We are faced with a very serious environmental crisis. Resource depletion, air pollution, biodiversity loss, climate change, etc., accompany each other, resulting largely from our current way of life. Manufacturing of nature-friendly products is far from being able to protect the environment; technologies can and should be made to change people's environmental behavior. However, designing and implementing such technologies tend to provoke ethical and political concerns about human freedom and autonomy. Determining how to take advantage of behavior-steering technology without causing worries, therefore, is unavoidably a key challenge to practitioners as well as theorists. This dissertation takes on this challenge by synthesizing philosophical insights about human-technology interactions, psychological explanations for the nature of human behavior, and political theories concerning democracy and social reform. It not only offers a conceptual basis in arguing for the necessity and feasibility of behavior-steering technology, but also develops a piecemeal-behavioral approach to realizing this goal. Through the chapters of this dissertation, it becomes clear that, above all, to address the environmental crisis eventually requires us to conceive of a more realistic image of human beings and to rebuild contemporary ethical and political frameworks based upon it.
Design for Green

Ethics and Politics for Behavior-Steering Technology
DESIGN FOR GREEN

ETHICS AND POLITICS FOR BEHAVIOR-STEERING TECHNOLOGY

DISSERTATION

To obtain
the degree of doctor at the University of Twente,
on the authority of the Rector Magnificus,
prof.dr. T.T.M. Palstra,
on account of the decision of the Doctorate Board,
to be publicly defended
on Wednesday 13 November 2019 at 10:45 hrs

by Ching HUNG

born on 22 September 1982
in Chiayi, Taiwan
# Contents

Acknowledgments vii

1 Introduction 1
   1.1. Houston, We Have a Problem 1
   1.2. Technology and the Environment 2
   1.3. Technology and Users 4
   1.4. Objectives and Research Questions 5
   1.5. Outline of the Dissertation 6

2 A Focal Shift from Knowing to Doing 11
   2.1. Environmental Education: An Unsuccessful Attempt 11
   2.2. The Gap between Knowing and Doing 13
   2.3. The Asymmetry Concerning Anthropocentrism 17
   2.4. A Possible Way Out 21
   2.5. Concluding Remarks 24

3 Classifying Behavior - Steering Technology 25
   3.1. The Classification According to What is Mediated 25
   3.2. The Classification According to the Modes of Interaction 27
   3.3. The Classification According to Force and Salience 29
   3.4. A Classification Made for Behavior-Steering Technology 31
   3.5. Informational Behavior-Steering Technology 32
   3.6. Material Behavior-Steering Technology 37
   3.7. Concluding Remarks 40

4 Ethical Concerns about Behavior-Steering Technology 43
   4.1. The Issue of Human Freedom 43
   4.2. Ethical Guidelines for Persuasive Technology 45
   4.3. Defense for Nudge 48
   4.4. In Need of Material Behavior-Steering Technology 53
   4.5. Human Freedom under Material Behavior-Steering Technology 56
   4.6. Concluding Remarks 59

5 Technology of Behavior: B. F. Skinner’s Thoughts and Defense 61
   5.1. A Science of Behavior 61
   5.2. How to Change Human Behavior: Walden Two 63
   5.3. Against Worries: Beyond Freedom and Dignity 68
Acknowledgments

A PhD is a long journey, especially for me as an overseas student. The cultural difference between Taiwan and the Netherlands even makes it more difficult. I could never arrive at the end without many people’s support and help. This is a note of gratitude to all of them.

First of all, I am unbelievably lucky and thankful to have Peter-Paul Verbeek as my supervisor and Michael Nagenborg as my co-supervisor. This work could not have been done probably without their unfailing faith in me and my research project. I am deeply indebted to Peter-Paul, who has always held a positive attitude toward my thoughts, responded in time to my requests concerning the official part of getting a PhD, and corrected the Dutch summary of this dissertation with carefulness for me. To Michael, who did best to put me on the right track and pointed me to Skinner’s work, I owe a sincere thank you for his constructive comments and practical suggestions on every version of my thesis draft. The teamwork of them struck a balance between giving me independence to pursue my interests and providing invaluable guidance.

I would also like to extend my genuine gratitude to the independent members of my doctoral committee: Philp Brey, Geke Ludden, Ibo van de Poel, Wijnand IJsselsteijn, and Daiwie Fu in particular, who said yes to a 14-hour flight from my homeland. I feel really grateful and honored that you all accepted to evaluate my work and to participate in my defense.

My thanks also go to two Dutch and two Taiwanese institutions which offered financial assistance in the past few years. Both the Department of Philosophy at Twente, as Ciano Aydin as its head, and the 4TU.Ethics, as Philip Brey as the director, have supported me with travel funds, making my attendances and presentations at several conferences and workshops possible. The scholarships from Taiwan Ministry of Education and the Delta Electronics Foundation (partnered with the Netherlands Trade and Investment Office) provided a basic financial condition, allowing me to start my research as an external PhD in the Netherlands.

During the years of my study, I have enjoyed the time with my PhD colleagues in our research group of human-technology relations. Many thanks to Jonne Hoek, Bas de Boer, Olya Kudina, and Melis Bas; our (academic) trip to Japan in 2016 was unforgettable, especially the Korean BBQ dinner and the karaoke night with Peter-Paul at Tokyo. I wish to thank Margoth González Woge
as well, who made a joint presentation with me at the Persuasive Technology 2017 conference. In addition to them, I would also like to thank Agata Gurzawska, David Lyreskog, Sophie van Baalen, and May-Li Mertens, the departmental PhD colleagues with whom I had pleasant get-togethers and chats.

The intellectual interactions with the PhDs from outside the Netherlands were also enjoyable. A thank you to Lumeng Jia, who discussed several times with me during her one-year visit at Twente in the academic year 2016. Also, a thank you to Shinya Oie, for our few but congenial discussions, particularly the one we did with a paper full of drawings and words at a local cafe in Osaka. I feel lucky too to have met Zhanxiong Liu and Pin Yan, with whom I have candid conversions about academic and personal life in the future.

Additional thanks should go to the departmental secretaries. Carola Bouwens in the beginning of my study, Petra Bruulsema during the years, and Ada Krooshoop for the closure, all have helped me with and through administrative procedures. I also appreciate Ning Huang’s effort to put the dissertation into the right format, making the layout clean and clear.

