section 4.3, can best be understood in this context as a person's ability to realize one's vital goals and the ability to exercise a cluster of basic human activities.

(2) The capability of being able to experience human emotions. Being able to experience love, grieve, longing, gratitude and justified anger, and not have one's emotional

development blighted by overwhelming fear and anxiety, or by traumatic events of abuse or neglect.

(3) The capability of being able to have social affiliations with others. To have the social bases of self-respect and non-humiliation, and to engage in various forms of meaningful social interactions with other human beings.

(4) The capability of being able to enjoy life, to play, laugh and have fun.

(5) The capability of practical reason, to plan and give direction to one's life.

(6) The capability of having control over one's material and political environment.

Functionings

The functionings that correspond with the selected capabilities include the following: the functioning of actually being in good health, both physically as well as mentally, the functioning of actually having emotional attachments to others and not be overwhelmed by experiences of fear or anxiety that limit ones emotional development and capacities, the functioning of actually having meaningful and respectful social affiliations and engage with others in social interactions, the functioning of actually enjoying oneself and laugh and play, and the functioning of making plans and give actual direction to one's life.

Important to note is that a technology such as the therapy chatbot should foremost be *capability* enhancing and it should ultimately be up to the user to choose whether or not to actually turn a capability into a functioning. However, a technology can steer someone in a certain direction; persuading the user to engage with the technology in a certain way and perform certain target behavior. For persuasion by technology to be ethically permissible, as I have argued elsewhere (Jacobs 2020), it is important that persuasion never (1) significantly blocks or burdens options, that (2) a person is aware of the fact that one is being intentionally influenced, and aware of the mechanisms of that influence, and (3) that the influence is in the best interests and in alignment with the personal goals of the person being influenced. If these conditions are met, then it is ethically permissible that a technology steers users in a certain direction, aiming that they actually turn certain capabilities into functionings.

Direct and indirect stakeholders

In the conceptual investigation, CSD practitioners as well identify the relevant direct and indirect stakeholders involved. The direct stakeholders are those people who will interact directly with the technology or the technology's output. The indirect stakeholders are those people who won't interact with the technology directly but who might be impacted by it (Friedman and Hendry 2019). There are various methods developed in the empirical social sciences to identify direct and indirect stakeholders, it exceeds the scope of this chapter to discuss these methods in detail but I like to refer here to the work of Sarah Spiekermann (2015, p.174-175) on identifying stakeholders. Furthermore, the inputs and values of stakeholders can be gathered with the help of multiple empirical methods such as interviews, surveys, workshops or focus groups (Friedman and Hendry 2019; Cenci and Cawthorne 2020).

Direct stakeholders involved with the design of an AI-based therapy bot for mental health improvement are people dealing with mental health difficulties and are willing to make use of an AI-based therapy bot; these direct stakeholders are the end-users of the technology. Indirect stakeholders involved are e.g., health care professionals such as psychologists and psychiatrists who might have to deal with patients that are also making use of an AI-therapy bot besides their services, or with patients that initially replace their services because they turn to the therapy chatbot first. Other indirect stakeholders are: family and friends of people suffering from mental health issues that will make use of the therapy bot; health insurance companies that want to reduce the costs of regular mental health care; employers that want to prevent (long-term) sick leave or burn outs amongst employees and are therefore interested in the services of a AI-therapy bot for their employees; and governmental policy makers that are working on policies to improve public mental health and wellbeing.

Once the stakeholders are identified, they are engaged in the process of discussing, selecting and specifying the relevant capabilities and functionings for the design context at hand.

Conversion factors & human diversity

After identifying the direct and indirect stakeholders, CSD practitioners need to identify the conversion factors that are at play for these stakeholders. That is: practitioners need to get clear what the abilities of these various stakeholders are to convert *the means of the therapy chatbot* into *functionings*, since people have various abilities to transform a resource into a functioning (Robeyns 2017).

Due to limitations of space, the focus here will be solely on conversion factors of the direct stakeholders: the people who are dealing with mental health difficulties and are willing to make use of an AI-based therapy chatbot. The following conversion factors can be identified upfront as important to take into consideration: [28]

Personal conversion factors that are relevant include (1) the nature and severity of the mental illness that a person is suffering from, which might have influence on the person's receptiveness to (online) therapy; (2) whether or not, or to what degree, people are capable of expressing their (troubling) thoughts and feelings into words and write them down in conversation with a chatbot; (3) whether or not people are digitally literate enough to use a smartphone or computer and interact with an online chatbot.

Social conversion factors that are relevant include (1) whether or not there are societal norms such that people are comfortable enough to honestly speak about their (troubling) thoughts and emotions with someone, or some*thing* in the case of the therapy bot; (2) whether or not there exist social stigmas associated with certain mental illnesses that make it hard for people suffering from these conditions to speak about their experiences; (3) the size and quality of the social network that is available to a person, does the person has (many) relatives, friends, or caregivers to rely on? (4) Are certain people's voices not 'heard' or undermined on account of their illness, gender, sexuality, race, class or age, or because society lacks the concepts to name and acknowledge the (mental) symptoms someone is suffering from? In order words, are there epistemic injustices (Fricker 2007) at play towards certain (groups of) people in society?

An environmental conversion factor that is relevant includes whether there is a digital infrastructure such that people can easily make use of an online application such as an AI-therapy bot at places that they deem comfortable and private enough to use the chatbot.

By explicitly focusing on the varieties of stakeholders' conversion factors (i.e., their personal characteristics, social settings and environmental factors), CSD practitioners account for the diversity of stakeholders involved. And as already indicated previously, accounting for diversity is of crucial importance because when certain stakeholders

[28] By no means is this an exhaustive list of all conversion factors that should be taken into consideration since conversion factors depend greatly on context. Thus, more relevant conversion factors will be discovered in the empirical investigation when empirical research is being conducted among stakeholders and prototype testing takes place.

are not brought up to a certain threshold level of a capability that has been identified as morally relevant, due to the fact that the stakeholder group lacks the required conversion factor(s) needed to turn the recourse(s) of a technology into a functioning, then the technology design not only fails these stakeholders, but might also be morally unjust.

At this point, it is worthwhile to take a step back, and recall for a moment that VSD was facing various challenges when the method is *not* complemented by an ethical theory such as the capability theory by Nussbaum, as was discussed in the introduction of this chapter. The reason why it is worthwhile to recall this is because two out of the three challenges that VSD was facing without complementation of an ethical theory such as Nussbaum's capability theory, are at this point addressed.

First of all, there was the challenge that practitioners of VSD fail to make explicit their own voice in the design process, including their values and normative commitments, and thereby cause the risk that unfounded and unjustified moral authority is being claimed (Borning and Muller 2012; Jacobs and Huldtgren 2018). Although in the latest literature on VSD Batya Friedman and David Hendry (2019) state that "designers are encouraged to make their own values, as well as the project vales, explicit and transparent throughout the design process" (p.38) it remains unclear how designers should do this. However, by indicating at the start of the design process that the moral principles that CSD practitioners are making use of are derived from Nussbaum's capability theory, this problem is addressed. From the start of the CSD process it is now explicit that practitioners are guided by the moral principle that all human beings are equal and worthy of a life worth living -which entails that people have access to ten central capabilities- and that human diversity should be respected in technology design and (structural) injustices diminished. Now, when practitioners at the start of a new design process formulate their values in direct response to the normative input of CSD, then their voice and the normative commitments that are being made in the design process are explicit and transparent.

Secondly, VSD on its own faces the challenge of conflating facts and values. That is because the values solicited from stakeholders are taken as leading values in the VSD process without questioning whether what *is* valued by stakeholders, also *ought* to be valued (Jacobs and Huldtgren 2018; Manders-Huits 2010). VSD practitioners often assume to know what to do in a *normative sense* when knowing *empirically* what values stakeholders hold. But as myself and Huldtgren have pointed out: "even if empirical

data is able to show *that* people hold certain values, it is not able to say anything about whether people *should* hold certain values" (Jacobs and Huldtgren 2018). In order to avoid the conflation of facts with values -i.e., the naturalistic fallacy- the normative input for the VSD analysis should be derived from ethical theory, as argued for by Manders-Huits (2010) and Jacobs and Huldtgren (2018). CSD addresses this problem by deriving its normative input from Nussbaum's capability theory and her list of ten capabilities, providing CSD practitioners with normative input to argue what *ought* to be valued and thereby addressing the naturalistic fallacy. Important to note is, however, that CSD advocates the "philosopher-investigator" strategy as put forward by Claassen (2011) and discussed earlier in section 4.2. To summarize, this entails that CSD practitioners will let the normative input of Nussbaum's capability theory be influenced by the results of empirical stakeholder analyses. For example, by using the tool of "capability cards" (Steen 2016) practitioners discuss with stakeholders which capabilities from Nussbaum's list to focus on in the technology design project and how to specify these capabilities in the design context. How the role of stakeholders exactly takes shape is discussed in more detail in the following paragraph.

Finally, a challenge that to this point remains for VSD as well as for CSD is the challenge how to make trade-offs in the case of value- or capability conflicts. This challenge will be discussed in detail in section 4.5.

Empirical investigation

The empirical investigation explores if the findings of the conceptual phase correspond with the experiences and values of the direct and indirect stakeholders. With the help of various empirical methods such as interviews, surveys or focus groups (Friedman and Hendry 2019) CSD practitioners develop an understanding of how stakeholders are experiencing current provisions and services of mental health care, what stakeholders currently value and what they are missing, what their initial impressions are of a therapy bot, and what conversion factors play a role for them. Furthermore, in conversation with the stakeholders, CSD practitioners further specify the selected capabilities, adding context-specific content to the abstract capabilities. Subsequently, based on these findings CSD practitioners can make prototypes of the envisioned technology that incorporates the results of the conceptual and empirical investigation (how this works is explained in the next paragraph). Then, after the making of the first prototype(s) there is, ideally, a second empirical investigation conducted in which the prototype is presented to the stakeholders and their assessment of the prototype is explored, i.e.: whether the stakeholders experience if the technology design sufficiently

accounts for the selected capabilities, whether all relevant conversion factors are sufficiently accounted for enabling them to actually turn the technology into a functioning, and whether they experience any conflicts between valuable capabilities in the technology design. Based on these findings, the prototype is adjusted up to the point that it finds its 'ideal' form, to the extent that is feasible.

Technical investigation

How are the selected capabilities providing normative guidance in the actual design process, i.e.: in the *technical* part of the design phase? In other words: how do the selected capabilities affect, or even dictate, what tangible design requirements should be met in the design of an AI-based therapy chatbot to help improve people's mental health?

Van de Poel (2013) has developed the notion of *value hierarchy* in order to facilitate the translation of values into concrete design requirements for the VSD method. Following the workings of Van de Poel's value hierarchy, we can re-conceptualize the hierarchy as *capability hierarchy*, as proposed by Oosterlaken (2015a).

According to Van de Poel "a values hierarchy is a coherence structure that is held together by two relations" (2013, p.254). First there is the relation of 'specification' by which values are translated into norms, and norms into design requirements. Secondly, there is the relation of 'for the sake of' by which design requirements are connected to the higher-level elements of norms and values. That is: a design requirement is fulfilled 'for the sake' of the norm, which is the higher-level element that dictates the design requirement, and the norm is fulfilled 'for the sake of' the value or capability which forms the highest-level element in the hierarchy. A basic outline of a capability hierarchy looks like this:

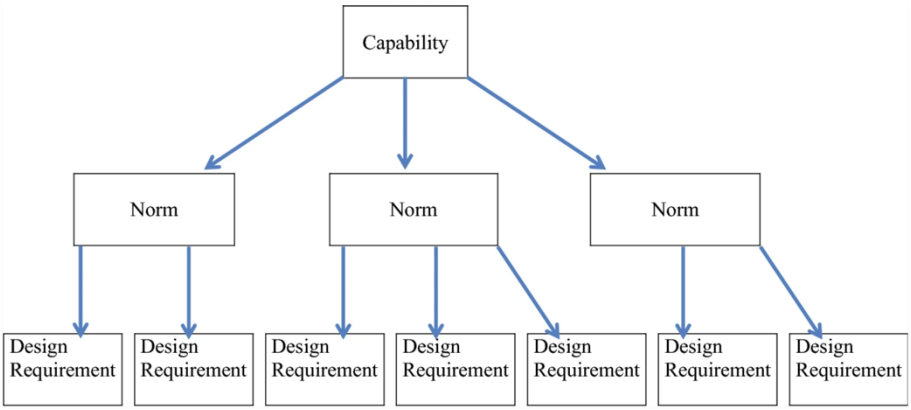

Figure 1. Basic outline of a capability hierarchy following the example of Van de Poel's (2013) value hierarchy.

Values, or in the case of CSD: capabilities, are relevant for evaluating the worth or goodness of certain things. However, values or capabilities do not directly imply certain prescriptions or restrictions for action, Van de Poel rightly points out (2013). Norms, contrary to abstract values or capabilities, do have a prescriptive character. Thus, the aim is to get from an abstract value or capability, to a more specific norm that articulates certain prescriptions for or restrictions on action. This translation is done with the methodological tool of specification, which is often used in mid-level theories. Specification adds context- or domain-specific content to abstract values or capabilities. There are two criteria for the adequacy of translating a value or capability into norms: the norms should be an appropriate response to the value or capability, and the norms should be sufficient to properly respond to or engage with the value or capability (Van de Poel 2013). For example: the capability of 'having social affiliations that are meaningful and respectful', which is identified as a relevant capability for the design of a therapy chatbot to help improve people's mental health, can be specified by adding context-specific content such as that people suffering from mental health problems often experience an absence of people they can talk to about their thoughts and emotions and who can provide them with emotional support. The addition of this context-specific information to the abstract capability of 'having social affiliations that are meaningful and respectful' helps us to formulate more specific norms that are appropriate and sufficient responses to the capability. Such norms could be: 'A person should be able to have confidential conversations' and 'A person should be able to speak about one's thoughts and emotions and be listened to'.

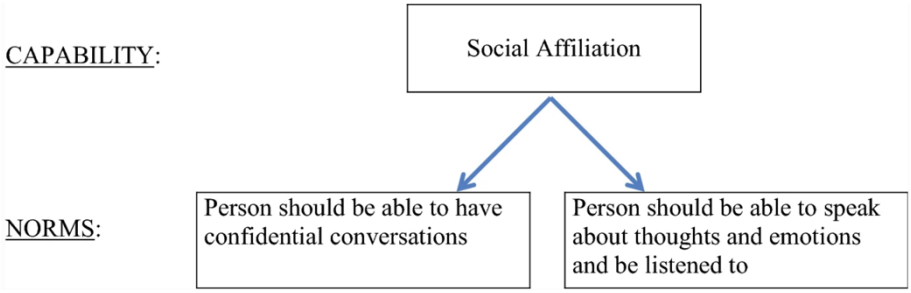

Figure 2. Detail of capability hierarchy

The next step is to translate these norms into concrete design requirements. The norm 'A person should be able to have confidential conversations' can e.g., be translated into the design requirement that the conversations between the user and the chatbot are end-to-end encrypted so that no one other than the user has excess to the data that is shared in the conversations, thereby assuring that the conversations are confidential and private.

A design requirement derived from the second norm 'A person should be able to speak about one's thoughts and emotions and be listened to' could be to program the chatbot in such a way that it encourages users with brief and daily 'check-ins' to ask how the user is doing and encourage them to take the time to tell their stories. For example, the chatbot Woebot sends out short daily messages asking users questions such as: 'tell me, what are you doing at the moment?' followed by an emoticon of a pencil. Or: 'got a second to reflect on what you're grateful for today?' These brief daily check-ins function as conversation starters and encourage users to take a moment to write down their thoughts and feelings.

Another design requirement derived from this norm could be that users can text 'SOS' when they want to talk to a human health care professional instead of talking to the AI-based chatbot in case of an alarming situation. Chatbots Woebot and Wysa direct users to international helplines when they type in 'SOS'. In addition, chatbot Wysa also offers users the option to subscribe for an additional fee to a personal coaching plan with a human coach.

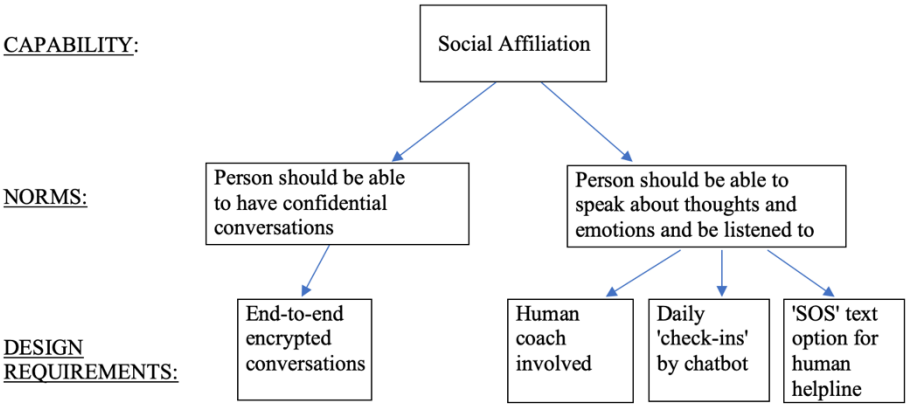

Figure 3. Detail of capability hierarchy

With the help of the capability hierarchy, CSD practitioners can translate the abstract capabilities into concrete design requirements. Subsequently, based on the identified design requirements, CSD practitioners then build (a) prototype(s) of the envisioned technology and empirically test with the stakeholders whether the prototype sufficiently facilitates the capabilities that are identified as important and whether the prototype accounts for the relevant conversion factors at stake for the stakeholders involved. Based on these empirical findings technical adjustments can be made.

4.5. Challenges for CSD

The CSD framework, however, faces some challenges, namely the challenge of (1) sufficiency, (2) dealing with capability conflicts, and (3) dealing with 'multistability'. I will briefly discuss these challenges and explain how they can best be dealt with.

A first challenge is how to determine whether a design ultimately meets the criteria of the capability hierarchy in a satisfactory way. That is: how can CSD practitioners determine whether they have accounted for all the norms derived from the selected capabilities, and subsequently whether they *sufficiently* accounted for all the design requirements that can be derived from these norms? This is a persistent issue that resurfaces in design approaches in general and there is no satisfying response to the issue of sufficiency. The most viable option for CSD is to retrace the capability hierarchy step by step and to move back and forth between the conceptual, empirical and technical investigations, making various iterations in consultation with the relevant

stakeholders to see whether the conditions of the capability hierarchy are sufficiently met.

A second challenge for CSD practitioners is how to deal with conflicts between capabilities in design contexts. That is: when two or more capabilities cannot be realized at the same time in a technology design, how should CSD practitioners prioritize one capability over the other?

For example, in the design case of the AI-driven therapy chatbot for mental health the following conflict between capabilities could occur: as discussed above, the capability of social affiliation is important in this design context. The capability of social affiliation can be translated in the norm 'a person should be able to have confidential conversations', which in turn could be translated into the design requirement that the conversations between the user and the chatbot are end-to-end encrypted so that no one other than the user has excess to the data that is shared in the conversations. However, the capability of health is also selected as an important capability in this design case and the capability of health can be translated into the norm that 'users suffering from mental health issues should be guided by health care professionals that have the expertise and skills to assist them'. This norm, subsequently, can be translated into the design requirement that the data shared by the user with the chatbot should also be monitored by qualified health care professionals in order for them to provide users with adequate help in cases of crisis.

Now, a conflict arises on the level of design options: one design option favors encryption of user data, while the other design option favors monitoring user data by health care professionals. Both design options refer to different capabilities: the design option of encryption refers to the capability of affiliation, while the design option of monitoring refers back to the capability of health. The question for CSD practitioners now is: what design option to favor and thus what capability to choose above the other?

Nussbaum has been very clear on the question, stating that each of the ten capabilities on her list is important and incommensurable, and thus cannot be traded off against another. In the case a capability conflict occurs and not all capabilities can be secured, "it becomes a purely practical question what to do next" Nussbaum has stated (2006,

p. 175).[29] In line with Nussbaum's position, Van de Poel has proposed "a solution to set thresholds for all relevant capabilities and to look for a design that reaches all these solutions" (Van de Poel in Oosterlaken 2015a, p.236). Let me elaborate on this solution. As the example above shows, a conflict occurs on the level of design options that in turn refer to norms, which refer to capabilities. Important to note is that the trade-off here should be made on the level of design options and this is thus the level to look for a solution. This is important to keep in mind since one can be convinced that a normative decision needs to be made on the level of capabilities, namely what capability needs to be traded off for the other, while a capability conflict can often "be tackled by means of technical innovations", Jeroen van den Hoven, Gert-Jan Lokhorst and Ibo van der Poel (2012, p.144) point out. Capability conflicts may very well have creative design solutions that enable us to expand *all* capabilities concerned rather than to make a trade-off between capabilities (Oosterlaken 2015a).

The conflict could be dealt with by various design solutions: think for instance of a design solution that ensures encrypted data-sharing while also offering users a 'red panic button' that users can press when they are in need of a human health professional to assist them in case of crisis. Another design solution could be to offer users encrypted data-sharing while also ask them at the start of using the chatbot to fill in an emergency contact that can be contacted in the case there is good reason to think the user poses a danger for themself or others. Yet another design solution could be to have users take a self-assessment test at the start of their use of the chatbot to determine the severity of their mental health issues and to determine whether or not the user is at an acceptable level to use the chatbot or whether a user's issues are so severe that one needs to be directed to a human health care professional. Thus, what CSD practitioners should do is come up with various design solutions to a capability conflict, which they subsequently pilot with the help of design prototypes among stakeholders in a second or third empirical iteration to find out what design solutions works best.
However, capability conflicts can present themselves also in different ways: not only can a capability conflict occur between two capabilities, there is also the possibility that a conflict occurs due to interpreting the same capability in two different ways. That is: various stakeholders could specify the same capability in different ways,

[29] As Robeyns (2016) has pointed out, "it is striking that very few proposals on selecting and weighing or aggregating have been worked out by philosophers based on foundational work on ethical theory." However, it exceeds the scope of this chapter to provide theoretical grounding that resolves the issue of conflicting capabilities.

seemingly leading to incompatible understandings of that capability and subsequently leading to different norms and design requirements that are derived from that understanding of the capability.

For example, as discussed earlier in this chapter, the capability of health can be specified in various ways. That is: the context and domain-specific content that can be added to the capability of health can differ greatly based upon what conception of health stakeholders adhere to. Thus, one stakeholder group, e.g., the designers of the therapy chatbot, could specify the capability of health using Nordenfelt's conception (1986) as people's ability to realize their vital goals. While the stakeholder group of health care professionals could specify the capability of health by Christopher Boorse's conception (1975) as statistical normality of functioning and the absence of disease.

The second stakeholder group, that adheres to the conception of the capability of health as the absence of disease, could derive from that conceptualization the norm that people should be able to live in absence of mental illnesses such as depression. While the first stakeholder group derives the norm that people should be able to undertake activities that are valuable to them, with or without the presence of a mental illness such as depression. Subsequently, the design requirements that follow from these different norms could point into conflicting directions, i.e.: either into the direction of curing disease, or into the direction of enabling people with (or without) disease to realize their abilities.

In such cases, it is important that CSD practitioners return to the process of specification, which forms the heart of the capability hierarchy. As already indicated, the process of specification consists of reducing the indeterminacy of an abstract capability into less abstract norms and subsequently into concrete design requirements. Reducing indeterminacy and narrowing the scope of capabilities and norms comes down to "spelling out where, when, why, how, by what means, to whom, or by whom" (Richardson 1990, p.289) an action is to be done or avoided. In jointly spelling out the where, when, why, how, by what means, to whom, or by whom something needs to be done or avoided, stakeholders that (initially) have different conceptions of a capability can come to an agreement via a process of deliberation on how to specify the abstract capability at hand, and subsequently into what norm(s) the capability can best be translated. In the case that differences remain on how to specify a capability, agreement can be sought on the level of norms and ultimately on the level of design requirements by looking for technical solutions that expand all conceptualizations of the capability and facilitate the various normative prescriptions that are identified.

Thus, by going through the process of specification that forms the core of the capability hierarchy, stakeholders are jointly deliberating on the relevant context- and domain-specific content that should be added to the abstract capabilities, with the aim to ultimately come to a joint solution on the design level. And thus again; capability conflicts (arising either between two capabilities or within the specification of one and the same capability by different stakeholders) may very well have creative design solutions that enable us to expand *all* (conceptions of) capabilities concerned, rather than to make a trade-off.

A last perpetual challenge for technology assessment in general is how to deal with unforeseen consequences of technology design. Anders Albrechtslund (2007) calls this the 'positivist problem', pointing out that in many design approaches the default position is "that the design of a technology will -more or less- correspond with the use of technology and that this relation does not pose a problem" (2007, p.68). However, the relation between design contexts and user's practice is often complex and unpredictable: technologies are frequently used in ways different than initially intended and technologies "can be conceived of differently according to cultural, historical and social contexts" (2007, p.68). Recent research, for example, has shown that people are "more compliant when a robot asks them to do something as compared with a person" (Fiske et al 2019, p.6). This could lead to the unforeseen and unwanted consequence that people share or do things that they otherwise would (and possibly should) not do. In the case of therapy chatbots, this could lead to the situation of people (over-)sharing personal and sensitive information without them properly realizing that the information is stored and used by the company behind the chatbot, and possibly even shared with third parties.

Don Ihde (1993) has pointed out that a defining characteristic of technology is its 'multistability'. This means that technology does not have a pre-existing essence or basic meaning apart from the use contexts it enters into. This defining characteristic of technology is what Ihde calls 'multistability'. The challenge design approaches in general face, and thus also CSD, is how to acknowledge and account for the multistability of technology. There is a fundamental openness implied in the concept of multistability, Albrechtslund (2007) points out, and this creates a difficulty for technology designers and developers to come up with a comprehensive list of potential human-technology relations and use contexts. The challenge is to "imagine potential use contexts and the ethical scenarios they create [...] and to envision as many multistabilities as possible

while designing technology in order to anticipate future ethical problems and dilemmas" (Albrechtslund 2007, p.70). What CSD practitioners should do in order to meet the challenge of the positivist problem is to acknowledge technology's multistability and the unpredictability of contexts of use, while at the same time account for active stimulation of "creative thinking to imagine the near-unimaginable" (Albrechtslund 2007, p.70) in all three investigations of the CSD process.

4.6. Concluding Remarks

This chapter presented the framework of Capability Sensitive Design, which consists of a merging of VSD with Nussbaum's capability theory. Various reasons have been given for why CSD can contribute to the normative assessment of technology design in general, and to technology design particularly for health and wellbeing. Subsequently, the general workings of CSD have been demonstrated by applying it to a hypothetical design case of a therapy chatbot. The chapter has shown how CSD addresses VSD's challenge that the voice of practitioners is obscured and thereby causing the risk that unfounded moral authority is being claimed, since the normative input in CSD is explicitly derived from the core claims of Nussbaum's capability theory and her list of ten capabilities. Furthermore, CSD has shown to be able to address VSD's challenge that stakeholder values are taken as leading values in the design process without questioning whether what *is* valued by stakeholders also *ought* to be valued, since within CSD it is normatively argued for that what *ought* to be valued are the selected capabilities based upon the normative core claims of Nussbaum's capability theory. Although the challenge of conflicting capabilities or values remains for both VSD and CSD, a strategy has been presented for how CSD could deal with occurring capability conflicts. To conclude, this chapter has presented the framework of CSD and illustrated its general workings, its merits over the standard VSD approach, and most importantly: what a new technology design for health and wellbeing such as a therapy chatbot would look like when attention is paid to capabilities right from the start of the design process. Now it is time for future empirical work on the application of CSD to determine its merits and limits in practice.

5 Bridging the Theory-Practice Gap: Design-Experts on Capability Sensitive Design

Naomi Jacobs and Wijnand IJsselsteijn[30]

Designers and engineers love to create new things, and by creating new things they are giving shape - in small or large ways - to the world we live in. Examples of technological innovations that have impacted the world and our daily lives are endless and their scope seems all-encompassing; ranging from the innovation of the steam engine, to household appliances such as the vacuum cleaner or smart home devices like Alexa and Siri, to the innovation of glasses, the MRI scanner, and AI-driven chatbots for mental health support. These are just a *very few* examples of technological innovations that have impacted the world we live in. Nowadays, many people are aware of the far-reaching impacts that technologies have on their daily lives. Lately, the awareness that designers and engineers can *actively shape* the societal and ethical effects of their designs seems to have increased as well (see e.g., literature on designing for the future by Reeves, Goulden and Dingwall 2016; Mazé 2016. As well as literature on designing for values by e.g., Flanagan and Nissenbaum 2014; Friedman, Kahn and Borning 2013; Winkler and Spiekermann 2018; Friedman and Hendry 2019). Many of the choices that designers make during design processes impact not only the functionality, usability or aesthetics of a technology, but also reflect and impact the values that might be represented, supported or undermined via the technology design. Designers can actively *design for values*. This increasing awareness has led to the development of multiple 'ethics by design' approaches (see e.g., Van den Hoven, Vermaas and Van de

---

30 This chapter has originally been published as: Jacobs, N. and IJsselsteijn, W. (2021). Bridging the Theory-Practice Gap: Design-Experts on Capability Sensitive Design. *International Journal of Technoethics*. Vol.12. No.2. pp.1-16. This article is based on a joint idea by Wijnand IJsselsteijn and me. IJsselsteijn contributed to the outline of interview questions that was used for the semi-structured interviews. I interviewed the design-experts, transcribed and coded the interviews and then together with IJsselsteijn made an analysis of the coded data and collated it into potential themes. Together we discussed and refined the potential themes in order to specify the overall story that the analysis tells. The final analysis has been written down by me and is presented in this chapter.

Poel 2015). These ethics by design approaches meet a growing need in the design and engineering community: a need for practical advice on how to consciously take into account the ethical implications of innovations during the design process.

The most prominent ethics by design approach is Value Sensitive Design (VSD): a design methodology that aims to address and account for values in a "principled and systematic manner throughout the technical design process" (Friedman and Hendry 2019, p.4; Friedman, Kahn and Borning 2013). However, despite being a highly promising approach to ethics by design, VSD has been criticized for not providing enough practical guidance in actual design processes; as well as for taking stakeholder values as leading values in design processes without questioning whether these stakeholder values *should* be valued; and for not being able to provide normative justification for making value prioritizations and trade-offs in the design process (Jacobs 2020; Jacobs and Huldtgren 2018; Borning and Muller 2012; Davis and Nathan 2015; Manders-Huits 2010).

Now, although there are some very promising examples of VSD applications (see e.g., Iversen et al (2020); Maathuis et al 2019; Oosterlaken 2014), the challenges that VSD currently faces make it a rather difficult approach to apply to concrete design processes. In the literature on VSD, various authors have argued that, in order to overcome its challenges, VSD should be complemented by an ethical theory (Jacobs 2020; Cenci and Cawthorne 2020; Jacobs and Huldtgren 2018; Manders-Huits 2010). Jacobs and Huldtgren (2018) have argued that VSD can best be complemented by a mid-level ethical theory, because mid-level action guiding principles are able to converge differences that occur on the highest level of moral theory, and because these mid-level principles are able to give action-guidance in concrete cases. The thesis that VSD should be complemented by a mid-level ethical theory, as argued for by Jacobs and Huldtgren (2018), is accepted at face value in this chapter.

In recent years, multiple scholars have set out to complement VSD with various mid-level ethical theories. Examples include Van Wynsberghe who has complemented VSD with an account of care ethics (Van Wynsberghe 2013) and Cawthorne and Van Wynsberghe (2019) who have complemented VSD with the theory of principlism. Furthermore, Cenci and Cawthorne (2020) aimed to refine VSD with a capability-based procedural ethics approach, and Jacobs (2020) has merged VSD with Martha Nussbaum's capability theory. Now, an in-depth analysis of which mid-level theory complements VSD "best" would exceed the scope of this chapter. Moreover, we hold

the opinion that it depends on the context of application which mid-level theory is best to complement VSD. That is, in the context of technology design for care robots, e.g., the theory of care ethics seems a very suitable companion to complement VSD (Van Wynsberghe 2013), while in the context of design for health and wellbeing technologies, Nussbaum's capability theory appears particularly well-suited (Jacobs 2020).

The context wherein the authors of this chapter operate, is the context of technology design for health, wellbeing and behavior change. As argued by Jacobs (2020), it is the capability approach (CA) (Sen 1985; 1992; 1999; 2009; Nussbaum 2000; 2006; 2011) that is especially well-equipped to ethically assess technology design for health and wellbeing. In short, that is because the CA has as its prime aim to expand people's capabilities, which corresponds particularly well with Lennart Nordenfelt's influential conceptualization of health as a person's ability to realize one's vital goals (Nordenfelt, 1986). As well as with the later refinement of that definition by Sridhar Venkatapuram (2013) of health as a person's ability to achieve or exercise a cluster of basic human activities or capabilities. Given these influential conceptualizations of health, the CA seems well suited to normatively assess health and wellbeing technology designs.[31]

The CA is an "open-ended and underspecified framework, which can be used for multiple purposes" (Robeyns 2017, p.29). When one specifies the approach and applies it for use with a particular purpose, we should speak of a capability *theory* (Robeyns 2017). What all capability theories have in common is that they understand wellbeing not in terms of the number of resources a person has, but instead by looking at what people are able to do and be, and thus the kind of life that they are effectively able to lead (Sen 1985; 1992; 1999; 2009). That is because quantities of, or access to, resources or goods don't tell us much about what a person is actually able to be and do in one's life and the real opportunities available to a person (Sen 2009). The freedoms and valuable opportunities that a person can choose from, are called *capabilities*. What people are *actually* achieving in terms of beings and doings, are referred to as *functionings*. Furthermore, it should not be taken for granted that resource provision leads to increased capabilities or functionings. Instead, the *conversion* of goods and services into what people are actually able to be and do is influenced by personal, social, and environmental *conversion factors* that determine the extent to which a person can transform a resource into a functioning (Robeyns 2017, p.45).