Last but certainly not least, I would like to thank my loved ones. They have been very patient with me, especially at times when I had to prioritize the research over everything else in my daily life. I owe a huge debt of gratitude to my dear parents and sister Hsin Hung for their endless warmth and support throughout and beyond this challenging work. A special thanks is reserved for my partner Ruijie Tang, who has given me heartfelt love, unfading encouragement, and wonderful companionship—it is about time for us to move onto the next stage of our life.
1 Introduction

*Actions speak louder than words.*

— An old English proverb

1.1. Houston, We Have a Problem

We are faced with a very serious environmental crisis. A host of environmental problems are generally associated with one another: resource depletion, air pollution, biodiversity loss, and extreme weather, to name only a few. Among them, climate change is probably the most pressing, regularly headlined problem in recent decades. The Intergovernmental Panel on Climate Change has made clear that “[i]n recent decades, changes in climate have caused impacts on natural and human systems on all continents and across the oceans” (IPCC 2015, 49). The rise of global temperature by just 1 or 2°C is enough to be menacing:

Some ecosystems and cultures are already at risk from climate change (*high confidence*). With additional warming of around 1°C, the number of unique and threatened systems at risk of severe consequences increases. Many systems with limited adaptive capacity, particularly those associated with Arctic sea ice and coral reefs, are subject to very high risks with additional warming of 2°C. (IPCC 2015, 72, italics in original)

Things can be much worse if the temperature goes higher:

The risks associated with temperatures at or above 4°C include severe and widespread impacts on unique and threatened systems, substantial species extinction, large risks to global and regional food security, consequential constraints on common human activities, increased likelihood of triggering tipping points (critical thresholds) and limited potential for adaptation in some cases (*high confidence*). (IPCC 2015, 77, italics in original)

The warming of the climate system is unequivocal. Although there are still some deniers and skeptics, the confirmation of such a problem is not seriously challenged because “the issues of scientific certainty most discussed by climate skeptics usually deal with the magnitude and timing of climate change, not with whether global warming is a real threat” (Brown 2002, 102). Moreover, not only
are the hazards of climate change widely acknowledged, but the main causes are also uncontroversial: it is humankind that is responsible for most environmental problems, including climate change (Doran and Zimmerman 2009). Since the 1750s, the concentration of CO₂ (carbon dioxide), the most famous greenhouse gas, has been increasing rapidly due to human activity. Numerous products geared to meeting human needs and wants are manufactured through industrial processes that emit CO₂. Animal husbandry to provide meat and milk for human consumption also creates a huge amount of CO₂ simply due to the breathing of animals, whereas animal feces can release other greenhouse gases such as methane. Most of the energy used by humans is produced by burning fossil fuels (coal and oil), which also emit a great deal of CO₂. How to mitigate climate change has become a primary issue for many countries.

1.2. Technology and the Environment

It is no accident that the beginning of dramatic CO₂ accumulation is often dated to the 18th century (Zalasiewicz et al. 2008). All the human activities mentioned above cannot be efficiently and effortlessly practiced without the help of technology, and the first explosive development of technology occurred exactly then, during the Industrial Revolution. Since then, as human productivity has increased, energy consumption and resource exploitation have grown considerably. Unsurprisingly, this close relationship between technological development and the environmental crisis implies that the latter can be attributed to the former. This is probably why, throughout the 20th century, anti-war protests and environmental movements were usually associated with one another. After seeing how technologically advanced weapons caused mass casualties in World War I and II, people also lost their faith in technology. The negative attitude toward technology, therefore, gradually spread to many societies and cultures.

Academia was no exception to such an atmosphere. The field of philosophy at that time also demonstrated a pessimistic view toward technology. For instance, Martin Heidegger, whose philosophy of technology is now identified as classical, saw modern technology as “enframing” (Gestell) in which the world becomes a “standing-reserve” (Heidegger 1977). Stated more simply, modern technology reveals everything (human beings included) as, or turns everything into, a resource ready for exploitation. As shown by Heidegger’s famous example, building a hydroelectric dam on the Rhine River would turn it into an energy resource, rather than a source for philosophical inspiration or cultural
pride. Likewise, Hans Jonas suggested that modern technology can be so powerful in bringing harm to nature and human beings that we need to formulate ethical principles which obligate us to take responsibility (Jonas 1984). This pessimistic view of technology is still prevalent. Today, most people view technology as a part of the crisis rather than a solution to it. However, it is likely unrealistic to think of returning to a world with only primitive forms of technology. It is more practical and more feasible to find out what technology can do to help with environmental problems.

Technology is a double-edged sword. While not denying that some technology has damaged the environment, optimistic engineers and designers believe that technology can be continuously improved by removing its negative effects. If paper has to be made from trees, let us develop a technology that can produce and increase the quality of recycled paper. If cars emit too much CO$_2$, we need to make them more fuel efficient. If household appliances over-use electricity, we should increase their energy efficiency through improved technology. If buildings are not carefully designed to dissipate heat in summer and retain heat in winter, we must reconsider their forms, structures, and materials, and use improved building technologies. Technological products which have unintentionally been made environmentally unfriendly can be improved. Implementing nature-friendly technology, therefore, is a fix for the environmental crisis and a hope for our future.

Unfortunately, sometimes nature-friendly products still produce unfriendly results. For example, the user of an energy-saving heater may keep the heater turned on for longer periods than the user of a general heater. Energy-saving heaters tend to give users the mistaken impression that they are saving energy while running. That is to say, nature-friendly products may mistakenly relieve users from adhering to nature-friendly practices by misleading them in the way they use a product. Such a phenomenon is called the “rebound effect” and can be very troublesome (Tenner 1997). When the gains created by implementing nature-friendly technology are reduced, or overridden, by the losses incurred by improper use of that technology, the total benefit may be questionable. The rebound effect, accordingly, shows the limits and flaws inherent in applying nature-friendly technology.
1.3. Technology and Users

While nature-friendly technology looks for solutions to environmental problems, it tends to leave out lifestyle considerations. Nature-friendly products do not question users’ purposes or expectations in using them, such as comfort and convenience. As a result, there is nothing changed in user behavior—people still do what they have been doing. This is where the rebound effect sneaks in and often makes the effort ineffective. Attempting to implement nature-friendly technology apparently does not necessarily guarantee nature-friendly practices by users (Midden 2006). Given that the net result is the sum of technology and the practices of users rather than either of them individually, we need to pay attention to the habits of users as well. If people walk or bike more, CO₂ emissions will decrease even though the latest emission-reducing technology may not have been added to gasoline-powered vehicles. If people regularly turn off heaters upon leaving their homes, energy can be saved even though their heaters may be too old and inefficient to receive Energy Star¹ labeling. To tackle environmental problems, we need nature-friendly users in addition to nature-friendly technology.