[31] For a more detailed discussion on why the CA is especially suited for ethical evaluation of health and wellbeing technologies, please see chapter 4.

This chapter takes 'Capability Sensitive Design' (CSD) under review. CSD is a merging of VSD with Martha Nussbaum's capability theory (2000; 2006; 2011). Oosterlaken (2013) has analyzed the applicability and potential added value of Nussbaum's capability approach for the design of technical artifacts, which she dubbed 'Capability Sensitive Design' (CSD). This work on CSD was recently refined by Jacobs (2020), by explicitly merging VSD (Friedman and Hendry 2019) with Nussbaum's capability theory (Nussbaum 2000; 2006; 2011).

Now, why have we chosen to review CSD and not the complementation of VSD with another capability theory, such as the capability-based procedural ethics approach by Cenci and Cawthorne (2020), for example? The main reason we chose to reflect on CSD, is because Nussbaum's capability theory clearly and explicitly indicates a substantive normative foundation.[32] This substantive normative foundation entails that all people are morally equal and deserve a life worth living, which entails that there are ten capabilities which are entitlements for every human being. In CSD, Nussbaum's list of ten capabilities forms the normative foundation together with the two moral principles of respecting human diversity and diminishing (structural) injustices in technology design (Jacobs 2020). As argued for by Jacobs and Huldtgren (2018) and Jacobs (2020), VSD needs complementation of a substantive ethical theory that provides grounds of justification and argumentation for moral claims and considerations, because without such substantive ethical commitments, VSD faces the challenges of obscuring the voice of its practitioners and thereby claiming unfounded moral authority; taking stakeholder values as leading values in the design process without questioning whether what *is* valued by stakeholders also *ought* to be valued; and not being able to provide normative justification for making value prioritizations and trade-offs in the design process. By explicitly complementing VSD with this substantive normative foundation of Nussbaum's capability theory, the CSD framework is able to provide sources of justification and argumentation for moral claims and considerations, which are needed to make principled judgments, to attend to a set of bounded and principled values, and to avoid conflating facts with values. The capability-based procedural ethics approach by Cenci and Cawthorne (2020) is -just like VSD- a procedural approach that doesn't make substantive ethical commitments, and therefore is not able to meet the challenges that VSD faces, like CSD is able to (Jacobs 2020).

[32] Of course, we must also emphasize here that one of the authors is also a developer of CSD (Jacobs 2020).

As mentioned, CSD takes Nussbaum's list of ten capabilities as its normative foundation, together with the moral principles of respecting human diversity and diminishing (structural) injustices in technology design (Jacobs 2020). The ten capabilities on Nussbaum's list include:

> "being able to live a normal length of lifespan; having good health; maintain bodily integrity; being able to use the senses, imagination, and think; having emotions and emotional attachments; possess practical reason to form a conception of the good; have social affiliations that are meaningful and respectful; express concern for other species; being able to play; have control over one's material and political environment" (Nussbaum 2000, p.33).

The aim of CSD is to proactively pay attention to the expansion of one or more of these capabilities from the start of a design process. CSD provides a clear normative foundation for design practices, with practical ethical guidance on what capabilities matter and why, and how these capabilities can be expanded in concrete technology design.

The added value for designers adopting CSD, is that (1) it enables them to overcome the abovementioned challenges that VSD faces, as well as that (2) CSD provides them normative guidance and justification in designing for justice and diversity in design processes, and (3) CSD provides designers with Nussbaum's list of capabilities that functions as input -as an 'awareness tool'- into the design process, directly engaging designers with the ethical implications of their design.

However, while CSD may be a promising approach to ethics by design, it has so-far only been developed from a theoretical point of view (Oosterlaken 2013; Jacobs 2020). We have yet to explore whether the framework is able to bridge the 'theory-practice gap', that is, whether CSD is practically applicable and of value in actual design practice. The current paper explores the perceptions of design-experts with regard to CSD and in relation to their own design practices. In this way, we aim to critically assess the extent to which CSD may be successful in bridging the theory-practice gap, and are seeking ways in which to improve the value of CSD to designers. To this end, the authors have conducted interviews with various design-experts to explore what these design-experts think of designing for values, what methods -if any- they use for incorporating values in their current design practices, what they regard as the strengths and weaknesses of CSD, and, most importantly, whether they think CSD could be of practical use to their design (research) practices.

5.1. Methodology

For this study, nine design-experts were interviewed. The selection of these nine design-experts was based on having a balance between experts on design *practice* and design *theory*, having mid-career experts as well as senior experts, and including experts with knowledge on health technology and/or behavior change theory, since the CSD framework was initially developed for health-related behavior change technologies (Jacobs 2020). Six experts were female and three were male. The backgrounds of the experts range from a mid-career designer at a medium-sized design agency, to a senior design-researcher at a health technology company, to a professor in design. Seven out of the nine experts were explicitly working on design (theory) for behavior change and seven out of nine experts have worked on technology design for health and wellbeing.

Furthermore, seven design experts explicitly indicated that they conduct stakeholder analyses or employ other forms of end-user involvement including user-centered or participatory design methods during design (research) trajectories (see Bekker & Long 2000). All design-experts were working in the Netherlands at the time of the interviews. The interviews were conducted 'one-on-one' and lasted approximately one hour. The interviews took place in the spring of 2020, during the Covid-19 pandemic, and were therefore all conducted online, via teleconferencing software. The interviews were semi-structured, which means there was an outline of interview questions (i.e., the interview guide) that formed the basis for the interviews with all the experts. However, each interview was flexible to an extent in order to adapt to the conversation and probe for underlying motivations. Between interviews, some gradual changes were introduced to gain more information on particular topics of interest, whereas less fruitful or saturated avenues of questions were truncated. The interview started with short introductions. Subsequently, the interviewer introduced the purpose of the interview, asked whether audio recordings could be made of the interview, and obtained informed consent. The first set of questions were warmup questions, where participants were asked to talk about their design research and experience. The interview guide furthermore addressed the following topics: encountering ethical dilemmas in design practice, actively designing for values, ethics in design education, and CSD and its practical applicability.

The data gathered from the interviews are thematically analyzed by the authors. Thematic analysis is a method for identifying, analyzing and reporting themes within the gathered data (Braun and Clarke 2006). The semi-structured interviews were recorded, transcribed, and then coded by the authors, after which the codes were collated

into potential themes. The authors then extensively discussed and refined the potential themes in order to specify the overall story that the analysis tells. The final analysis is presented in this chapter.

Important to note is that the author who conducted the interviews is also one of the people who developed the CSD framework (Jacobs 2020). The design-experts were aware that the author was also the developer of CSD and that the author stands favorably towards CSD. The design-experts were also aware that the author was theoretically committed to CSD, as well as that the author believed in the added value of ethics by design approaches and encouraged mutual exchange between the domains of ethics and design. Having said this, the author did explicitly invite constructive criticisms that would allow a critical assessment of CSD.

The main objective of the study was to let the design-experts reflect upon the question whether the CSD framework could be of practical value in their design (research) practice. At the time of the interviews, CSD was a new framework within the field of ethics by design. The design-experts did not have any prior knowledge or experience with the CSD framework, nor were they familiar with Nussbaum's capability theory. None of the participants in the study had reflected on the CSD framework prior to the interview, and were introduced to CSD for the first time by the interviewer.

The way in which the design-experts were introduced to CSD was first by a very brief introduction into the basic rationale of Nussbaum's capability theory (2000; 2006; 2011), as also mentioned in the introduction of this chapter. The experts were then introduced to specific aspects of the framework, namely Nussbaum's list of ten capabilities, followed by an excerpt of the capability hierarchy (Van de Poel 2013; Oosterlaken 2015; Jacobs 2020) which is a tool to help translate the abstract capabilities into tangible design requirements.

As a general reflection on the conversations with the design-experts, all design-experts easily engaged with questions on the social impact technology design can have, on the role that values play in design practices, and on the necessity for designers and engineers to have ethical awareness in their design practices. When it came to introducing the various aspects of CSD; all experts were responsive to the CSD framework. There were sometimes short pauses in the conversations where the design-experts had to take a moment to reflect on how aspects of the framework might relate to certain aspects

of their own practices. Next, we will present a detailed analysis of the main themes that emerged from the interviews with the design-experts. These themes are characterized as three interconnected questions: (1) 'Can We Design for Values?', (2) 'Can We Apply CSD Flexibly?', and (3) 'How to Give Guidance and Avoid (Too Many) Restrictions?'. After discussing these themes in detail, the chapter concludes with a discussion on the insights drawn from the themes in relation to the question how CSD can be of practical value to design practice.

5.2. Can We Design for Values?

All of the interviewed design-experts showed themselves aware of the far-reaching impact that technologies have on society and daily life. All design-experts furthermore showed an increasing awareness of the role designers have in shaping the societal impacts of their technology designs. One design-expert remarked:

> "*It is absolutely necessary that designers start to think about and reflect upon the effect that their designs might have on people, society and wellbeing; on the bigger picture.*" – director at a university-wide design lab.

Many of the design-experts (five out of nine) indicated that they have worked on design projects wherein they explicitly designed for one or more ethical values. Examples mentioned by the design-experts include e.g., a large design project for ambient assisted living that aimed at increasing the value of social connectedness among the people who would eventually be using the ambient assisted living technology. Other examples include the design of a smart bedside lamp that aimed to increase the values of creativity and social empowerment in interaction with the user, and the design of a digital platform that aims to increase the autonomy of people who experience mental health difficulties. The majority of the design-experts easily recalled one or several design projects they had worked on that explicitly aimed to increase one or more ethical values by means of the technology design, without explicitly indicating that they were conducting some form of *ethics by design* (Van den Hoven, Vermaas and Van de Poel 2015).

Six out of nine of the design-experts was explicitly asked whether they had ever encountered an ethical dilemma during a design (research) project, and all six participants who were asked the question could recall encountering one or more ethical dilemmas in their design practices. An example of an ethical dilemma encountered by a designer working at a medium-large design agency revolved around the question

how much autonomous decision-making power people with serious mental health difficulties should be given by the digital platform the design agency was designing, since *too much* autonomous decision-making power could potentially cause these people to harm themselves or others. An assistant professor in industrial engineering working on a digital platform for people to exercise more recalled encountering the ethical dilemma of whether or not they could 'trick' users of the platform with false ratings in order to persuade those users to try harder and exercise more. As a third example, an ethical dilemma mentioned by both a psychologist working at a health technology company, as well as by an associate professor in design, is the dilemma of what authority a team of designers and researchers has to determine what a healthy lifestyle consist of and what is 'healthy' and 'good' for other people. One of them remarked:

> *"Who are we […] to prescribe to others what is healthy or what is good for them?"* – associate professor in design.

What the above seems to show is that the interviewed design-experts are aware that technology design can have social and ethical implications. In addition, many of the experts indicate that designers can deliberately support certain values in a technology design process. Many of the experts indicate to have actively supported one or more ethical values in a design process. Without, however, explicitly making use of an existing approach or framework for 'ethics by design' (Van den Hoven, Vermaas and Van de Poel 2015). Furthermore, the experts that were asked ever encountering an ethical dilemma in their design (research) practice could all recall one and showed themselves aware of the fact that such ethical dilemmas form a part of design processes. Although none of the experts indicated to be explicitly working with an approach for 'ethics by design', and none of the experts used an ethical theory or an ethical framework in their design (research) practices, all of the experts showed themselves aware of, and sensitive to the ethical aspects of design practice.

5.3. Can We Apply CSD Flexibly?

After having discussed various ethical aspects of design practice, the experts were given a very brief introduction into the basic rationale of Nussbaum's capability theory (2000; 2006; 2011). None of the experts were familiar with Nussbaum's capability theory prior to the interview. After the brief introduction, the design-experts were asked whether they thought if 'thinking in terms of expanding capabilities' would be

applicable to their design (research) practice? All nine design-experts indicated that thinking in terms of 'expanding capabilities' fits well with their views on design practice, saying:

> "*It's kind of a 'direct match' in a way*" – professor in design.

> "*It is an exciting idea*" - designer working at a medium-sized design agency.

However, when the design-experts were introduced to the list of ten capabilities developed by Nussbaum (2000) that forms the basis of CSD (Jacobs 2020), the reactions were a lot more diverse. The design-experts were shown the list of ten capabilities and were asked whether they think this list could form a starting point for a design process. The initial reactions of the design-experts ranged from recognition, to pointing out the demanding and intellectual character of the list, to raising the issue of whether or not the list is (sufficiently) complete, and to questioning who has the authority to decide what capabilities are -and should be- on the list.

For example, a designer working at a medium-sized design agency recognized many of the ten capabilities from a similar list that she works with at the design agency[33], a senior design-researcher at a health technology company noticed that the list of ten capabilities is familiar to a list of twelve human needs and values that she uses at her company, and an associate professor in design recognized elements from behavior change frameworks that she makes use of in her design research practice. However, what she also noted is that the list of ten capabilities is highly intellectual and demanding, and that it requires a lot of people if you want them to be capable of doing all these things. Another expert responded that the capabilities on the list sound good, but questioned where the authority comes from to decide what capabilities should be on the list.

> "*Who decides who decides what capabilities are actually on the list and how the capabilities should be interpreted? Who decides what a 'normal length of lifespan is*" - professor in design research.

The director of a university-wide design lab made a similar remark, questioning who decides what having 'good' health actually entails? The same expert then remarked that the list of ten capabilities is perhaps too paternalistic, because it prescribes what

[33] This is the taxonomy of human goals by Ford & Nichols (1987).

capabilities are good for people and constitute their wellbeing. Perhaps people themselves thinks otherwise, the expert remarked.

> "*I think some of the capabilities on the list are too paternalistic…saying 'this is good for you', while people themselves might think otherwise.*" - director of a university-wide design lab.

In addition, the same expert pointed out that some of the capabilities on the list are formulated in a way that might be way too abstract and difficult to understand for most people. This can make it difficult to engage these people in reflection on the capabilities, the director of a university-wide design lab remarked.

A professor in design pointed out that the list of capabilities could form a nice lens to reflect on a design process, but also indicated that the list doesn't seem suitable as a starting point for a design process.

> "*I wouldn't start a design from that. But perhaps use it to take a step back and reflect on it.*" – professor in design.

Furthermore, the same expert questioned whether the list is complete, i.e., whether there are capabilities that are not on the list that should be. The expert then quickly added that 'completeness' should never be the aim of designers in a design process. A psychologist at a health technology company also questioned whether the list of ten capabilities is complete and covers all relevant capabilities. Furthermore, she questioned whether all of the ten capabilities are as relevant in every design process.

> "*The list of ten capabilities could form a starting point. However, if it would cover everything… and if it would always be these ten capabilities…I find it hard to say. It also depends, say if you look at the health solutions that I work on, helping chronically ill patients to manage their condition and their health, well than these type of capabilities are really important. But when you design, I don't know, a shaver… […] than it might be a bit* over the top." - psychologist at a health technology company.

The issues raised by some of the design experts on the paternalistic, intellectual and elitist character of Nussbaum's list, as well as the issue of authority and the question of completeness of the list, mirror some of the main criticisms that Nussbaum's capability list has received in scholarly literature (see e.g., Claassen 2011). How Nussbaum has responded to these criticisms, and how CSD deals with these criticisms, is discussed in Jacobs (2020).

After presenting the design-experts with the list of ten capabilities, the experts were introduced to the tool of the 'capability hierarchy' (Van de Poel 2013; Oosterlaken 2015; Jacobs 2020)[34]. The capability hierarchy is a tool to help translate the abstract capabilities into tangible design requirements. The capability hierarchy aims to assist designers to translate an abstract capability first into prescriptive norms and from there into concrete design requirements. The following excerpt from a capability hierarchy derived from Jacobs (2020) was presented to the participants. The excerpt shows a detail of the capability hierarchy applied to a hypothetical design case of an AI-driven mental health chatbot:

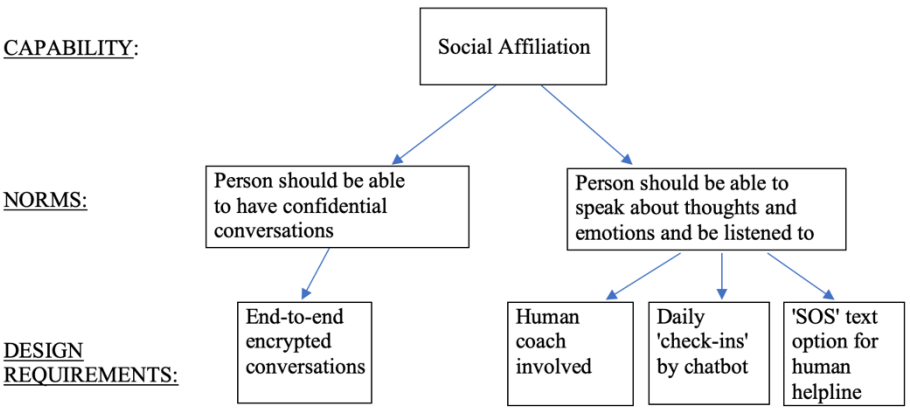

Figure 1. Detail of capability hierarchy (Jacobs 2020)

Even though most design-experts felt that the capability hierarchy can go some ways in bridging the gap between abstract capabilities and more concrete ethical norms, they were more skeptical with regard to the direct applicability of the capability hierarchy in a design process. The experts indicated that the capability hierarchy remains a very abstract tool that doesn't directly match with their own design practices or design tools used. For example, one expert remarked that:

> "*Such a capability hierarchy can help designers on their way, to make that translation*" […] *but this is not typically how designers work*" - associate professor in design.

Another expert indicated that a tool such as the capability hierarchy can be very useful in order to make a translation from abstract capabilities to concrete design

[34] The capability hierarchy is based on Ibo van de Poel's (2013) 'value hierarchy'. The concept of the capability hierarchy was first proposed by Oosterlaken (2015) and developed by Jacobs (2020).

requirements, even stating that such a tool is "*a must*". However, the expert questions how norms will eventually resonate with the needs of end-users.

A designer at a medium-sized design agency indicated to find the capability hierarchy an interesting tool, but pointed out that it is a very abstract tool that requires a lot of design skills in order to be applied satisfactorily.

> "*Such a capability hierarchy looks like an interesting translation process*" […] "*But I think it does require considerable design qualities to translate those abstract capabilities into norms. I think that's a tough step.*" - designer at a medium-sized design agency.

Many other experts indicated as well that the capability hierarchy remains an abstract tool that doesn't directly match to their own design practices. Many experts seemed to find it difficult to see how the capability hierarchy could provide concrete guidance in a design process. For example, an assistant professor in human-technology interaction mentioned that the translation from norms to design requirements still seems very complex.

> "*I find it very complicated.*" […] "*I find it difficult to know where those norms exactly come from.?*" - assistant professor in human-technology interaction.

A professor in design remarked that there is again -as with the list of ten capabilities- a matter of exhaustiveness and completeness. That is, how do designers know that they have mapped all the norms that correspond with a certain capability? A similar remark was made by another expert, who wondered how a designer can guarantee to have executed the hierarchy completely and exhaustively?

> "*Are these all the norms, or are there also many other norms? And where do these norms exactly come from? […] And then the design requirements […] is this the right design requirement? Or are there more? Where do they come from and who decides that it is those? And in addition: how can you guarantee that the norm is met when you fulfill these design requirements?*" – professor in design research.

In reaction to the capability hierarchy, many of the design-experts raised questions concerning the completeness and exhaustiveness of the capability hierarchy, and many emphasized that the translational steps between capabilities, norms and design requirements require considerable design qualities. Furthermore, another issue that was stressed by various experts in response to the capability hierarchy concerned iteration. Many of the experts emphasized the importance of iteration in design processes and in tools such as the capability hierarchy. A professor in design research

e.g., stated that design processes should always enable to critically re-examine all of the choices made thus-far, and possibly even come back to choices made earlier in the design process if it turns out that something doesn't quite work out as planned or works out *even better* than planned. Such re-examination and iteration should take place on all levels of the design process and is a crucial aspect of a design process. Another expert argued along similar lines and remarked that the subdivision of the capability hierarchy looks really good, but questioned where the required iteration would come in.

> "*How does it account for iteration? Now it seems as if you move from top to bottom. […] But you can't go from norms to design requirements and 'there you are'. You have to continuously move back and forth.*" - director of a university-wide design lab.

Another expert pointed out the importance of flexibility in design practices, and indicated that it is important to apply the capability hierarchy in a flexible manner.

> "*The most important thing is to be flexible about the moment you introduce this into the design process. When you look at design processes, they can come about in many ways and an idea may develop in various ways until you say 'this is what we are going to make'. It now seems that the capability hierarchy tool is mainly theoretical and to be applied when you don't really know yet what you will design, so you start thinking about what design requirements you could have. But that isn't always how it works of course. […] A design doesn't always start with a blank piece of paper, so when will you introduce this hierarchy then?*" - associate professor in design.

5.4. How to Give Guidance but Avoid (Too Many) Restrictions?

During the interviews, various design-experts brought up the issue of making value prioritizations and value trade-offs in design processes, as well as how to deal with value dynamics and with value change. The experts addressed that these complicated ethical issues are highly relevant in the context of design practice, but at the same time indicated that they have no clear-cut guidelines on how to make such difficult choices concerning value prioritization or trade-offs, or how to deal with value dynamics and change in practice. The experts pointed out that at first glance, CSD doesn't seem to provide clear guidelines on how to deal with these issues. One of the experts explicitly indicated it as a weakness of CSD that it doesn't seem to provide clear guidance on the issue of value/capability prioritization, stating that:

> "*The fact that you can't clearly prioritize one capability over another makes it very hard to use*" - assistant professor in human-technology interaction.

Another expert wondered how the various capabilities on Nussbaum's list relate to each other, and what effect that would have on iterative design processes.

> "*I wonder, do these capabilities also have effect on each other? How do I find out if a capability has effect on other capabilities in the design process and how do I deal with that in an iterative design process? Value dynamics… changing values, how does that work with capabilities, are they independent of each other?*" - associate professor in design.

Both design-experts seemed to indicate that it would be desirable if CSD provided clear ethical guidance on these issues; i.e., indicating how designers should prioritize between capabilities in a design process, as well as how designers should deal with capability dynamics in an iterative design process. However, this need for direction and guidelines on how to make value/capability prioritizations and how to deal with value dynamics and value change in design processes seems to be somewhat at odds with the emphasis that almost all design-experts put on the importance of flexibility, and context-dependent interpretation and decision-making for design processes during the interviews. One expert e.g., clearly stated that designers should be made aware of the issues of value/capability prioritizations, trade-offs, value change and dynamics, but should not be prescribed what to do in cases such an issue occurs.

> "*It is more important that you make designers aware [of value change, capability dynamics and prioritization etc.] and that they have to find ways to deal with that, than that you would prescribe designers how they should do that*" - associate professor in design.

What the interviews with the design-experts seemed to lay bare is somewhat of a tension between a need for clear ethical guidelines in design processes on the one hand, and the fear of interference with the necessary flexibility, iteration, context-adaptability and creative freedom that are crucial to successful design processes on the other hand. In relation to the CSD framework, many of the design-experts seemed to regard the list of ten capabilities and the capability hierarchy-tool as welcome and inspiring ethical guidance, however they were also seen as somewhat rigid aspects of the CSD framework that might restrict, or even form a threat to, the flexible and iterative process that is necessary for design practice. This tension seems to be well-captured in the following quote by one of the design-experts:

> "*The ways in which people value things changes continuously, you have to anticipate that if you want your innovations for the future to make sense.*" […] "*However, the fact that the list by Nussbaum is always very stable makes it a good starting point for a design process*" - senior design researcher at a health technology company.

5.5. Discussion

The authors have entered into dialogue with various design-experts in order to explore designers' take on ethics by design, the strengths and weaknesses of CSD, and the practical usefulness of CSD to their design (research) practices. With this study, by entering into dialogue with design-experts, the chapter aims to contribute to bridging the 'theory-practice gap' between ethics of technology and design practice.

One insight from of the study is that all of the experts showed themselves aware of, and sensitive to, the ethical aspects of design practice. All experts indicated to have worked on at least one design (research) project wherein one or more ethical values were actively supported in the design process. None of the experts, however, made use of an ethical framework in order to actively support ethical values in their design (research) practices. Nevertheless, the interviews made clear that a genuine need for some sort of practical ethical guidance exists in current design practice. Practical guidelines or directions on how to account for the ethical implications of innovation in a design process seemed to be welcomed by the experts, specifically with regards to making value/capability prioritizations in a design process, and how to deal with value dynamics and value change.

A second insight from the study is the importance all experts attributed to flexibility, iteration, and context-adaptability for a successful design process. All experts stressed how crucially important an iterative and flexible process is for a satisfactory design result. This, however, seemed to be somewhat at odds with the previous outcome, i.e.; designers' need for clear ethical guidelines in design processes. There exists a tension between the need for ethical guidance on the one hand, and stressing the importance of flexibility and adaptability on the other hand. This tension played out clearly in relation to certain aspects of the CSD framework. With regards to CSD, many of the design-experts seemed to understand the list of ten capabilities and the capability hierarchy-tool as a welcome and inspiring form of ethical guidance for a design process, however, they also understood these aspects to be somewhat rigid. Many of the experts seemed to think that the list of capabilities and the capability hierarchy tool might restrict, or even threaten, the flexible and iterative process that is necessary for design practice.

It seems appropriate at this point to clarify the role of the list of ten capabilities and the capability hierarchy within CSD. The list of ten capabilities within CSD is

intended as an adaptive, dynamic and open-ended list that foremost functions as an 'awareness tool' and as input into the design process to enable the design team to engage with the ethical implications of their design. The list of ten capabilities should thus not be interpreted as static, but instead as a dynamic tool between the design team and the object or system that is being designed. However, considering the feedback from the design-experts, one might wonder whether Amartya Sen's procedural account of the CA (Sen 1985; 1992; 1999; 2009) would not be more useful to designers than Nussbaum's capability theory? As often discussed in the scholarly literature on the CA (see e.g., Robeyns 2017), one of the main differences between Sen's procedural account and Nussbaum's capability theory, is that Sen doesn't provide a list of capabilities, but instead leaves the selection of capabilities open to the deliberation process. Nussbaum's list -in contrast to Sen's open, procedural approach- is then often presented in the literature as being static and definite (for this discussion, see e.g., Claassen 2011. For a recent endeavor wherein VSD is complemented with Sen's procedural capability account, see Cenci and Cawthorne 2020). However, Nussbaum has always emphasized that her list is open for discussion and revision (Nussbaum 2000, p. 77) and therefore is not to be understood as static nor definite. Capability-scholar Claassen (2011) e.g., has argued that Nussbaum's list can best be understood as a piece of input into the deliberation process. In the context of design i.e., as a stimulus for debate in the design process.

Taking the feedback from the design-experts into account, it becomes apparent that the CSD framework should emphasize more clearly that Nussbaum's list is intended as an adaptive, dynamic and open-ended list that foremost functions as input into the design process to enable the design team to engage with the ethical implications of their design. As pointed out by Jacobs (2020), Nussbaum deliberately formulated the ten capabilities on her list at an abstract level, precisely in order to make room for the activities of specification and deliberation (Nussbaum 2000, p.79) by citizens, stakeholders and designers. Designers, together with the various stakeholders involved, need to provide context-specific specification of the capabilities, adding content to the abstract capabilities on the open-ended list (Jacobs 2020). A practical tool to assist this endeavor is the set of "capability cards" developed by Marc Steen (2016) which provides input to designers and stakeholders to discuss, select and specify the capabilities from Nussbaum's list on which to focus specifically in the design context at hand (Jacobs 2020). In this way, Nussbaum's list of ten capabilities thus offers a piece of input into the deliberation process, providing us with a stimulus for the debate in design process (Claassen 2011; Jacobs 2020).

With regard to the capability hierarchy, its primary aim is to function as a tool to help designers become aware of the relationship between capabilities, norms and design requirements, and subsequently assist them to make the translational steps between the three dimensions. Designers should be free, however, to creatively adapt the tool as they deem necessary and implement the tool in a flexible way that gives way to an iterative design process.

A related outcome of the study was that various designers indicated a need for ethical guidance on how to make value prioritizations in a design process, and how to deal with value dynamics and value change. Nussbaum has argued that the capabilities on her list are all equally important and incommensurable (Nussbaum 2006). In the case a capability conflict occurs, Nussbaum has stated that "it becomes a purely practical question what to do next" (2006, p. 175); a practical question that depends on the specificities of the context at hand. Important to keep in mind is that Nussbaum's list is intended for setting the basic requirements of a society that allows for all humans to live a dignified life. When CSD places Nussbaum's list in the context of design, the list functions as an instrument to assure that designers expand capabilities via their technology design that contribute to people's wellbeing. There is no reason, however, to suppose that a technology design should promote *every* capability that is on the list, since that would not be feasible. Instead, when applying CSD, a design team can choose one or more capabilities from the list (and if needed, even add capabilities that are not yet on the list) that they will design for in the specific design context at hand. The design team should choose which of the chosen capabilities they expect to be most relevant to the design context, and which ones will be most likely less relevant. Of course, the relevance of capabilities may change during the design process due to findings from e.g., stakeholder analyses, and designers should be able to be flexible on the matter. Also, some of the chosen capabilities might turn out to be very explicitly endorsed in the design, while others may be more implicitly present in the eventual design. In light of the occurrence of a capability conflict in design, it is with context-specific weighing and balancing that a design team should decide which capability to prioritize in the design over another. Although it might not be feasible for a design to support *all* capabilities on the list, designers should aim to never seriously undermine one of the capabilities.

What significance do these study's outcomes have for design practice and ethics by design in general, and the further development of CSD in particular? Most

importantly, the interviews with the design-experts laid bare that CSD's systematic and orderly ethical guidelines are at odds with the nature of the design process, which is iterative, interactive, flexible, non-linear, and messy at times. It therefore seems important to clearly distinguish between (1) the process of design and (2) the outcome of that process (the actual design) and its justification. While the process of design is iterative, interactive, flexible, non-linear, and messy at times, the justification of the actual design can be, on the other hand, systematic, orderly, reflective and could be conducted in terms of adherence to rules and guidelines.

From the interviews, CSD seems primarily relevant for the justification of the final design; to provide designers with tools to increase awareness of the relevant ethical issues that could be at stake in the design, to reflect on the various steps and decisions taken so-far in the design process, and in the final justification of the design in which designers reflect upon to what extent their design lives up to the design requirements and how capabilities and values are -or could be- affected by the design, both positively or negatively. More precisely; CSD endorses the ten capabilities identified by Nussbaum (2000) as to have moral value and the list forms the normative starting point of a design process. Designers, together with multiple stakeholders and with help of e.g., the "capability cards" (Steen 2016) discuss, select and specify the capabilities from Nussbaum's list on which to focus in the specific design context at hand. Decisions made in the design process that enable the expansion of these capabilities in the final design are normatively justified. At the same time, *if* the technology design at a certain point in the process would *fail* to expand the selected capabilities for (a part of) the intended user group, CSD signals that there is an injustice at play and the designers need to question and revise their design choices.

CSD appeared to be not so relevant for steering the design process, given the iterative, interactive, flexible, non-linear, and sometimes messy nature of that process. However, referring back to the first-mentioned insight of this study that all experts showed themselves aware of, and sensitive to, the ethical aspects of design practice, it could very well be that CSD enables designers to increase this awareness and sensitivity. That is: during the design process, CSD can help raise designers' awareness and sensitivity to ethical tensions and choices that need to be made related to values and capabilities in the design process. This increased sensitivity brought about by CSD would then accompany designers throughout the process. CSD might thus not be so relevant for *steering* the design process, it could very well *accompany* designers in the messy, iterative and interactive process of design.

Lastly, a study of this kind has a number of inherent limitations. First of all, the perspective of the interviewer was not neutral, as one of the authors developed CSD (Jacobs 2020). We are aware of this bias, and have aimed to present CSD as objectively as possible, and have explicitly invited critical commentary and reflections from the design-experts. Nevertheless, given the nature of polite conversation, the known favorable position of the authors towards CSD may have given rise to more positive evaluations of CSD than an interviewer with a more neutral or negatively inclined stance would have otherwise elicited. Secondly, we are also aware that, although the sample of experts was intentionally drawn to be representative of different levels of experience as well as different design perspectives (e.g., academic versus industry versus design lab), our sample is limited in that all participants were employed in the Netherlands at the time of the interviews (although not all are Dutch natives), and most respondents represent a predominantly Western-European background. It would be very interesting to extend our work in the direction of cross-cultural comparison of design practices, as CSD may resonate differently across design cultures. Almost all experts had a focus on health technology and were experts in design for behavior change; this specific focus might have had effect on the experts' positions towards the perceived importance of designing for values, since the domains of health and behavior change are domains with significant known implications for the people affected by the designs. Indeed, the experts' focus on health technology and design for behavior change might have had an effect on their attitudes towards (the relevance of) CSD and the type of values and capabilities they understood as important.