People’s behavior has to be directed toward a more sustainable way of life. But can technology help to change users’ environmental behavior? If so, how? Let me provide a case study I completed several years ago (Hung 2009). In Taiwan, there is a four-floor office building occupied by an electronics company (Delta Electronics) designed and constructed to be “green”.² It is the first office building to have been awarded the “Gold-Level Certification for Green Architecture” for its accomplishment in reducing energy consumption. Besides the use of water-saving and energy-efficient appliances and ecological materials, what contributes most is the design of the stairs and the elevator. The stairs conspicuously colored with bright yellow, orange, and jacinth are located at the center of the entrance hall. In contrast, the elevator is installed away from the entrance hall, in a dark, inconspicuous corner. Such a design significantly improves use of the stairs and reduces that of the elevator. In other words, most of the employees and visitors take the stairs rather than using the elevator. As a

¹ Energy Star is a symbol for energy efficiency, backed by American government. See the website: https://www.energystar.gov (accessed September 11, 2017).

² To avoid word repetition, the words “sustainable”, “green”, and “nature-friendly” are used interchangeably in the present dissertation.
result, the building saves a huge amount of energy, and monthly electricity bills are usually much lower than those of other conventional office buildings. In this case, sustainability is achieved not only by using nature-friendly products but also by influencing user behavior.

Technology, as we learned from the case above, can be designed to guide, shape, or change user behavior. Such technology is usually called behavior-steering technology, defined as “a technology in which one of its main functions is to make its users behave in a way that is not necessarily desired by the user but that is desired by some other party in control of the technology” (Brey 2006, 357). Behavior-steering technology, therefore, can be a complement to the approach of using nature-friendly technology. Rather than creating nature-friendly technology, this complement aims to create nature-friendly users through technological design.

However, not everyone agrees upon the approach of using behavior-steering technology to achieve sustainability. Here let me use my personal experience as an example. Several years ago, I joined two conferences to present my further study on the case of the stairs-elevator design (Hung 2013). At one conference, some audiences expressed their concerns about technological power: “Don’t you think it’s a power over people?” and “The approach you are suggesting is problematic, because human freedom is invaded by purposive technology!” At the other conference, a different type of concern was raised: “Those users are not really sustainable; their pro-environmental behavior is not ‘heartful!’” and “How do you know such an approach will not be abused for the opposite purpose?” All these objections and doubts posed a serious challenge to the approach of using behavior-steering technology to achieve sustainability. To be honest, I had no idea how to respond on both occasions; later however, these questions motivated me to think further and became the starting point of this research.

1.4. Objectives and Research Questions

This research is an exploration of the approach focused on behavior-steering technology. With an eye on the environmental crisis, two objectives can be juxtaposed: on the one hand, I would like to develop a theoretical basis on which the approach of behavior-steering technology for sustainability can be well established; on the other hand, I shall offer design recommendations by which engineers and designers can put the approach into practice. To these ends, research questions are formulated as follows: 1) Why is behavior-steering
Design for Green

technology necessary for mitigating environmental problems? 2) What kind of behavior-steering technology is most helpful and why is that? 3) How can we respond to ethical concerns that tend to emerge concerning the design and implementation of behavior-steering technology? 4) What would a political framework that can accommodate the practice of behavior-steering technology look like?

As one may tell from the objectives mentioned above, this research is largely oriented toward problem-solving. Moreover, the research questions point toward implementing an interdisciplinary style, so to answer each question in detail, I shall not limit myself to any specific academic area. Any study that helps find answers to the research questions, be it from the field of engineering, humanities, or social sciences, can be the resource to seek out. Indeed, as we will find in the coming chapters, theories from philosophy of technology, psychology, behavioral economics, philosophy of science, and political philosophy will be combined to synthesize a general defense for the approach of behavior-steering technology.

1.5. Outline of the Dissertation

To answer the research questions, the rest of the present dissertation comes in eight chapters. Chapter 2 aims to establish as basis for the following chapters to develop further arguments for behavior-steering technology. By inquiring into the “unsuccess” of environmental education, I will argue that use of behavior-steering technology is a necessary part of improving people’s environmental behavior. Environmental education, as a widely adopted strategy dealing with the environmental crisis, represents a general but problematic tendency that neglects the gap between what people know and what people do. Such a tendency, as I will argue, comes mainly from an asymmetric effort in overcoming anthropocentrism. While the “knowing” dimension of anthropocentrism has been largely corrected by environmental ethics, the “doing” dimension has not yet been touched. Here, the concept of the Anthropocene will be a hint to make non-doing-anthropocentrism possible. People’s behavior will be proposed as the direct area of focus; behavior-steering technologies will suggest a more practical

---

3 I use the word “unsuccess” instead of “failure” because environmental education is still an ongoing practice and we cannot deny the possibility that some day it may become successful. However, the result of such a practice so far is not satisfying.
approach to dealing with environmental problems as a complement to the problem-solving benefits of environmental education.

Before taking advantage of behavior-steering technologies, it is essential to have knowledge of their types, forms, and mechanisms. Therefore, in Chapter 3, I will introduce three classifications outlining the relationships between humans and technology. Although they can equip us with basic understanding about behavior-steering technologies, they appear too complicated to fit into the present dissertation and may miss something important in explaining how behavior-steering technologies work. I then will suggest a new but simpler classification, useful for the development of further augments. It categorizes behavior-steering technologies as one of two types by distinguishing their mechanisms: One is “informational” and the other “material”. The difference in mechanisms results in establishing the pros and cons of each type. While the former, including persuasive technology and nudge, is mild and more welcomed by designers, the latter is much more effective but also provokes concern. With several examples provided, it will become clear that the degree of ethical concern regarding a behavior-steering technology is correlative to the mechanism it follows.

Following the classification above, Chapter 4 will elaborate on the ethical concerns of each type and how they have been received. The main issue here concerning behavior-steering technologies is the loss of users’ freedom. As we shall see, to protect a user’s freedom or to ease related worries, there have been ethical guidelines developed for persuasive technology and defensive arguments constructed for nudge. However, they may restrict the potential of behavior-steering technologies to change environmental behavior and therefore are not at all effective in tackling the environmental crisis. As I will argue, on the one hand to improve environmental behavior unavoidably requires intervening in the behavior of the users of a certain technology. On the other hand, informational behavior-steering technologies are far from enough to produce a significant change. In other words, the approach of using material behavior-steering technology is indispensable. This, however, leads us to a serious dilemma: It becomes not possible to design and implement truly effective behavior-steering technologies without invading users’ freedom. Such a dilemma implies, as I will point out finally, a need to revise our current understanding of human behavior.