As a reflection on method, the use of the interview method and subsequent thematic analysis allowed us to go into depth and probe topics interactively and dynamically. This yielded a rich and detailed set of insights, in the respondents' own words, which we believe to be representative for the range of attitudes towards CSD as a potential ethics-by-design tool. At the same time, interview methods deploying a relatively small number of participants cannot be used to seek out larger trends in the population, or specify insights depending on background variables such as age or experience. For such population-level generalizability of results, other methods, such as large-scale surveys, can be deployed. Finally, the proof of the proverbial pudding is in the eating. That is, when designers will have had hands-on experience in using CSD as a design tool in practice, they will be able to reflect on their own lived experiences, thus further shaping, enriching and contextualizing their (initial) perceptions, opinions and attitudes. Given that CSD is currently a new ethics-by-design method, we believe the

insights provided in this paper are a valuable point of departure for further work in this direction.

6 Concluding Remarks

This thesis has examined how ethics can be proactively integrated in the design of health and wellbeing technologies, specifically, health-related behavior change technologies. The thesis closely focused on vulnerable people and the ethical significance of vulnerability for the design of these technologies. The ethics by design framework CSD that is developed and put forward in this thesis is a design framework that is uniquely able to mitigate vulnerability in design, account for human diversity and counter (structural) injustices that manifest in technology design. CSD is developed theoretically and applied to a hypothetical design case, as well as examined empirically to see if and how CSD can be of practical use for designers in their design practice.

The thesis contributes to the field of ethics by design in general and to the further development of Value Sensitive Design in particular. With developing CSD, this thesis has strengthened the normative foundation that an ethics by design framework can have, specifically in the domain of design for health and wellbeing for vulnerable people. There are, however, a number of challenges raised in this thesis that have not yet been resolved. Let me once more reflect upon these challenges and propose ways forward for future research.

Chapter 2 discussed that in order to assure that the interests and needs of vulnerable users are taken into account, designers have to elicit the needs of their prospective vulnerable users during the design process. Further research on elicitation tools for vulnerable people, however, has been beyond the scope of this thesis. Now, although this thesis did not include explicit research on elicitation tools for vulnerable people, the developed framework of CSD could be helpful in finding solutions. As argued in chapter 2, in order to assure that the interests and needs of vulnerable users are taken into account, designers have to keep two crucial aspects in mind. These are: taking real life contexts into account and supporting stakeholders in their communication with designers. What CSD offers is a fine-grained understanding of people's individual characteristics and the social and environmental context they are in by putting a strong focus on conversion factors. With this emphatic focus on peoples' conversion factors, CSD is able to accurately take vulnerable users' real-life contexts into account, which is crucial for adequate elicitation of vulnerable people's interests and needs.

In chapter 5, design-experts were introduced to the list of ten capabilities developed by Nussbaum (2000) that forms the basis of CSD. The reactions of the experts to the list were, however, rather diverse and not unanimously positive. Some experts recognized the capabilities on Nussbaum's list as important values that they at times already aim to endorse in their design practices. Others, however, pointed out the demanding character of the list and questioned its paternalistic and elitist character. These points of critique mirrored some of the main criticisms that Nussbaum's capability list and her selection of capabilities has received in scholarly literature (see e.g., Claassen 2011; Byskov 2017). In the scholarly literature on the CA there exists a longstanding dispute between proponents of, roughly, Nussbaum's normative substantive capability theory and Sen's procedural-deliberative capability approach (see e.g., Claassen 2011; Byskov 2017; Robeyns 2017). This dispute also exists in the literature on ethics by design, namely between ethics by design approaches that are complemented with Nussbaum's substantive capability theory (see e.g., Oosterlaken 2013; Jacobs 2020) and ethics by design approaches that are complemented with the procedural-deliberative capability approach of Sen (see e.g., Cenci and Cawthorne 2020).

I have argued in chapter 4 and 5 why I have chosen to complement VSD with Nussbaum's normative substantive capability theory, instead of Sen's procedural-deliberative version of the CA. However, not all design-experts seemed convinced of the choice for Nussbaum's theory, and more specifically of using her pre-defined list of ten capabilities. In light of this critique, it is worthwhile to discuss the dispute between Nussbaum's normative substantive capability theory and Sen's procedural-deliberative version of the CA in relation to ethics by design in more detail here, and point to a possible solution to the issue.

The most prominent proponents of Sen's procedural-deliberative version of the CA combined with VSD are Cenci and Cawthorne (2020). These authors argue that the intrinsic participatory-deliberative character of "value-based technical design in democratic settings" is inconsistent with front-loading substantive ethics approaches based on fixed universal values that are used to substantiate VSD (p.2634). Instead, they propagate to complement VSD with Sen's procedural-deliberative version of the CA, with the aim not to: "provide a definite, complete list of universal values and normative principles", but rather to put emphasis on "the *transparency*, *correctness*, and *inclusiveness* of the ethical procedure behind the choice of foundational values and goals underlying technological design" (p.2647). According to Cenci and Cawthorne, "Sen's capability approach's normative and meta-ethical foundations are largely in keeping with an

explicit and extended use of *participatory-deliberative* methods in the task of eliciting shared plural incommensurable values of ethical and social importance" (2020, p.2634).

The main critique of Cenci and Cawthorne on complementing VSD with substantive normative theories -such as CSD does- is that such approaches rely too heavily on the reasoning of theorists and disregard cooperative-participatory-democratic ideals. Furthermore, they argue that within such approaches, paternalistic methods are used to select the normative ideals and principles of justice and that the list of supposedly universal values is obtained too rigidly (2020, p.2644). This clearly resembles the critique raised by some of the design-experts. A procedural-deliberative approach, such as Sen's capability approach as put forward by Cenci and Cawthorne (2020), appears to circumvent these critiques. Therefore, these authors argue that VSD should be refined not with a substantive ethics such as Nussbaum's capability theory, but with Sen's procedural-deliberative version of the CA (Cenci and Cawthorne 2020).

However, I believe these critiques can -to a high extend- be refuted, as well as that there are some strong arguments in favor of complementing VSD with Nussbaum's capability theory, as discussed in chapter 3 and 4. As argued for in these chapters, VSD faces various challenges that have to do with the fact that VSD is a procedural-deliberative approach itself that does not make any substantive ethical commitments. I have argued that VSD needs complementation with a substantive mid-level ethical theory that provides grounds of justification and argumentation for moral claims and considerations, in order to solve these challenges. Nussbaum's capability theory provides such a normative substantive mid-level ethical theory and is able to solve two of the three challenges faced by VSD, as discussed in chapter 4.

The critique that CSD would rely too heavily on the reasoning of theorists and thereby disregards the participatory and democratic ideals of design can be refuted by pointing to the 'philosopher-investigator's strategy' (Claassen 2011) that CSD endorses. By taking on the philosopher-investigator approach (Claassen 2011), as discussed in chapter 4, designers assure that the list of capabilities that they design for is not the mere result of isolated reflection but is also thoroughly informed by empirical study and stakeholder analysis. Thereby, the design process draws upon the knowledge of (many) others and thoroughly respects the participatory and democratic ideals of a design process.

The critique that CSD makes use of paternalistic methods to select normative principles, values and capabilities can be refuted by pointing to fact that Nussbaum has

always emphasized that her list of capabilities is open for discussion and revision (Nussbaum 2000, p. 77) and therefore is not to be understood as paternalistic, rigid nor definite. Nussbaum's list is intended as a piece of input into the deliberation process, i.e.; as a stimulus for debate in the design process (Claassen 2011), inviting people to contribute, revise and adapt the list as necessary for the design process at hand.

However, it has become apparent that for many people -including the authors Cenci and Cawthorne (2020) as well as many of the interviewed design-experts- this deliberative nature of Nussbaum's list is not immediately clear. Based on the findings of chapter 5, it needs to be made much more explicit that Nussbaum's list is intended as an adaptive, dynamic and open-ended list that foremost functions as input into the design process. In other words; that designers together with stakeholders are encouraged to discuss and deliberate on the list of capabilities during the design process, and that they themselves will have to provide context-specific specification to the abstract capabilities on the open-ended list (Jacobs 2020). A practical tool to assist them in this endeavor is the set of "capability cards" developed by Marc Steen (2016), as proposed in chapter 4 and 5. The capability cards provide input to designers and stakeholders to discuss, select, adapt, and specify the capabilities from Nussbaum's list to the design context at hand (Jacobs 2020). In this way, Nussbaum's list of ten capabilities thus offers a piece of input into the deliberation process, providing designers and stakeholders with a stimulus for debate in the design process (Claassen 2011).

Another remaining challenge for CSD that has been discussed in chapter 4, is how to determine whether a design ultimately meets the criteria of the capability hierarchy in a satisfactory way. How can CSD practitioners determine whether they have accounted for all the norms derived from the selected capabilities, and subsequently whether they sufficiently accounted for all the design requirements that can be derived from these norms? This concern also surfaced in chapter 5, where various design-experts remarked that there exists an issue of exhaustiveness and completeness with regard to the capability hierarchy. A professor in design questioned how designers could know that they have mapped all the norms that correspond with a certain capability. Another expert wondered how a designer can guarantee to have executed the hierarchy completely and exhaustively.

The proposed solution to this challenge is found in chapter 4, where I have argued that multiple iterations of the capability hierarchy will enable designers to check whether the conditions of the capability hierarchy are sufficiently met. However, the power of this proposed solution is in the eating of the pudding. Whether extensive

iterations of the various steps of the capability hierarchy will eventually satisfy the sufficiency challenge is only to be found out once we empirically examine this in practice. Such empirical investigation has been outside the objective of this thesis, but would be a necessary next step in future research on the practical applicability of CSD.

A further remaining challenge for CSD that has been discussed in chapter 4, is how designers should deal with capability conflicts in design contexts. This issue of prioritization in case of capability conflicts has been raised by multiple design-experts in chapter 5 as well. Various experts pointed out the lack of clear-cut guidelines on how to make value prioritizations or trade-offs when capability conflicts occur, and remarked that CSD doesn't seem to provide the needed guidelines on how to deal with these issues. What made the issue even more challenging is that the design-experts on the one hand indicated a need for clear-cut guidelines on how to deal with capability conflicts, but on the other hand feared interference with the necessary flexibility, iteration, context-adaptability and creative freedom that are crucial to successful design processes.

The solution to dealing with capability conflicts that I have proposed in chapter 4 has been to advise designers to return to the process of specification that forms the core of the capability hierarchy. By repeatedly going through the process of specification, designers together with their stakeholders jointly deliberate on the relevant context- and domain-specific content that should be added to the abstract capabilities, with the aim to ultimately come to a joint solution on the design level. What is apparent in this solution is that the strength of the solution is to be found in practice. We can only find out if this solution indeed works by empirically examining the repeated process of specification in design processes by designers and their stakeholders. Thus, here as well, a crucial next step in future research is to empirically examine whether and how designers are successful in dealing with capability conflicts when they make use of the proposed solution.

When designers gain hands-on experience in using CSD in practice, they will be able to further shape, enrich and contextualize the framework. Future research directions that I envision will first and foremost consist of empirical research on the lived experiences of designers with CSD. Such future empirical research will most-likely provide us with valuable insights that enable us to further strengthen CSD and increase its practical applicability in design practices.

Future relevant *theoretical* research would be to explore the added value of CSD not only for the domain of technology design for health and wellbeing, but for other domains as well, such as e.g., technology design for education, social robotics, or environmental sustainability. Furthermore, it would be relevant to examine the value of CSD in inter-cultural design contexts, e.g., within a design team that consists of people from different cultural backgrounds who design a technology for a global market. Although CSD is based on the normative foundation of Nussbaum's capability theory who has always argued that the ten basic capabilities on her list are of universal value (Nussbaum 2000), it is important to investigate both theoretically as well as empirically whether CSD is indeed applicable in inter-cultural design contexts, or that there are cultural differences that require certain adaptations of CSD.

Epilogue

It's a bright morning in early autumn and Emma pushes open the swinging doors of the eHealth Design Lab. Clutched under her arm is a large map with sketches. She will show them to her team for the first time today.

As senior design strategist at the design lab, she is in charge of developing a new design for an app that encourages people to behave more empathetically towards each other. In the past year, Emma has often been amazed at the lack of empathy in society and the hardening of public discourse. People seem to have lost the ability to listen to each other and this increasingly worries her.

Shortly after the elections in early spring, her design team was approached by a non-profit organization with a mission to promote democratic values in society. The organization reported on a growing inability by different societal groups to listen and speak to each other. The corona crisis had accelerated this inability and caused a further degradation of democratic values. In response, the non-profit had teamed-up with a national organization for the promotion of the mental wellbeing of young people. Various reports showed that during the pandemic the mental health of young people had declined. The multiple lockdowns had impeded social contact and many young people felt increasingly isolated, which impacted their mental wellbeing.

Together, both organizations had come to Emma's design lab to explore the possibility of developing a service that enables young people to expand their social circles. They wanted something to stimulate an empathetic and open-minded attitude amongst young people in order to facilitate mutual understanding within communities. Ideally, they reasoned, such a service would contribute to democratic values in society as well as enhance people's mental wellbeing. They had come to the eHealth Design Lab because of its strong reputation in designing applications that effectively change people's attitudes and behaviors to improve their wellbeing.

Their first meeting took place at the end of spring. The directors of both organizations had come to the Design Lab to meet with Emma and the designers on her team Karim, Minke and Sefora.

During that first meeting, Minke made an impassioned plea on how important it is to design something with *real* societal impact. Minke was the youngest designer on the team and she never missed a change to emphasize the importance to *design for good*.

It was Emma's colleague Sefora who then told everyone about this new design tool. It was some kind of framework to design for the social good. To design for values like equality, justice and diversity Sefora had said. She had heard about it at a recent health by tech conference where there was a panel session on designing for justice.

It struck Emma that a new generation of designers had emerged for whom it comes so natural, *self-evident* even, to think about the societal and ethical impacts of their designs.

Back when she started, it was rather different. It had often been a *struggle* to convince her fellow-designers -let alone clients and stakeholders- of the importance to think about ethical considerations in design. Besides, there weren't many *tools* to support you in designing for values. At a certain point, Emma came across *Value Sensitive Design*. She thought it looked promising and she hoped it would provide her with concrete steps to really do ethics by design. But it turned out not to be so easy after all. Value Sensitive Design didn't provide her with specific information about how to select the values for which to design, it simply didn't seem to offer tools to distinguish important values from the mere preferences and whims of her clients and stakeholders. Emma also recalls that one time she and her team tried to design for *twelve* values, because all twelve values were indicated equally important by their stakeholders. That was impossible. What her team missed most was some sort of clarity, some *guidance* from the value sensitive framework that helped them decide on what values needed to be prioritized…

♦

Not long after their initial meeting, Minke reported back on the new ethics by design framework. It was something called *capability sensitive design*, she told the team, developed at Eindhoven University of Technology. Apparently, it was developed by an ethicist, which Emma found rather surprising.

Emma noticed that her first reaction was somewhat skeptical. A design framework developed by someone who is not a designer herself. But an *ethicist*. Although Emma agrees on the importance of ethics by design, she could not help but notice a slight feeling of agitation. She realized it probably had to do with her recent encounters with

the Ethics Review Board. Their review procedures can be so *rigid* and *time consuming*… She felt frustration bubbling up inside her again. *Anyway*, she thought to herself, judging by Minke and Sefora's plea, she had to give it a fair shake, *right*?

Minke and Sefora emphasized that the team's current design context focuses on how to persuade people to behave more empathetically. To have people give room to *diverse* voices, opinions, and feelings. In other words, to stimulate people to *listen to others* and engage and connect with them, to find ways to be *open minded* towards other people's opinions, thoughts and feelings even if they are completely different from yours. A design approach that strongly focuses on *human diversity* and that stimulates designers to do justice to human dignity and equal opportunity, like *capability sensitive design*, wouldn't that be a really good fit with our design challenge, Minke and Sefora asked rhetorically.

Although Emma wasn't quite convinced *yet*, she decided to practice what she was about to preach and to take on an open-minded attitude. Together with the team they decided to use the capability sensitive design framework *flexibly*. To only use those aspects of the framework that they judge relevant and applicable to their design context.

♦

Minke and Karim are already seated at the large table in the middle of the room, chatting lively with a coffee in hand. Emma places her map with sketches and her bag on the table and rushes over to the kitchen to make herself an espresso. She is a little nervous. The co-design workshop takes place today.

She invited a group of about twenty participants. Among them are various people from the *pro-democracy* non-profit and from the mental youth health organization. She also invited youngsters from various backgrounds between the age of 18 and 25-years old. Through social media her team tried to reach out to young people from all kinds of social circles and different political affiliations. This wasn't easy to do. She is happy they managed to invite ten youngsters who would *actually* come.

The main aim of today's workshop is to discuss and clarify the project's ultimate goal. Emma is really curious what the young people will say about empathy and open mindedness. What *they* think a desired situation would look like.

Hopefully, together they will be able to clarify the main goal of the project during the first part of the workshop. They have the whole morning to come up with something. In the afternoon they will move on to *capabilities*. Emma will introduce everyone to the new terminology. She will ask the participants *which capabilities* should be enhanced by the design. She is especially nervous about this second part of the workshop. Emma wonders whether the participants will pick up on the idea of *thinking in terms of capabilities*. Luckily, they have a capability-cards set to help the participants clarify and select capabilities they find important. She has placed her hopes on these cards.

It's a little after 10 o'clock and everyone is seated at the large table – it's such a pleasure that this is possible again now that everyone is vaccinated.

Emma closes the door and walks over to the presentation screen in front of the room.

"Thank you all so much for coming out today and joining us at the eHealth Design Lab. I'm very happy that we can meet in person again. I think it's safe to say that living in times of a pandemic has left its marks on all of us…

Reports of various world-wide health organization have taught us that people felt increasingly isolated and alone during the pandemic. And this has caused many people to feel gloomy, or even depressed.

What research has furthermore shown, is that feelings of isolation, loneliness and depression decrease our ability to empathetically engage with others. These feelings of loneliness and isolation have caused us to be very much caught up in our own realms. Making it increasingly difficult to extend beyond our own bubbles and connect with others outside our direct scopes…

Although we are carefully recovering from the crisis and people are emerging from their isolated bubbles and resume their social contacts, we do see deepening cracks and divisions running through society. Misha, director of the organization *pro-democracy* and our partner in this project, will shortly give a presentation on the effects of the pandemic on our shared democratic values and the increasing polarization in public discourse…"

Emma goes on to tell everyone that Misha, together with Gloria from the organization of mental health for youth, have come to the eHealth Design Lab with the request to design an application that would stimulate young people to *engage* and connect to others, to find ways to be *open minded* towards other people's opinions and persuade them to behave more empathetically towards each other.

"Together with my design team at the eHealth Design Lab," Emma continues, "we have decided to use a *new design tool* to assist us in this design project. It's the first time that we use this new tool and we want to try it out with you. So, it's a real *scoop*."

She explains that this new design tool is called *capability sensitive design* and that it enables them to start a design process by thinking about what capabilities we -the *designers*, *stakeholders and prospective users*- want to see enhanced by the application.

"In other words," she says, "it helps us answer the question what possibilities we want the application to facilitate and enhance."

Emma goes on to explain that during the afternoon session they will use *capability cards* to evoke discussion on what capabilities to select for the design.

"But first, it's time for a presentation by Misha from *pro democracy*. Misha, can I give the floor to you?"

"Yes, of course, thank you."

"Let me introduce to you Misha, who will give a short, and as I'm told an interactive presentation on recent examples of polarization in public discourse. Misha, the floor is yours…"

♦

The lunch break is over and people are slowly returning to the main room. Some are chatting with soft voices; others are quietly taking a seat at the table. Minke and Sefora have placed sets of capability cards on each corner of the table. Together they will lead the capability cards session this afternoon. Emma takes a seat at the left-back corner of the table. To her right sits a young woman named Lauren. It struck Emma how confident and self-aware this young woman had come across during the morning

session. When she spoke, she was so *candid* and *calm* at the same time. A striking combination, Emma thought, especially at such a young age. To her left sits Ciro. At first glance Emma thought he was a bit shy but as the morning progressed, Ciro tuned out to be a lively speaker with lots of little jokes up his sleeve.

"Welcome back everyone" Sefora kicks off. "This afternoon you will participate in a small group session wherein you use *capability cards*. Minke and I will quickly introduce you to the cards and explain the aim of this session. Most important, however, is that you just *feel free* to speak what's on your mind, ask questions about *anything* and freely voice what you think and feel is important to you."

Sefora turns to Minke who has a stack of cards in her hands. "That's totally right", Minke says as she takes over. "This afternoon session is about you feeling free to voice what you think is important in the design of this new application that we will create. In the morning session it became clear that most of us find it important that the application promotes *open mindedness* and that it enables people to come into contact with others, even with people who don't share our worldviews, opinions, or values. Now, in this afternoon session, we are going to map these findings to *capabilities*…"

Minke gives the word back to Sefora who goes on to tell that at the core of *capability sensitive design* lies the idea that every human is *equally worthy* and deserves a life worth living. To ensure that people are equally able to live a worthy life, we need to look at what people are able to be and do; their *capabilities*. Only if people have the freedom to exercise the capabilities important to them, they are able to life a worthy and joyful life. "Capabilities are the *freedoms* that people have to be and do what they find important", Sefora explains.

"Now, what I have here…", Minke explains as she waives her hand with which she holds a stack of cards, "is a card set with capabilities envisioned on them. The philosopher Martha Nussbaum has created a list of *ten capabilities* that are important for all people to be able to live a dignified life. These ten capabilities are envisioned on the cards that I'm holding right here. You also find these card sets on each corner of the table. What Sefora and I have done is we added some capabilities to the card set that we think will in addition be particularly relevant to our design context. Also, we have put in some *blank cards*. If, during the session you come up with a capability that you

find important that is presently lacking, you can write that capability on a blank card and add it to the mix…"

"I will now divide you into groups of four", Sefora continues, "and I invite you to take a close look at the cards and read the instructions on the cards. Each card gives a short description of a capability accompanied by a question that should give you food for thought, and hopefully evokes a discussion amongst the four of you. Minke and I will alternate between the groups to assist you in the process…"

♦

As Minke walks around the room she overhears snippets of conversations. What strikes her is that all groups got off to a flying start. Each group studied the cards and then quickly engaged in lively conversations. Minke takes a halt at Karim's group. Carefully without interrupting their conversation, she grabs a chair and joins them. The group consists of Karim, her fellow designer at the lab; Ciro, a witty boy of about 20-years old; Dhevan from the *pro- democracy* organization; and Dorelia, a woman in her mid-twenties with a colorful headscarf wrapped around her dark curls.

Dorelia is holding a card in her hand. When Minke slides to the top of her chair she can see what's on it. She reads: *the capability to have social affiliations that are meaningful and respectful*. Dorelia is telling the group that during the pandemic she has mainly been in contact with her three closest friends. She tells them that occasionally she would meet up with her friends in a park or so… but mostly they would organize *zoom dates* or meet on the app Clubhouse. Although she enjoyed spending time with her friends and even felt she really needed them to stay sane, she did miss something… She says she's not quite sure what it was. But after some months into the pandemic, she started to feel a bit *empty* and *indefinable*… Dorelia pauses for a moment and then quickly shrugs her shoulders saying *I don't know*…

Ciro joins in and says he has also been using Clubhouse. He says he finds the app fun and that it has given him easy opportunities to meet *new people*, something he missed the most during the pandemic. Talking to strangers in Clubhouse really was something that helped him keep things *light* and *fun*, he tells the group. During the pandemic he felt empty at times as well, just like Dorelia had described. And engaging in

conversations - just talking about *anything* really- with new people had helped him overcome feelings of loneliness.

Dhevan says he is also on Clubhouse. What he likes most about it, he tells the rest, is that it enables him to interact with all kinds of people he didn't know before, and to swap ideas and thoughts. He says he finds it exciting to get in touch with people he normally wouldn't meet otherwise.

Karim tries to capture what the others just told. He asks them whether they would agree that the *merit* of an app like Clubhouse is in meeting strangers and swapping ideas in a respectful environment? The others nod in agreement. "So, do you agree that this would be an important feature of a new application, to ensure that *this capability of having social affiliations that are meaningful and respectful* is enhanced by the app?" Karim asks. "Yes, I would say so", Dhevan replies. Dorelia and Ciro both hum in approval. "*Meaningfulness* is, of course, a *flexible* concept" Ciro says jokingly. "I mean, what is *meaningful*, right? I just like to chat from time to time, and *that's* meaningful to me, you know? Just having a chat with someone, as long as it's respectful, that's fine by me." Dorelia nods, "yeah, I agree. *Meaningful* could mean *so many things*. But respectful, yeah, I find that really important…"

While trying not to make any noise, Minke gets up from her chair and walks towards the group at the corner of the table where she hears Emma talking. She quietly joins to listen in on how their conversation is going…

♦

Emma treats herself to a seat in the first-class compartment. It has been a long day. Exciting *and* exhausting at the same time. Once she has settled comfortably in her seat, she takes out her notebook and places it on the small table in front of her. She has her pen ready in her hand but she decides to gaze out of the window for a little while…

She thinks about the afternoon session of the co-design workshop. With her group Emma had spent most time on a card about the capability *to possess practical reason to form a conception of the good*. She had engaged in a *somewhat heated* discussion on what it means to form a conception of *the good*. Lauren had started off the conversation by asking how we are to understand 'practical reason.' What does it entail exactly, she

had asked, and what if you don't have it, like a young child perhaps, or someone who has cognitive impairments, are these people unable to form a conception of *the good*, she had asked? And is there something like *the* good, she had asked. There are so many people out there and most of them have *completely different* conceptions of *the good*, don't they? Daniël, another young participant, had at that point suggested to *just skip* the card since it gave rise to so many *difficult questions*, he had said while looking at Lauren. Misha, however, indicated that this difficulty probably hinted at something important that they needed to map out. Lauren had nodded in agreement. Not only are there *many conceptions* of the good, Misha had continued, but I would say that there is virtually *nothing more annoying* than having someone else telling you what is *good* and what not. Daniël recalled that he often had felt miserable as a child precisely for that reason. How *great* it had been to finally turn eighteen and be able to decide things for himself. Having someone else telling you what is right is just *horrible*, Lauren had said. Indeed, Misha had said, it is important that people may hold different conceptions of the good and that they are able to come up with such conceptions *themselves*. For a democratic society that's is extremely important, he had said, adding that it is crucial that people are able to form their own views, values and beliefs and that these may differ from the views that others hold.

It had been Emma who had pointed back to the capability card, suggesting that the capability on the card could then perhaps best be interpreted as saying that every person should be able to form a conception of the good for him- or herself, with help of practical reasoning. This practical reasoning then, she had added, would enable us to talk about our conceptions of the good with each other, to engage in discussion, without necessarily agreeing with each other.

At the centre of a large sheet of paper Emma had then started to write down 'the good'. She gave Daniël, Lauren and Misha a pen, each with a different color. She had asked them to write down words or concepts around the centre that they associated with 'the good' based on the conversation they just had. Daniël had written down 'my own', while Lauren wrote down 'plurality', and 'diversity', and Misha added 'egalitarianism' and 'discussion'.

While they were all bent over the sheet of paper, Lauren had said that they shouldn't forget those people for whom practical reasoning might be harder... Some people are simply less able to express themselves, she had said, and these people should

be especially supported by what we will design, to assist them with giving voice to their thoughts and ideas, *right*?

♦

Emma slowly manages to pull her gaze from the moving landscape. She wonders whether working with the capability cards had provided their group discussion with clarity, or whether it had only made their conversation more complicated. In any case, it had become clear that the capability *to possess practical reason to form a conception of the good* was in no way unambiguous nor straightforward to understand. To be honest, Emma had pondered on the formulation herself before, finding it rather *vague* and *indeterminate*… What she did like, *though*, is that it had sparked a lively discussion amongst the participants, perhaps *precisely* because of that indeterminate formulation. And although their discussion hadn't been particularly easy, she is sure that it has yielded many threads that her design team can work with during the design process.

References

Albrechtslund, A. (2007). Ethics and Technology Design. *Ethics and Information Technology*. Vol. 9. No.1. pp.63–72.

Arras, J. (2006). Theory and Bioethics. *The Stanford Encyclopedia of Philosophy*. Zalta, E. N. (ed.) Accessed on 1-12-2016 at http://plato.stanford.edu/archives/sum2016/entries/theory-bioethics/.

Barak, A. and Grohol, J. (2011). Current and Future Trends in Internet-Supported Mental Health Interventions. *Journal of Technology in Human Services*. Vol.29. pp.155-196.

Bekker, M. and Long, J. (2000). User Involvement in the Design of Human-Computer Interactions: Some Similarities and Differences Between Design Approaches. *People and Computers XIV—Usability or Else!* London: Springer. pp.135-147.

Berdichevsky, D. and Neuenschwander, E. (1999). Towards and Ethics of Persuasive Technology. *Communications of the ACM*. Vol. 42. No.5. pp.51-58.

Bernal, P. (2014). Privacy, Autonomy and The Internet. In: *Internet Privacy Rights: Rights to Protect Autonomy*. Cambridge Intellectual Property and Information Law. Cambridge: Cambridge University Press.

Beauchamp, T.L. and Childress, J. F. (2013). *Principles of Biomedical Ethics*. Oxford: Oxford University Press.

Boorse, C. (1975). On the Distinction Between Disease and Illness. *Philosophy and Public Affairs*. Vol.5. pp.49–68.

Borning, A. and Muller, M. (2012). Next Steps for Value Sensitive Design. CHI 2012 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Accessed on 2-12-2016 at http://dl.acm.org/citation.cfm?doid=2207676.2208560.

Blumenthal-Barby, J.S. (2012). Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts. *Kennedy Institute of Ethics Journal*. Vol.22, No.4. pp.345-366.

Braun, V. and Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*. Vol. 3. pp.77-101.

Brighouse, H. and Robeyns, I. (eds.) (2010). *Measuring Justice: Primary Goods and Capabilities*. Cambridge: Cambridge University Press.

Burri Gram-Hansen, S. (2013). Towards an Approach to Ethics and HCI Development Based on Løgstrup's Ideas. *INTERACT '09 Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I*. pp.200-203.

Buse, C. E. (2010). E-Shaping the Ageing Body? Computer Technologies and Embodiment in Later Life. *Ageing & Society*. Vol. 30. No. 6. pp.987-1009.

Butler, J. (2004). *Precarious Life: The Powers of Mourning and Violence*. London: Verso.

Butler, J. (2009). *Frames of War: When is Life Grievable?*. London: Verso.

Calvo, R., Milne, D., Hussain, M., Christensen, H. (2017). Natural Language Processing in Mental Health Applications Using Non-Clinical Texts. *Natural Language Engineering*. Vol. 23. pp.649-685.

Cawthorne, D. and Cenci, A. (2019). Value Sensitive Design of a Humanitarian Cargo Drone. *International Conference on Unmanned Aircraft Systems (ICUAS)* IEEE. pp.1117-1125.

Cawthorne, D. and Robbins-van Wynsberghe, A. (2019). From HealthDrone to Frugal Drone: Value Sensitive Design of a Blood Sample Transportation Drone. *EEE International Symposium on Technology and Society (ISTAS)*. DOI: 10.1109/ISTAS48451.2019.8938019.

Cenci, A. and Cawthorne, D. (2020). Refining Value Sensitive Design: A (Capability-Based) Procedural Ethics Approach to Technological Design for Well-Being. *Science and Engineering Ethics*. Vol.26. No.6. pp.2629–2662.

Cenci, A. and Hussein, A. (2019). Epistemic and Non-epistemic Values in Economic Evaluations of Public Health. *Journal of Economic Methodology*. Vol. 27. No.1. pp.66-88.

Claassen, R. (2011). Making Capability Lists: Philosophy Versus Democracy. *Political Studies*. Vol.59. pp.491-508.

Coeckelbergh, M. (2009). Health Care, Capabilities, and AI Assistive Technologies. *Ethical Theory and Moral Practice*. Vol. 13. pp.181-190.

Coeckelbergh, M. (2010). Human Development or Human Enhancement? A Methodological Reflection on Capabilities and the Evaluation of Information Technologies. *Journal of Ethics and Information Technology*. Vol. 13. pp.81-91.

College voor de Rechten van de Mens (2020). Zwangerschapsdiscriminatie is een probleem dat helaas nog niet minder wordt. Accessed on 30-06-2021 at https://mensenrechten.nl/nl/toegelicht/zwangerschapsdiscriminatie-een-probleem-dat-helaas-nog-niet-minder-wordt

Crawford, R. (2006). Health as a Meaningful Social Practice. *Health*. Vol.10. No.4. pp.401-420.

Criado-Perez, C. (2019). *Invisible Women, Exposing Bias in a World Designed for Men*. London: Vintage.

Custers, B. (2016) Click Here to Consent Forever: Expiry Dates for Informed Consent. *Big Data & Society*. Vol.3. No.1. pp.1-6.

Davis, J. (2009). Design Methods for Ethical Persuasive Computing. *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*. pp.1-8.