I will take upon the challenge in Chapter 5 by turning to the work of the psychologist B. F. Skinner, who scientifically studied closely and carefully the relationship between organism behavior and the environment where the behav-
ior takes place. The fact that changing experimental settings can change animal behavior implies that human behavior could be similarly altered by providing a revised environment. Skinner’s novel, *Walden Two*, explains how this idea works and offers practical suggestions. This book has provoked serious objections that are very similar to those against the approach of behavior-steering technology. His later book, *Beyond Freedom and Dignity*, argues against the objections by questioning the nature and usefulness of the conceptions of human freedom and dignity. Skinner’s thoughts and defense, as I will argue, can help us in many ways. On the one hand, his theory of behavior enables us to further understand why environmental problems can hardly trigger behavioral changes. On the other hand, by taking his “behavioral engineering”, we can largely ease worries about users’ freedom and thereby make room for the approach of behavior-steering technology. However, Skinner’s idea is not flawless. In taking advantage of his technique of behavioral modification, we also bump into an epistemological obstacle concerning the environmental crisis: It seems less than possible to justify the design and implementation of green behavior-steering technologies by simply saying that green is good.

To deal with this obstacle, Chapter 6 will begin with two real-world Skinnerian communities. One is Twin Oaks in America; the other is Los Horcones. They were both built to imitate and realize Skinner’s behavioral utopia. However, while the former is stuck in a political predicament, the latter is vigorous and still growing. By comparing their practices, I will show the importance of focusing on behavior rather than social structure and the necessity of adopting a scientific view on human behavior. I will further argue, by probing into Skinner’s design of *Walden Two*4, that due to experimental nature of behavioral engineering, the key to making behavioral change effective is to keep the population of those who are to be changed a manageable size. Accordingly, there is nothing more suitable than a community as a place to apply behavior-steering technologies. At this point, I will connect Skinner’s behavioral engineering to the philosopher Karl Popper’s “piecemeal engineering”. While the former offers practical reasons for staying small, the latter points us to the epistemological limitation of planning large-scale social reform. Popper’s epistemological arguments not only strengthen the approach of Skinner, but also help us to overcome the obstacle left by the previous chapter. Referring to Popper allows us to alter

4 When “Walden Two” is not in italic, it refers to the fictional community itself, not the novel.
people’s environmental behavior without knowing whether green is good. Such a piecemeal-behavioral form of engineering can underpin the approach of applying behavior-steering technology to the environmental crisis.

To make the approach practical and workable, in Chapter 7 I will synthesize the points having been made and formulate a list of seven design recommendations as a take-away strategy for designers. Moreover, I will apply the list to a famous green community, Village Homes, analyzing its designs and practices. As we shall see, the community embodies both the idea of piecemeal-behavioral engineering and the approach of behavior-steering technology. I will focus on its traffic design, arguing that it is crucial to the achievement of the community. As a material behavior-steering technology, it does not only have a prominent effect in improving the residents’ environmental behavior but also makes the other nature-friendly design possible. Without such a concrete, tangible, and inexorable design, the whole community would never be green. Village Homes vividly demonstrates the usefulness and feasibility of the concepts, perspectives, and arguments developed in and through the previous chapters for tackling environmental problems.

Chapter 8 attempts to take the whole discussion to the level of politics. The most serious challenge that the environmental crisis poses concerns the current political system. While tackling the environmental crisis requires people to act collectively, our liberal democracy prioritizes the freedom of individuals to behave in their own way. The practice of behavior-steering technology is also limited in similar fashion. To enable further action and to enlarge the possibility of behavior-steering technologies, I will on the one hand point out that we have been building our politics on a problematic image of humans, and on the other hand, by referring to the political theorist Chantal Mouffe, propose “agonistic democracy” as a potential substitution for liberal democracy. Such a democracy, as it neither looks for inclusive consensus nor denies power as an essential to human society, is not only compatible with a better understanding of human behavior, but is also full of potential to widen the scope for the design and implementation of behavior-steering technologies.

To conclude in Chapter 9, I will initially run through the entire line of my argument by recapitulating chapter by chapter. It shall become clear that for tackling environmental problems, the approach of behavior-steering technology developed in this dissertation offers not only flexibility for design activities but also tenacity in refuting objections. In addition, I will also put this research into a broader context, discussing its implications for sustainability advocacy, design
Design for Green

in general, and technology studies. Three concepts: “educational environment”, “design to discourage”, and “politicalizing technology”, will be suggested as a starting point for future work.
2 A Focal Shift from Knowing to Doing

The environmental crisis is largely a product of human activities. The lifestyle of many people today is highly energy-consumptive and waste-productive. Changing people’s environmental behavior has become a primary task for the advocates of sustainability; a conventional strategy for such a task is to awaken people’s “environmental consciousness”. However, as I will argue firstly in this chapter, this strategy, environmental education as its representative, often runs into difficulties because it relies too much on the assumption that what people know is the generator of what people do. As a consequence, it neglects not only the gap between human “knowing” and “doing” but also non-human factors that impact human behavior. Moreover, while anthropocentrism has been largely rectified in the dimension of human “knowing”, the “doing” dimension of anthropocentrism is still there, causing an asymmetric employment of technology between humans and nature. By referring to the concept of the Anthropocene, I will argue that it is necessary to look at human behavior itself in order to overcome such asymmetry. Accordingly, behavior-steering technology will be proposed as a promising approach to solve environmental problems.

2.1. Environmental Education: An Unsuccessful Attempt

Compared to well-established subjects, such as physics and sociology, the history of environmental education is relatively short. It can be traced back to the early 20th century, when its two predecessors were primarily established in North America. The first is “nature study” in the 1910s, teaching students to appreciate and embrace nature through fables or moral lessons; the second is well known as “conservation education” in the 1930s, which focused more on scientific methods to manage and solve environmental issues. Catalyzed by the environmental movements in the late-1960s and early-1970s in America, the two predecessors converged into a clearly defined field that supported, as well as learned from, the movements. Its name, given by William B. Stapp, who is now credited as its founder, appeared for the first time in the first issue of its official journal in 1969:
This new approach, designed to reach citizens of all ages, is called “environmental education.” We define it in this way: Environmental education is aimed at producing a citizenry that is knowledgeable concerning the biophysical environment and its associated problems, aware of how to help solve these problems, and motivated to work toward their solution. (Stapp 1969, 31)

Three years later (1972), Stapp was appointed as the first Director of Environmental Education for UNESCO (United Nations Educational, Scientific and Culture Organization). In the same year, the UN Conference on the Human Environment declared that environmental education is a must-use approach to handle global environmental problems. These are the indicators that the goal and function of environmental education have been widely recognized and highly trusted. Currently, environmental education includes not only school-based teaching, but also other efforts to educate the public, such as community activities and mass media campaigns. Although the summary above seems too brief to catch the details, it is sufficiently clear to show that environmental education represents the common strategy that has been, and still is, adopted to deal with the problem of people’s unsustainable lifestyle.