Davis, J. and Nathan, L.P. (2015). Value Sensitive Design: Applications, adaptations, and critiques. In Hoven, van den J., Vermaas, P.E., and Poel, van de I. (eds.) *Handbook of Ethics, Values, and Technological Design*. Dordrecht: Springer. pp.11-40.

Delano, M. (2015). I tried tracking my period and it was even worse than I could have imagined. Accessed on 1-9-2019 at https://medium.com/@maggied/i-tried-tracking-my-period-and-it-was-even-worse-than-i-could-have-imagined-bb46f869f45.

Dickinson, E. (2011). GDP: A Brief History. Accessed on 18-12-2019 at: https://foreignpolicy.com/2011/01/03/gdp-a-brief-history/.

Dodds, S. (2008). Inclusion and Exclusion in Women's Access to Health and Medicine. *International Journal of Feminist Approaches to Bioethics*. Vol.1. No.2. pp.58-79.

Doyal, L. and Gough, I. (1984). A Theory of Human Needs. *Critical Social Policy*. Vol. 10. pp.6-38.

Fineman. M.A. (2008). The Vulnerable Subject: Anchoring Equality in the Human Condition. *Yale Journal of Law and Feminism*. Vol.20. No.1. pp.1-23.

Flanagan, M. and Nissenbaum, H. (2014). *Values at Play in Digital Games*. Cambridge: MIT Press.

Fogg, B.J. (2003). *Persuasive Technology, Using Computers to Change What We Think and Do*. San Francisco: Morgan Kaufmann Publishers.

Ford, M. E. and Nichols, C.W. (1987). A Taxonomy of Human Goals and Some Possible Applications. In Ford, M. E. and Nichols, C.W. (ed.). *Humans as Self-Constructing Systems: Putting the Framework to Work*. New York: Routledge. pp.289-312.

Frediani, A.A. and Boano, C. (2012). Processes for Just Products: The Capability Space of Participatory Design. In Oosterlaken, I. and Van den Hoven, J. (eds). *The Capability Approach, Technology and Design*. Springer: Dordrecht. pp.203-222.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Friedman, B. (ed.) (1997). *Human Values and the Design of Computer Technology*. Stanford: CSLI Publications.

Friedman, B. and Hendry, D. (2019). *Value Sensitive Design, Shaping Technology with Moral Imagination*. Cambridge: The MIT Press.

Friedman, B. and Kahn Kahn Jr. P. H. (2003). Human Values, Ethics, and Design. In: Jacko, J. A. and Sears, A. (eds). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. Mahwah: Lawrence Erlbaum. pp.117-1202.

Friedman, B., Kahn Jr. P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In: *Early Engagement and New Technologies: Opening Up the Laboratory*. Ed: Doorn, N., Schuurbiers, D., van de Poel, I., Gorman, M.E. Dordrecht: Springer. pp.55-95.

Friedman, B., Nathan, L., Kane, S., and Lin, J. (2011). Envisioning Cards. Seattle: University of Washington.

Friedman, B. and Nissenbaum, H. (1996). Bias in Computer Systems. *ACM Transactions and Information Systems*. Vol.14. No.3. pp.330-347.

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

Gordon, J. S., Rauprich, O. and Vollmann, J. (2011). Applying the Four-Principle Approach. *Bioethics*. Vol. 25. No. 6. pp.293-300.

Gough, I. (2014). Lists and Thresholds: Comparing the Doyal-Gough Theory of Human Needs with Nussbaum's Capability Approach. In Comim, F. and Nussbaum, M. (eds). *Capabilities, Gender, Equality, Towards Fundamental Entitlements*. Cambridge: Cambridge University Press. pp.357-381.

Haenssgen, M. J. and Ariana, P. (2017). The Place of Technology in the Capability Approach. *Oxford Development Studies*. Vol.46. pp.98-112.

Hancke, B. (2016). Mitigating Risk and Facilitating Access to Capabilities: The Role of Affect in the Design of an ICT-tool for Queer Youth in Asia. *Emotion, Space and Society*. Vol.18. pp.35-43.

Harwell, D. (2019). Who else is tracking your pregnancy?. Accessed on 1-9-2019 at https://www.washingtonpost.com/technology/2019/04/10/tracking-your-pregnancy-an-app-may-be-more-public-than-you-think/.

Haverkamp, B., Bovenkerk, B., Verweij, M. (2018). A Practice-Oriented Review of Health Concepts. *Journal of Medicine and Philosophy*. Vol.43. pp.381-401.

Hendl, T., Jansky, B., en V. Wild. (2020). From Design to Data Handling: Why mHealth Needs a Feminist Perspective. In Loh, J. and Coeckelbergh, M. (eds.) *Feminist Philosophy of Technology*. Stuttgart: J.B. Metzler. pp.77-103.

Hoermann, S., McCabe, K., Milne, D., and Calvo, R. (2017). Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review. *Journal of Medical Internet Research*. Vol. 19. pp.1-10.

Hoven, Van den J. (2008). Moral Methodology and Information Technology. In Himma, K. E., and Tavani, H. T. (eds.). *The Handbook of Information and Computer Ethics*. New Jersey: Wiley. pp.49-68.

Hoven, Van den J. (2012) Human Capabilities and Technology. In Oosterlaken, I., Van den Hoven J. (eds). *The Capability Approach, Technology and Design*. Springer: Dordrecht. pp.27-36.

Hoven, Van den J., G.-J. Lokhorst., and I. van de Poel. (2012). Engineering and the Problem of Moral Overload. *Science and Engineering Ethics*. Vol. 18. pp.143-155.

Hoven, Van den J., Vermaas, P. E., and Van de Poel, I. (2015). *Handbook of Ethics, Values and, Technological Design. Sources, Theory, Values and Application Domains*. Dordrecht: Springer.

Huber, M., Knottnerus, J., Green, L., Van der Horst, H., Jadad, A., Kromhout, D., Leonard, B., Lorig, K., Loureiro, M., van der Meer, J., Schnabel, P., Smith, R., van Weel, C., Smid, H. (2011). How Should We Define Health? *BMJ*. Vol. 343. doi:10.1136/bmj.d4163.

Hurst, S. A. (2008). Vulnerability in Research and Health Care: Describing the Elephant in the Room?. *Bioethics*. Vol. 22. No.4. pp.191-202.

Ihde, D. (1993). *Philosophy of Technology: An Introduction*. New York: Paragon House.

IJsselsteijn, W., De Kort, Y., Midden, C., Eggen, B., & Van Den Hoven, E. (2006). Persuasive Technology for Human Well-Being: Setting the Scene. *International Conference on Persuasive Technology Berlin*. Heidelberg: Springer. pp.1-5.

Iversen, N., Birkved, M., and Cawthorne, D. (2020). Value Sensitive Design and Environmental Impact Potential Assessment for Enhanced Sustainability in Unmanned Aerial Systems. *IEEE International Symposium on Technology and Society 2020*.

Jacobs, N. and Huldtgren, A. (2018). Why Value Sensitive Design Needs Ethical Commitments. *Journal of Ethics and Information Technology*. Vol.23. No.1. pp.23-26.

Jacobs, N. (2020). Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People. *Bioethics*. Vol. 34. pp.519-526.

Jacobs, N. (2020). Capability Sensitive Design for Health and Wellbeing Technologies. *Science and Engineering Ethics*. Vol. 26. No.6. pp.3363-3391.

Johnstone, J. (2007). Technology as Empowerment: A Capability Approach to Computer Ethics. *Journal of Ethics and Information Technology*. Vol. 9. pp.73-87.

Kaltenthaler, E., Brazier, J., De Nigris, E., Tumur, I., Ferriter, M., Beverly, C., Parry, G., Rooney, G., Sutcliffe, P. (2006). Computerised Cognitive Behaviour Therapy for Depression and Anxiety Update: A Systematic Review and Economic Evaluation. *Health Technology Assessment*. Vol. 10. pp.1-168.

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I. (2019). Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. *Biomedical Informatics Insights*. Vol. 11. pp.1-9.

Lanzing, M. (2019). *The Transparent Self: A Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*. PhD Dissertation. Eindhoven University of Technology.

Lazzaro, S. (2017). Is this Soap Dispenser Racist? Controversy as Facebook employee shares video of machine that only responds to white skin. *Daily Mail*. Accessed on 17-12-2019 at: https://www.dailymail.co.uk/sciencetech/article-4800234/Is-soap-dispenser-RACIST.html

Levine, C., Faden, R., Grady, C., Hammerscmidt, D., Eckenwiler, L. Sugarman, J. (2004). The Limitations of 'Vulnerability' as a Protection for Human Research Participants. *American Journal of Bioethics*. Vol.4. No.3. pp.44-49.

Levinson, C.A., Fewell, L., Brosof, L.C. (2017). My Fitness Pal Calorie Tracker Usage in the Eating Disorders. *Eating Behaviors*. Vol.27. pp.14-16.

Luna, F. (2009). Elucidating the Concept of Vulnerability: Layers not Labels. *International Journal of Feminist Approaches to Bioethics*. Vol.2 No.1. pp.121-139.

Lupton, D. (1995). *The Imperative of Health: Public Health and the Regulated Body*. London: Sage.

Lupton, D. (2013). Quantifying the Body: Monitoring and Measuring Health in the Age of mHealth Technologies. *Critical Public Health*. Vol.23. No.4. pp.393-403.

Maathuis, I., Niezen, M., Buitenweg, D., Bongers, I. L., and van Nieuwenhuizen, C. (2019). Exploring Human Values in the Design of a Web-Based QoL-Instrument for People with Mental Health Problems: A Value Sensitive Design Approach. *Science and Engineering Ethics*. Vol. 26. pp.871-898.

MacIntyre, A. (1999). *Dependent Rational Animals: Why Human Beings Need the Virtues*. Chicago: Open Court.

Mackenzie, C. (2013). The Importance of Relational Autonomy and Capabilities for an Ethics of Vulnerability. In: Mackenzie, C., Rogers, W., Dodds, S. (eds.) (2013). *Vulnerability, New Essays in Ethics and Feminist Philosophy*. Oxford: Oxford University Press. pp.33-59.

Mackenzie, C., Rogers, W., Dodds, S. (2013). Introduction: What is Vulnerability, and Why Does it Matter for Moral Theory?. In Mackenzie, C., Rogers, W., Dodds, S. (eds.) (2013). *Vulnerability, New Essays in Ethics and Feminist Philosophy*. Oxford: Oxford University Press.

Manders-Huits, N. (2010). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*. Vol.17. No.2. pp.271-287.

Manders-Huits, N. and Zimmer, M. (2009). Values and Pragmatic Action: The Challenge of Introducing Ethical Intelligence in Technical Design Communities. *International Review of Information Ethics*. Vol.10. No. 2. pp.37-45.

Mazé, R. (2016). Design and the Future: Temporal Politics of 'Making a Difference'. In R. C. Smith, K. T. Vangkilde, M. G. Kjaersgaard, T. Otto, J. Halse, and Binder, T. (eds.). *Design Anthropological Futures*. London: Bloomsbury. pp.37-54.

Meek Lange, M., Rogers, W., Dodds, S. (2013). Vulnerability in Research Ethics: A Way Forward. *Bioethics*. Vol. 27. No. 6. pp.333-340.

Mink, A., Parmar, V.S., Kandachar P.V. (2014). Responsible Design and Product Innovation from a Capability Perspective. In: Van den Hoven J., Doorn N., Swierstra T., Koops BJ., Romijn H. (eds.) *Responsible Innovation 1*. Springer: Dordrecht. pp.113-148.

Murphy, C. and Gardoni, P. (2012a). The Capability Approach in Risk Analysis. In: Roesser, S., Hillerbrand, R., Sandin, P., Peterson, M. (eds.) (2012). *Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk*. Dordrecht: Springer. pp.980-959.

Murphy, C. and Gardoni, P. (2012b). Design, Risk and Capabilities. In Oosterlaken I. and Van den Hoven J. (eds.) *The Capability Approach, Technology and Design*. Springer: Dordrecht. pp.173-188.

Nichols, C. and Dong, A. (2012). Re-conceptualizing Design Through the Capability Approach. In Oosterlaken I. and Van den Hoven J. (eds.) *The Capability Approach, Technology and Design*. Springer: Dordrecht. pp.189-201.

Nordenfelt, L. (1986). Health and Disease: Two Philosophical Perspectives. *Journal of Epidemiology and Community Health*. Vol. 41. pp.281–284.

Nussbaum, M. (2000). *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

Nussbaum, M. (2006). *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge: Harvard University Press.

Nussbaum, M. (2011). *Creating Capabilities*. Cambridge: Harvard University Press.

O'Neill, O. (2003). Some Limits of Informed Consent. *Journal of Medical Ethics*. Vol. 29. No.1. pp.4-7.

Oosterlaken, I. (2012). Inappropriate Artefact, Unjust Design? Human Diversity as a Key Concern in the Capability Approach and Inclusive Design. In Oosterlaken, I., Van den Hoven J. (eds.) *The Capability Approach, Technology and Design*. Springer: Dordrecht. pp.223-244.

Oosterlaken, I. (2013). *Taking a Capability Approach to Technology and Its Design*. PhD Dissertation. TU Delft.

Oosterlaken, I. (2015a). Human Capabilities in Design for Values. In Hoven, van den J., Vermaas, P.E., and Poel, van de I. (eds.) *Handbook of Ethics, Values, and Technological Design*. Dordrecht: Springer. pp.221-250.

Oosterlaken, I. (2015b). Applying Value Sensitive Design (VSD) to Wind Turbines and Wind Parks: An Exploration. *Science and Engineering Ethics*. Vol. 21. pp.359-379.

Orji, R., Moffatt, K. (2016). Persuasive Technology for Health and Wellness: State-of-the-Art and Emerging Trends. *Health Informatics Journal*. Vol.24. No.1. pp.66-91.

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

Petty, R. E., Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Communication and Persuasion*. New York: Springer. pp.1-24.

Poel, Van de I. (2012). Can We Design for Wellbeing? In Brey, P., Briggle, A. and Spence, E. (eds.) (2012). *The Good Life in a Technological Age*. London: Routledge.

Poel, Van de I. (2013). Translating Values into Design Requirements. In: Michelfelder, DP., McCarthy, N. and Goldberg, DE. (eds.) (2013) *Philosophy and Engineering: Reflections on Practice, Principles and Process.* Dordrecht: Springer. pp.253-266.

Pommeranz, A., Detweiler, C., Wiggers, P., Jonkers, C. (2012). Elicitation of Situated Values: Need for Tools to Help Stakeholders and Designers to Reflect and Communicate. *Journal of Ethics and Information Technology*. Vol.14. No.4. pp.285-303.

Rawls, J. (1971). *A Theory of Justice*, Cambridge: Harvard University Press.

Reeves, S., Goulden, M., and Dingwall, R. (2016). The Future as a Design Problem. *Design Issues*. Vol. 32. pp.6-17.

Richardson, H. S. (1990). Specifying Norms as a Way to Resolve Concrete Ethical Problems. *Philosophy and Public Affairs*. Vol.19. pp.279-310.

Robeyns, I. (2009). Justice as Fairness and the Capability Approach. In Basu, K. and Kanbur, R. (eds.) (2009). *Arguments for a Better World. Essays for Amartya Sen's 75th Birthday*. Oxford: Oxford University Press. pp.397–413.

Robeyns, I. (2016). The Capability Approach. *The Stanford Encyclopedia of Philosophy*. Zalta, E. N. (ed.). Accessed on 18-12-2019 at: <https://plato.stanford.edu/archives/fall2018/entries/capability-approach/>.

Robeyns, I. (2017). *Wellbeing, Freedom and Social Justice: The Capability Approach Re-Examined*. Cambridge: Open Book Publishers.

Ross, W. D. (1930). *The Right and the Good*. Oxford: Clarendon Press.

Sax, M. (2020). Optimization of What? For-profit Health Apps as Manipulative Digital Environments. *Journal of Ethics and Information Technology*. https://doi.org/10.1007/s10676-020-09576-6

Schechner, S. and Secada, M. (2019). You Give Apps Sensitive Personal Information. Then They Tell Facebook. *Wall Street Journal*. Accessed on 28-04-2021 at: https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636

Sen, A., (1985). Wellbeing, Agency and Freedom: The Dewey Lectures 1984. *Journal of Philosophy*. Vol. 82. pp.169–221.

Sen, A., (1992). *Inequality Re-examined*, Oxford: Clarendon Press.

Sen, A. (1999). *Development as Freedom*. New York: Knopf.

Sen, A. (2009). *The Idea of Justice*. London: Allen Lane.

Sharkey, L. (2018). A Twisted Comparison Game: How Fitness Apps Exacerbate Eating Disorders. Accessed on 26-06-2019 at: https://broadly.vice.com/en_us/article/pammjn/a-twisted-comparison-game-how-fitness-apps-exacerbate-eating-disorders.