It is believed that the practice of environmental education can lead to positive changes in people’s environmental behavior (Bamberg and Möser 2007; Wals et al. 2014). In other words, environmental education can be regarded as successful if the people who have been taught by it start biking or walking rather than driving, taking the stairs rather than elevators, eating vegan rather than meat, and so on. Ideally, as long as such changes (the demand side) can be achieved, the problem of over-production (the supply side) will be solved as well. Environmental education, therefore, is tasked as the role of fulcrum, by which small changes in people’s behavior can produce a world of sustainability.

However, the small changes planned and promoted by environmental education have not arrived yet. A study, which counts observed behaviors instead of self-reported ones, shows that only a small number of learners who joined the lectures of environmental education altered their anti-environmental behavior afterward (Osbaldiston and Schott 2012). Another study also found that, among the students who have obtained environmental literacy through environmental

---

5 The word “biking” or “cycling” used in this dissertation refers mainly to “riding bicycles as a means of transport”; it is not about using bicycles for recreation, exercise, or sport, nor has it to do with motorcycles.
education, few behave pro-environmentally (Hsu and Roth 1998). Similarly, most learners who were given outdoor-experiencing classes do not begin behaving in a nature-friendly way (Thapa 2010). Evidently, environmental education does not work as well as expected, and the goal to improve people’s environmental behavior has not yet been reached.

Why is environmental education incapable of improving people’s environmental behavior and what can be done to help it do so?

2.2. The Gap between Knowing and Doing

The root of environmental education’s lack of success is its optimistic reliance on the assumption that what people know will determine what people will do. Here, “what people know” refers not only to knowledge about, but also attitudes toward, nature. These two elements have been embodied respectively in the two predecessors of environmental education mentioned above. While the nature studies looked for solutions to people’s attitudes, the conservation education appealed to scientific knowledge about the world of nature. Environmental education tasks itself with the integration of both, thereby improving people’s environmental consciousness. This task has been clearly stated by an international declaration:

Environmental education is a process aimed at developing a world population that is aware of and concerned about the total environment and its associated problems, and has the attitudes, motivations, knowledge, commitment and skills to work individually and collectively towards solutions of current problems and the prevention of new ones. (Tbilisi Declaration 1977, emphasis added)

It is not hard to tell from the quotation above that environmental education presumes a seamless flow from environmentally “informedness” to environmental behavior. Such a flow begins from knowledge, moves to attitude or motivation, and ends at behavior. For example, if a person is informed with scientific data that his/her car’s CO₂ emission has already killed several polar bears, s/he will be motivated to bike or walk. Or, if a student is taught that trees have intrinsic value in themselves, s/he is likely to feel sympathy and therefore print his/her

---

6 For gender-neutrality, I will use the words “s/he” and “his/her” in this dissertation.
homework on both sides of paper. Such a flow can be illustrated as a simplified model below (Figure 1):

![Diagram of the flow: Knowledge (facts and values about nature) → Attitude → Motivation → Environmental Behavior]

Figure 1. The assumption underlying environmental education

The first arrow in the flow is not implausible. Generally, a person's attitude is closely related to his/her evaluation of someone or something. If a person knows that someone or something is valuable, s/he tends to hold a positive attitude toward them (Petty and Wegener 1998). Moreover, it is argued that ethical principles by themselves have motivational force on the person who makes decisions according to them (Williams 1979). This is probably why environmental ethics is so important to environmental education. From the perspective of environmental education, as long as nature's intrinsic value is successfully argued or ethical principles for sustainability are well established, people will then have positive attitudes toward nature or be motivated to behave pro-environmentally. It is no accident, therefore, that the first two blocks in the figure, as the two predecessors, are integrated into a target for environmental education. This first sub-flow is exactly as mentioned earlier “what people know”; the “knowing” part of the entire flow requires further discussion. To be sure, this knowing part can be split because the arrow within it is somehow not guaranteed. However, the first arrow is rarely questioned by the practitioners of environmental education, and regarding the entire flow, it is not as problematic as the second one.

The last block refers to environmental behavior, which is now called the “doing” part of the entire flow. The second arrow, between the knowing part and the doing part, is where environmental education encounters its difficulties. As several studies have observed, the correlation between the practices of environmental education and the pro-environmental behavior of learners is surprisingly low (Kaiser, Wölfing, and Fuhrer 1999; Staats, Wit, and Midden 1996; 1996; Zelezny 1999). In other words, the improvement of the knowing part is far from

---

Although so-called moral “internalists” in ethics believe that the action of accepting moral principles itself has motivating force, those who are “externalists” usually request other factors to build the arrow (Svavarsdóttir 1999).
producing changes in the doing part. One may doubt those studies, arguing that their results are limited to their samples or that they just chose the wrong groups as the samples. However, a survey can refute this doubt. It was held in 22 countries and found that the high levels of environmental attitudes coexist with the low levels of environmental behavior (Dunlap, Gallup, and Gallup 1993). One may raise another doubt, saying that the above result is because the improvement of the knowing part is not strong enough. But this second doubt also does not hold. A long-term study, which has made annual surveys for 10 years, shows that even though the public’s levels of pro-environmental attitudes have been gradually growing, people still perform low levels of pro-environmental behavior (Coyle 2005). In sum, there is a persistent gap interrupting the flow from pro-environmental knowing to doing, and the arrow in between them is not as operative as assumed.

In fact, the presence of such a gap is not exclusively linked to the promotion of nature-friendly behavior. Any attempt to change human behavior will run into the same issue if it presumes a similar flow in which human knowledge plays the role of initiator. Indeed, the phenomenon of the “knowing-doing gap” has also been identified in the field of business administration. A study, in which the term was coined, shows that in many organizations, knowledge that has been learned often fails to induce behavior consistent with it (Pfeffer and Sutton 2000). Moreover, as concluded by another study, for any organization that wants to solve its problems, the knowing-doing gap is far more important and attention-worthy than the gap between ignorance and knowing (Pfeffer and Sutton 1999). In short, the ubiquitous knowing-doing gap constantly shows up everywhere and all the time. If it is a severe problem to solve, we may need to first know why the gap emerges.

There are many reasons for the presence of the knowing-doing gap. With regard to the environmental crisis, imagine the following scenario. A person just learned from environmental education that climate change is caused by the accumulation of CO₂ and that a certain proportion of CO₂ comes from driving cars. Moreover, s/he has been convinced by ethical arguments that nature has intrinsic value and that s/he has the responsibility for the contemporaries, future generations, and all other living organisms. Accordingly, s/he cherishes the beauty and vitality of nature very much and is motivated to bike for commuting. However, s/he still drives to work on her/his workdays, although s/he has already made a commit to her/himself several times that next time s/he will jump on her/his bike. Why? Maybe because, as her/his friends said, s/he is a person of
Design for Green

weak will, or s/he lives in a rainy city and her/his office is a bit far from her/his house. Probably, the answer is surprisingly simple: no bike paths in the city, which makes her/him always feel unsafe while biking. There can be more reasons. No matter what the reason, her/his will to behave pro-environmentally is not followed by an actual action. Such a scenario shows, apparently, that the flow from knowing to doing can only be completed under certain conditions.

Indeed, behaving morally often requires external support (Lindenberg 1983), and “in many cases altruistic and biospheric values need to be supported” (De Groot and Steg 2009, 63). Without support, be it physical or non-physical, informedness cannot lead to any long-lasting pro-environmental behavior (Schultz 2002; Stern 1999). Unfortunately, relying on the assumption that what people know generates what people do, environmental education pays too much attention to the former and ignores the fact that the lack of external support can create a gap in front of the latter. Worse, focusing exclusively on pro-environmental knowing may create a “reversal”. Doubtlessly, most people look for consistency between their beliefs and their behavior; however, when a person’s beliefs and behavior conflict, s/he is likely to alter the former rather than the latter (Festinger 1962). This means, accordingly, that once a person who has learned to value nature finds his/her behavior to be nature-unfriendly or hardly pro-environmental, s/he may revoke his/her positive attitudes toward nature. In short, with its special focus on knowing, environmental education keeps it blind to the factors that create the gap between knowing and doing, and may waste its efforts so that learners not only behave anti-environmentally as before, but also merely parrot what they have learned back to their teachers.

Many have previously noticed this troubling gap between environmental knowing and doing; however, the breakdown of the arrow from the knowing part to the doing part has been attributed to the failure of teaching techniques or the educational system (Saylan and Blumstein 2011), and the reliance on the assumption of the flow is not called into question. Moreover, while acknowledging the gap as a severe challenge, it has been proposed to shift the target of environmental education from people’s behavior to attitudes alone because “[b]ehavior, in contrast, is more influenced by non-personal formative influences” (Eilam and Trop 2012, 2237). According to such a proposal, whether or not environmental education can successfully improve people’s environmental behavior is no longer the main concern, and what can be taught are general attitudes concerning the world of nature, rather than specific attitudes connected to the related behavior; this may, predictably, make environmental education rather
A Focal Shift from Knowing to Doing

more unhelpful for advocating sustainability (Heberlein 2012). If the improvement in people’s environmental behavior remains the goal of environmental education, the inflexible sticking to the assumed flow and the special targeting to the knowing part have to be adjusted.

2.3. The Asymmetry Concerning Anthropocentrism

The tendency to focus exclusively on knowing rather than doing reflects, and probably results from, how the problem of anthropocentrism has been dealt with. Anthropocentrism is often considered as the root of the environmental crisis (Naess 1973; Plumwood and Sylvan 1980; Plumwood 2001). On the one hand, it refers to the phenomenon that human beings regard themselves as entitled to use non-human resources without constraints, where only human well-being counts; the well-being of animals, plants, or ecosystems is rarely considered, unless they can benefit humans. On the other hand, it implies the moral thinking that human responsibility is for humans alone, i.e., if we have any responsibility to support sustainability, it is because and only because we are responsible for the future generations who will need the same resources. In other words, anthropocentrism holds an instrumental view on nature and therefore leaves no room for nature to be in the ethical system.

Environmental ethics, underpinning and collaborating with environmental education, has developed two “ethical shifts” to overcome anthropocentrism. The first shift starts with traditional ethics that concerns only human-human relations and arrives at a revised ethics that also takes human-organisms relations into account. Because this revised ethics recognizes that living entities, including animals and plants, can also be moral subjects, it is sometimes called biocentrism. The movement of animal liberation, which has been promoted by Peter Singer (2009), can be regarded as representative of this shift. The second shift goes from the biocentrism to ecocentrism, in which interactions between humans and non-human nature are considered as a part of the ecological system. The idea of “land ethics” by Aldo Leopold (1968) is usually categorized as ecocentrism due to its holism in prioritizing the sustainability of the entire ecosystem rather than that of any specific species within it. Although there are
differences between the two shifts, apparently the trend of environmental ethics has been, and still is, toward non-anthropocentrism. As environmental ethics by its nature is philosophical, the effort against anthropocentrism is no doubt a conceptual work. It is more about human knowing than doing. In other words, what has been overcome is knowing-anthropocentrism rather than doing-anthropocentrism. Therefore, it seems no accident that, in taking up the fruit of environmental ethics, environmental education focuses itself on the knowing part rather than the doing part of the flow. Not only is a non-anthropocentric attitude very basic subject matter in the environmental education courses, but non-anthropocentric ethics is also an important research topic in the field (Nash 1989). However, as already pointed out, focusing too much attention on environmental knowing unfortunately makes environmental education itself an unsuccessful attempt to improve people’s environmental behavior.

While knowing-anthropocentrism has been the target to “knock out”, doing-anthropocentrism is still there to fight against. This is what I call the “asymmetry” of overcoming anthropocentrism. Environmentalists have been working hard to build non-knowing-anthropocentrism, but the goal of improving people’s environmental behavior still seems far from being reached, not because the effort is insufficient, but rather due to a lack of equivalent effort on overcoming doing-anthropocentrism. In other words, a similar shift toward non-doing-anthropocentrism has not yet taken place; to make it happen, there are several questions to answer: How can we overcome doing-anthropocentrism? What does non-doing-anthropocentrism look like? If non-knowing-anthropocentrism has been achieved by endowing or discovering the intrinsic value of nature, can we do the same to become non-doing-anthropocentric?

---

8 To be sure, few still believe that anthropocentrism can be a route for protecting and maintaining nature. For example, Grey (1993) argues that the problem is not anthropocentrism itself, but the “shallow” version of it, which considers only short-term, sectional, and self-regarding interests. Another example is Düwell & Bos (2016), who argue that the concept of universal human rights can be an approach to build environmental ethics, even though nature in such an ethics has only instrumental value. However, this approach has problems with defining basic human needs (Jamieson 2014) and difficulties in being politically inclusive (Goodin 2007). Besides, Epting (2010) suggests the position of “weak anthropocentrism”, whose environmental values are included in human values and should be embodied in technologies as well. This position, apparently, implies the advent of nature-friendly technology and can be related to the discussion on “doing-anthropocentrism” below.
To answer these questions, we can start with the idea of Anthropocene. The term “Anthropocene” was first coined and spread by Crutzen and Stoermer (2000), implying that we have left the epoch of the Holocene and entered into a new one that is highly related to Homo Sapiens. The Anthropocene is a descriptive term referring to the fact that the rapid changes of the Earth are largely caused by human beings rather than any other organisms or natural forces. To be sure, the beginning date of the Anthropocene has not been fully determined yet. Some believe that it can be traced back to the Industrial Revolution in the early 18th century (for example, Zalasiewicz et al. 2008), but others argue that it should be set in the mid-20th century, the time that many environmental problems extended globally (for example, Waters et al. 2016). However, both sides agree that human activities have profoundly influenced the atmospheric, hydro-spheric, biospheric, and geospheric processes within the earth’s ecosystem. Whether the date was the early 18th or mid-20th century, there is a strong connection between the date and technological development. This is probably why the human transformation of the environment in the 21st century is believed to be more dramatic than ever before. The investment in technological innovation and the speed of technological development since the Second World War have been growing so fast that it is plausible to say that we are in the era of the “Great Acceleration” (Steffen, Crutzen, and McNeill 2007; Steffen et al. 2011).