Sharon, T. (2017). Self-tracking for Health and the Quantified Self: Re-articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. *Philosophy & Technology*. Vol.30. No.1. pp.93-121.

Simon, J., Wong, P.-H., and Rieder, G. (2020). Algorithmic Bias and the Value Sensitive Design Approach. *Internet Policy Review*. Vol.9. No.4. pp.1-16.

Smids, J. (2012). The Voluntariness of Persuasive Technology. In Bang, M., Ragnemalm, E.L. (Ed.). *PERSUASIVE 2012. LNCS*, Vol. 7284. pp.123–132.

Spahn, A. (2011). And Lead Us (Not) into Persuasion...? Persuasive Technology and the Ethics of Communication. *Science and Engineering Ethics*. Vol. 18. No. 4. pp.633-650.

Spiekermann, S. (2015). *Ethical IT Innovation, A Value-Based System Design Approach*. New York: Auerbach Publications.

Steen, M. (2016). Organizing Design for Wellbeing Projects: Using the Capability Approach. *Design Issues*. Vol.32. pp.4-15.

Sterba, J. P. (2005). *The Triumph of Practice over Theory in Ethics*. Oxford: Oxford University Press.

Susser, D., Roessler, B., Nissenbaum, H. (2018). Online Manipulation: Hidden Influence in a Digital World. *SSNR Electronic Journal*. Vol.4. No.1. pp.1-45.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge*. New Haven: Yale University Press.

Thaler, R.H. (2015). The Power of Nudges, for good and bad. *The New York Times*. Accessed on June 8th 2019 at: https://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html

Tiffany, K. (2018). Period-tracking apps are not for women. Accessed on 1-9-2019 at: https://www.vox.com/the-goods/2018/11/13/18079458/menstrual-tracking-surveillance-glow-clue-apple-health.

Toboso, M. (2010). From Disability to Functional Diversity: ICT and Amartya Sen's Approach. In Kalantzis-Cope P. and Gherab-Martín K. (eds). (2010). *Emerging Digital Spaces in Contemporary Society*. London: Palgrave Macmillan. pp.102-104.

Toboso, M. (2011). Rethinking Disability in Amartya Sen's Approach: ICT and Equality of Opportunity. *Ethics and Information Technology*. Vol. 13. pp.107-118.

Turner, B. S. (2006). *Vulnerability and Human Rights*. Pennsylvania: Pennsylvania State University Press.

Venkatapuram, S. (2013). Health, Vital Goals, and Central Human Capabilities. *Bioethics*. Vol.27. pp.271–279.

Weizenbaum, J. (1966). ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM*. Vol.9. pp.36-45.

Winkler, T. and Spiekermann, S. (2018). Twenty Years of Value Sensitive Design: A Review of Methodological Practices in VSD projects. *Journal of Ethics and Information Technology*. DOI: 10.1007/s10676-018-9476-2

Wolff, J. and de-Shalit, A. (2007). *Disadvantage*. Oxford: Oxford University Press.

World Health Organization (2001). Mental Disorders Affect One in Four People. Accessed on 17-12-2019 at: https://www.who.int/whr/2001/media_centre/press_release/en/

World Health Organization (2006). Constitution of the World Health Organization. Accessed on 18-12-2019 at: https://www.who.int/governance/eb/who_constitution_en.pdf

World Health Organization (2015). *World Health Organization Fact Sheet N°310: The Top 10 Causes of Death*. Accessed on 26-06-2019 at: http://www.who.int/mediacentre/factsheets/fs310/en/

World Health Organization (2019). Depression. Accessed on 17-12-2019 at: https://www.who.int/en/news-room/fact-sheets/detail/depression

Wynsberghe, van A. (2013). Designing Robots for Care: Care Centered Value Sensitive Design. *Science and Engineering Ethics*. Vol.19. No.2. pp.407-433.

Wynsberghe, van A. and Robbins, S. (2013). Ethicists as Designer: A Pragmatic Approach to Ethics in the Lab. *Science and Engineering Ethics*. Vol.20. No.4. pp.947-961.

Yetim, F. (2011). A Set of Critical Heuristics for Value Sensitive Designers and Users of Persuasive Systems. *ECIS 2011 Proceedings, Helsinki*.

Zheng, Y. (2007). Exploring the Value of the Capability Approach for E-Development. *Proceedings of the 9th International Conference on Social Implications of Computers in Developing Countries*. São Paulo, Brazil. Accessed on 18-12-2019 at: https://pdfs.semanticscholar.org/25d9/199ff1d14fb70f1d7f946c60d3d82a1b12a3.pdf

Zheng, Y. (2009). Different Spaces for E-Development: What Can We Learn from the Capability Approach? *Information Technology for Development*. Vol. 15. pp.66-82.

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*. Vol. 30. No. 1. pp.75-89.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Flight for a Human Future at the Frontier of Power*. New York: Public Affairs.

Summary

Health-related behavior change technologies help people prevent, or reduce, unhealthy behaviors. But this type of persuasive technologies can give rise to ethical concerns, especially for vulnerable people. As of yet, however, no study in the field of persuasive ethics has explicitly focused on the ethical concerns that arise with the design and use of these technologies for vulnerable people. This is striking because these technologies are designed to help people change their attitudes or behaviors; something that is often particularly valuable for vulnerable people in order to better cope with their vulnerabilities.

In this thesis, I discuss the ethical concerns that arise with the design and use of health-related behavior change technologies for vulnerable people. I argue that ethics by design can proactively remedy these concerns. The dissertation starts with the identification of the two most prominent ethical concerns that arise for vulnerable people making use of health-related behavior change technologies. These two ethical concerns are: (1) sufficiently accounting for users' needs and interests and (2) respecting users' autonomy. Subsequently, I argue that designers could account for these concerns by means of 'ethics by design'. Ethics by design consists of the idea that design decisions shape and affect the set of interactions and constraints of a technology to users and that these decisions can support or undermine ethical values. By proactively articulating value considerations at the start of a technology design process, designers can account for the values that they build into their design.

The most prominent and influential ethics by design approach is Value Sensitive Design (VSD). Although VSD is a promising approach, I identify multiple challenges that VSD faces and argue that all of these challenges arise from the fact that VSD lacks a solid normative foundation. Subsequently, I argue that VSD can overcome these challenges when VSD is complemented by an ethical theory. That is because an ethical theory can provide sources of justification and argumentation for moral claims and considerations, which are needed to make principled judgments, to attend to a set of bounded and principled values, and to legitimize value trade-offs during the design process.

Subsequently, I argue that in the context of technology design for health and wellbeing promotion, Martha Nussbaum's capability theory is a suitable ethical theory to complement VSD. This complementation, as I put forward in this dissertation, results in the design framework Capability Sensitive Design (CSD); a framework that is able to overcome most of VSD's challenges and that is able to normatively assess technology design in general, and technology design for health and wellbeing for vulnerable people in particular.

Ultimately, the dissertation explores how to bridge the theory-practice gap by entering into dialogue with various design-experts on ethics by design in general, and CSD in particular. An empirical study, consisting of thematic interviews with nine design-experts, is conducted in order to explore design-experts' experiences with ethics by design, to understand what they regard as the strengths and weaknesses of CSD, and to see if CSD could be of practical use to their design practices.

Samenvatting

Persuasieve gezondheidstechnologieën kunnen mensen helpen om ongezond gedrag te voorkomen of verminderen. Dit type technologie brengt echter ook ethische zorgen met zich mee, met name voor kwetsbare mensen. Tot nu toe is er in het domein van de persuasieve ethiek nog geen onderzoek gedaan naar de ethische zorgen rondom het ontwerp en gebruik van persuasieve technologieën specifiek voor kwetsbare mensen. Dat is opvallend aangezien deze technologieën zijn ontworpen om mensen te helpen hun gedrag te veranderen, wat met name voor kwetsbare mensen waardevol kan zijn om beter met hun kwetsbaarheden om te gaan.

In dit proefschrift bespreek ik de ethische problemen die met het ontwerp en gebruik van persuasieve gezondheidstechnologieën kunnen ontstaan voor kwetsbare mensen. Vervolgens beargumenteer ik dat *ethics by design* deze problemen proactief kan oplossen.

Om te beginnen identificeer ik de twee meeste prominente ethische problemen die ontstaan voor kwetsbare mensen als zij gebruikmaken van persuasieve gezondheidstechnologieën. Deze twee ethische problemen zijn: (1) rekening houden met de behoeften en belangen van de technologie gebruikers, en (2) het respecteren van de autonomie van gebruikers.

Ik beargumenteer vervolgens dat de ontwerpers van persuasieve gezondheidstechnologieën het ontstaan van deze ethische problemen grotendeels kunnen voorkomen door ethics by design toe te passen. Ethics by design is gestoeld op het idee dat ontwerpkeuzes invloed hebben op de interacties tussen een technologie en haar gebruiker, en dat die ontwerpkeuzes morele waarden kunnen ondersteunen of ondermijnen. Door proactief op morele waarden te reflecteren aan het begin van een technologisch ontwerpproces kunnen ontwerpers expliciet morele waarden in hun ontwerp implementeren.

De meest prominente en invloedrijke *ethics by design* approach is Value Sensitive Design (VSD). VSD is een veelbelovend maar ik laat in hoofdstuk 3 zien dat er meerdere problemen zijn met VSD die allemaal voortkomen uit het gegeven dat het VSD aan een solide normatieve basis ontbreekt. Vervolgens beargumenteer ik dat deze

problemen opgelost kunnen worden door VSD te combineren met een ethische theorie. Een ethische theorie zorgt namelijk voor een normatieve basis van waaruit solide argumentatie voor- en rechtvaardiging van morele overwegingen gemaakt kunnen worden. Dit is nodig bij het maken van weloverwogen besluiten en bij het legitimeren van *trade-offs* tussen waarden in het ontwerpproces.

Vervolgens beargumenteer ik dat in de context van technologieontwerp voor de promotie van gezondheid en welzijn de capability theorie van de filosoof Martha Nussbaum bij uitstek een geschikte ethische theorie is om VSD te vergezellen. De combinatie van Nussbaum's capability theorie met VSD resulteert in het design framework Capability Sensitive Design (CSD). In hoofdstuk 4 presenteer ik CSD; een framework dat in staat is om technologieontwerp in het algemeen normatief te evalueren, en technologieontwerp voor gezondheid- en welzijn technologieën voor kwetsbare mensen in het bijzonder. Bovendien, zo beargumenteer ik, is CSD in staat om de meeste problemen van VSD op te lossen.

Tot slot onderzoek ik hoe we het beste een brug kunnen slaan tussen theorie en praktijk. In hoofdstuk 5 ga ik in gesprek met verschillende ontwerp-experts over *ethics by design* in het algemeen en CSD in het bijzonder. Met een empirische studie, bestaande uit thematische interviews met negen ontwerp-experts, breng ik de ervaringen van ontwerpers met *ethics by design* in kaart, onderzoek ik wat zij ervaren als de voor- en nadelen van een framework zoals CSD en onderzoek ik welk praktisch potentieel zij zien van CSD in hun eigen ontwerppraktijk.

About the Author

Naomi Jacobs (1990) completed her PhD in Ethics of Technology at Eindhoven University of Technology (TU/e) between 2016 and 2021. During her time at TU/e she has worked as a PhD candidate and teaching assistant. Before coming to TU/e she graduated from the Research Master Philosophy at Utrecht University with her thesis 'On the Viability of a Virtue Ethics Approach to Bioethics'. Early 2020, she has been a research fellow at the philosophy department of the Royal Institute of Technology at Stockholm, Sweden, where she worked with prof. dr. Barbro Fröding. In the summer of 2021, she will start as an Assistant Professor in bioethics at the philosophy department of the University of Twente.

List of Publications

Academic Publications

Jacobs, N. and IJsselsteijn, W. (2021). Bridging the Theory-Practice Gap: Design-Experts on Capability Sensitive Design. *International Journal of Technoethics*. Vol.12. No.2. pp.1-16.

Jacobs, N. (2020). Capability Sensitive Design for Health and Wellbeing Technologies. *Science and Engineering Ethics*. Vol. 26. No.6. pp.3363-3391.

Jacobs, N. (2020). Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People. *Bioethics.* Vol. 34. No.5. pp.519-526.

Jacobs, N. and Evers, J. (2019). De Kwetsbaarheid van Femtech. *Podium voor Bio-Ethiek*. Vol. 26. No.4. pp.13-15.

Jacobs, N. and Huldtgren, A. (2018). Why Value Sensitive Design Needs Ethical Commitments. *Journal of Ethics and Information Technology*. Vol.23. No.1. pp.23-26.

Professional and Popular Publications

Jacobs, N. (2019). Een Museum voor het Ongeluk. In *Mobiliteit Ontmoet Filosofie*. Uitgave van Kennisplatform CROW en Rijkswaterstaat Ministerie van Infrastructuur en Waterstaat.

Jacobs, N. (2018). Boven op de Toverberg. In *Onszelf voorbij, kijken naar wat we liever niet zien*. (Ed. Doeland, L., Jacobs, N. en de Mul, E.). Amsterdam: Uitgeverij de Arbeiderspers.

Jacobs, N. (2017). Viering van het Vreemde. In *Ruimte voor een Stad in Balans: Essays over Rotterdam*. Uitgave van Gemeente Rotterdam.

Jacobs, N. (2015). Mixofobie en de Stad. In *Filosofen Agenderen de Stad*. Uitgave van Platform 31 en het Ministerie van Binnenlandse Zaken.

Simon Stevin Series in Ethics of Technology
Delft University of Technology, Eindhoven University of Technology, Univeristy of Twente & Wageningen University
Editors: Philip Brey, Anthonie Meijers, Sabine Roeser and Marcel Verweij

Books and Dissertations

Volume 1: Lotte Asveld, *'Respect for Autonomy and Technology Risks'*, 2008

Volume 2: Mechteld-Hanna Derksen, *'Engineering Flesh, Towards Professional Responsibility for 'Lived Bodies' in Tissue Engineering'*, 2008

Volume 3: Govert Valkenburg, *'Politics by All Means. An Enquiry into Technological Liberalism'*, 2009

Volume 4: Noëmi Manders-Huits, *'Designing for Moral Identity in Information Technology'*, 2010

Volume 5: Behnam Taebi, *'Nuclear Power and Justice between Generations. A Moral Analysis of Fuel Cycles'*, 2010

Volume 6: Daan Schuurbiers, *'Social Responsibility in Research Practice. Engaging Applied Scientists with the Socio-Ethical Context of their Work'*, 2010

Volume 7: Neelke Doorn, *'Moral Responsibility in R&D Networks. A Procedural Approach to Distributing Responsibilities'*, 2011

Volume 8: Ilse Oosterlaken, *'Taking a Capability Approach to Technology and Its Design. A Philosophical Exploration'*, 2013

Volume 9: Christine van Burken, *'Moral Decision Making in Network Enabled Operations'*, 2014

Volume 10: Faridun F. Sattarov, *'Technology and Power in a Globalising World, A Political Philosophical Analysis'*, 2015

Volume 11: Gwendolyn Bax, *'Safety in large-scale Socio-technological systems. Insights gained from a series of military system studies'*, 2016

Volume 12: Zoë Houda Robaey, *'Seeding Moral Responsibility in Ownership. How to Deal with Uncertain Risks of GMOs'*, 2016

Volume 13: Shannon Lydia Spruit, *'Managing the uncertain risks of nanoparticles. Aligning responsibility and relationships'*, 2017

Volume 14: Jan Peter Bergen, *Reflections on the Reversibility of Nuclear Energy Technologies*, 2017

Volume 15: Jilles Smids, *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis*, 2018

Volume 16: Taylor William Stone, *Designing for Darkness: Urban Nighttime Lighting and Environmental Values*, 2019

Volume 17: Cornelis Antonie Zweistra, *Closing the Empathy Gap: Technology, Ethics, and the Other*, 2019

Volume 18: Ching Hung, *Design for Green: Ethics and Politics for Behavior-Steering Technology*, 2019

Volume 19: Marjolein Lanzing, *The Transparent Self: a Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*, 2019

Volume 20: Koen Bruynseels, *Responsible Innovation in Data-Driven Biotechnology*, 2021

Volume 21: Naomi Jacobs, *Values and Capabilities: Ethics by Design for Vulnerable People*, 2021

Simon Stevin (1548-1620)

'Wonder en is gheen Wonder'

This series in the philosophy and ethics of technology is named after the Dutch / Flemish natural philosopher, scientist and engineer Simon Stevin. He was an extraordinary versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbours, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview, which did not make him popular in religious circles. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically he was a pragmatic rationalist for whom every phenomenon, however mysterious, ultimately had a scientific explanation. Hence his dictum 'Wonder is no Wonder', which he used on the cover of several of his own books.