Although the term is descriptive, it has been argued in both humanities and social sciences that the idea of the Anthropocene has prescriptive implications. For example, it may present itself as an ideology by which “humanity is re-inserted into ‘nature’ only to simultaneously be elevated within it and above it” (Baskin 2015, 19). Moreover, “[t]he Anthropocene discourse touts the unavoidable merger of the human-natural, which, according to its reports, calls us to the high road of becoming good managers of the standing reserve” (Crist 2013, 144). In other words, the idea is regarded as a return of anthropocentrism, by which humans will be legitimized to master non-human nature. Although such arguments against the Anthropocene are understandable and plausible, they work only for the knowing dimension of anthropocentrism. As argued below, if we consider the doing dimension, it will become impossible to get rid of anthropocentrism.

When we talk about how to protect the environment or mitigate climate change, it is hard to avoid taking the position of managers as they are the ones who can act. No matter how strong the non-anthropocentric attitudes we have, humans are the ones dealing with the problems of resource depletion or climate
change. In other words, doing-anthropocentrism is *unavoidable* if we seek a sustainable future. What if, one may contend, we take the policy of “noninterference”, just leaving nature as it is, and stop taking advantage of it, but also repairing nothing? Is not such a policy non-doing-anthropocentric? Yes, but it is unrealistic and unpractical. As indicated by the term Anthropocene, humans have already become a primary force in shaping the Earth; it also implies that humans are competent to repair or recover the damage they have caused. If we choose just leave nature alone, the environmental problems are likely to continue deteriorating, especially given that the consequences of the environmental crisis can be delayed for several decades. For example, animal conservation to prevent human-caused extinctions will run into difficulties if we take *in situ* measures alone. The term “*in situ*” means that the animals to be protected are left in their original habitats and away from human interference as much as possible. However, it is unlikely to succeed because their pending extinctions result exactly from human-made damages to their habitats. Therefore, *ex situ* measures such as constructing natural-park-like zoos are indispensable (Keulartz 2017). In short, non-doing-anthropocentrism cannot be achieved by avoiding human interference in the age of the Anthropocene, unless we give up the opportunity to fix environmental problems.\(^9\)

If human interference is unavoidable in solving the environmental crisis, then what would be non-doing-anthropocentrism? We may learn from how knowing-anthropocentrism has been overcome. As mentioned earlier, people’s anthropocentric attitudes toward nature are corrected by environmental education through raising the status of nature. In other words, the origin of knowing-anthropocentrism is the asymmetric evaluation of humans and non-human nature: Humans are intrinsically valuable, but nature is not. By giving an equivalent to non-human nature, non-knowing-anthropocentrism becomes achievable because people are likely to consider both equivalently (see the left column of Table 1). However, we have not yet built any similar equivalency in the doing dimension of anthropocentrism. In the Age of Anthropocene, while we invent and adopt technological tools to fix non-human nature, we seldom do

\(^9\) In fact, those who object to the idea of the Anthropocene do not really oppose human interference with the environment. For example Baskin (2015) and Crist’s (2013), real concern in accepting the idea of the Anthropocene is clear: While scientists and engineers will have privileges to have a say, humanists and sociologists may be marginalized in solving the environmental crisis.
the same to humans, except for curing human disease. While technology is entitled to interfere nature’s “behavior”, such as the depletion of stratospheric ozone or the accumulation of green-house gases, such an entitlement is rarely applied to human behavior. Doing-anthropocentrism, therefore, originates from an unbalanced treatment between humans and non-human nature. Accordingly, to overcome doing-anthropocentrism requires treating both symmetrically; technology can be a remedy not only to unsustainable nature but also unsustainable humans (see the right column of Table 1).

Table 1. How to make non-anthropocentrism possible for human doing

<table>
<thead>
<tr>
<th>KNOWING</th>
<th>DOING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANTHROPCENTRISM</strong></td>
<td></td>
</tr>
<tr>
<td>humans</td>
<td>nature</td>
</tr>
<tr>
<td>technology</td>
<td></td>
</tr>
<tr>
<td>humans</td>
<td>nature</td>
</tr>
<tr>
<td><strong>NON-ANTHROPCENTRISM</strong></td>
<td></td>
</tr>
<tr>
<td>humans</td>
<td>nature</td>
</tr>
<tr>
<td>technology</td>
<td></td>
</tr>
<tr>
<td>humans</td>
<td>nature</td>
</tr>
</tbody>
</table>

2.4. A Possible Way Out

As argued in the previous sections, environmental education as an unsuccessful project represents the difficulties that the conventional strategy to promote people’s environmental behavior often confronts. On the one hand, presuming the flow from human knowing to doing overly emphasizes the change of the knowing part and neglects the disturbing presence of the knowing-doing gap. On the other hand, taking up the fruit of environmental ethics appeals to a more balanced evaluation between humans and nature in order to overcome knowing-anthropocentrism; however, it leaves the problem of asymmetric treatment for doing-anthropocentrism. Caught in such a “double trap”, environmental education’s enthusiastic endeavor does not come close enough to its goal. To escape the trap, we need not only to pay equivalent attention to the doing part of the flow, but also to figure out what technology can do to help.

As discussed earlier, the flow from environmental knowing to doing cannot be operative without certain conditions. Recall the imagined example of the person who has been taught to prefer biking but still drives. Thanks to environmental education, s/he knows very well what s/he can do to decrease CO₂
emission, but lacks external support to turn it into doing. In such a situation, technology can play the role of enabler mostly in a physical form. For example, if a network of bike paths can be nicely built in her/his city, his/her use of a bike may increase because the obstacle is gone. This is not an implausible proposal. An analysis on the high levels of cycling in three countries: Denmark, Germany, and the Netherlands, has strongly concluded that the secret to increasing the percentage of cycling is to provide bikers and potential bikers with a nice network of cycle paths (Buehler and Pucher 2008). That is to say, as an enabler, technology can bridge the knowing-doing gap and therefore make the flow run. However, such a form of technological help seems insufficient because it helps only those who have already learned from environmental education. For the others who cannot be easily reached by the practices of environmental education or who have not been convinced by environmental ethics, such a technology can be either redundant (“We have roads already!”) or annoying (“Bike paths take my driving space!”). In other words, these “uneducated” people will be left behind if we only look for technology as supportive help.

Fortunately, there is another way that technology can help. A growing body of studies in experimental psychology and behavioral economics has shown that human behavior is highly influenced by physical factors outside of human awareness. For example, the size of a plate can influence on how much a person eats at home or in a restaurant, but s/he may be entirely unaware of such an influence (Schmidt et al. 2013). In fact, there are too many factors that can impact people’s eating behavior without being noticed, which is why what people often do at the dining table is “mindless eating” (Wansink 2014). It means not only that people eat carelessly, but also that the behavior of eating itself has little, if not nothing, to do with human consciousness. Similarly, colors also impact users’ view on whether a website is trustworthy. “The use of vivid, high saturated colors”, for example, “is generally seen as promotional and aggressive when applied to ecommerce sites. This lowers the perceived trustworthiness of the online vendor” (Pelet and Papadopoulou 2011, 189). Although people know whether they trust or distrust a website, the emergence of such a trust or distrust however is outside of their knowing. The above studies clearly show that human knowing is not necessarily a precondition for human doing. Human behavior, therefore, can be initiated without mobilizing knowledge or attitudes.

To be sure, there are still other factors impacting on whether or not the promotion of cycling can be successful (Oldenziel et al. 2016); however, the role of physical designs are crucial.
Green behavior without environmental awareness provides an opportunity for technology to help. Sofas designed by a Dutch group called “Eternally Yours” can demonstrate this point well (Verbeek 2011). To stimulate the longer use of household sofas, an attractive pattern is sewn onto the upholstery of each sofa but covered with an ordinary surface. Such a sofa is used longer, as the pattern inside will become more and more visible. Unlike ordinary sofas, this sofa does not leave its users “bored” because it renews itself over time, thereby contributing to waste reduction; their pro-environmental behavior is the product of a sophisticated design rather than education. Such a result does not involve scientific knowledge or moral lessons to its users, but the behavior that environmental education aims to promote is undoubtedly created. To be sure, some of the users may be aware of the effect that the sofa can have on their behavior after or before using it, but that is hardly the reason for them to buy and cherish it. It entails typical behavior-steering technology, as already defined in Chapter 1. It works differently from the technology that is made to be an enabler. Rather than bridging the knowing-doing gap, behavior-steering technology can produce green behavior by bypassing the gap.

This feature makes behavior-steering technology the solution with the most potential to the double trap that has made environmental education an unsuccessful attempt to improve environmental behavior. On the one hand, the idea of behavior-steering technology does not presume human knowing as a generator for human doing, and thereby helps us to focus directly on environmental behavior instead of knowledge or attitudes. On the other hand, by changing human behavior through technical features, the practice of behavior-steering technology offers a more symmetric treatment to unsustainable nature and humans and therefore is likely to overcome doing-anthropocentrism. Moreover, behavior-steering technology can cover the role that supportive technology plays. Recall the example of the stairs-elevator design mentioned in Chapter 1. The electronic company that owns the building is actually well-known for its green advocacy. It often arrange lectures on environmental education for its employees, and its chairman is praised as a “green missionary” on mass media. Therefore, for the employees who have attended the lectures and obtained pro-environmental knowledge, the stairs-elevator design functions like an enabler to help them put their knowledge or attitudes into practice. But for the other employees and visitors, the same design is a behavior-steering technology due to its direct effect in changing their behavior from taking elevators to taking stairs. In short, behavior-steering technology can not only bridge, but also by-
pass, the knowing-doing gap, and such an approach has great potential to do what cannot be or has not yet been done by the environmental education in solving the environmental crisis.

2.5. Concluding Remarks

Environmental education, representing the conventional strategy to deal with the contemporary environmental crisis, aims to improve people’s environmental behavior by providing knowledge of the facts and values of nature. Such a strategy has run into difficulties because of the underlying assumption that human knowing is the initiator of human doing; it tends to ignore the gap between them and neglect the external factors that can support or create pro-environmental behavior. The failure of environmental education reflects, and results from, the asymmetric treatment toward knowing anthropocentrism and doing. While knowing-anthropocentrism has been overcome by the effort to evaluate humans and nature more equally, doing-anthropocentrism still reflects the unbalanced technological treatment toward human behavior and nature. Non-doing-anthropocentrism, therefore, can and should be achieved by altering human environmental behavior through the intervention of technology. With sophisticated design, technology can bridge the knowing-doing gap by helping those who have learned from environmental education to put their knowledge into practice; it can also bypass the gap by creating people’s nature-friendly behavior without appealing to related environmental knowing. Behavior-steering technology, according to its definition is aimed at human behavior; it has great potential to do the job and therefore is the best candidate to provide a complementary approach to environmental education, achieving what has not yet been achieved by the conventional strategy in solving environmental problems.
3 Classifying Behavior-Steering Technology

Behavior-steering technology can bridge or bypass the knowing-doing gap, which makes it potentially helpful for tackling the environmental crisis. Different behavior-steering technologies work in different ways, produce different effects, and raise different ethical concerns. In order to take advantage of behavior-steering technology and deal with related ethical concerns in the following chapters, this chapter aims to offer further understanding for behavior-steering technology. I will first introduce three classifications that have been made in the field of philosophy of technology to understand human-technology interaction in general. They have often been referred to analyze and guide the design of behavior-steering technologies. However, while they are undoubtedly useful, they do not really match the purpose of the present research. Therefore, I will develop a proper but simpler classification specifically for behavior-steering technology. By differentiating the mechanism by which it takes effect, a behavior-steering technology can be classified as “informational” or “material” behavior-steering technology. I will also provide various examples to illustrate this classification and discuss the pros and cons of each type. This classification, as we shall see, helps us to grasp the correlation between the degree of ethical concerns and the mechanism of behavior-steering technology.

3.1. The Classification According to What is Mediated

As argued convincingly by Peter-Paul Verbeek (2005), rather than asking what the essence of various technologies is, a more relevant question today is “what things do”. The answer to this question is “mediation”. Technology influences experience, what the world is for humans on the one hand, and praxis, what humans do in the world, on the other hand. For the former, technology does not faithfully deliver all the details of the world, but may amplify some parts or reduce others. For example, a pair of glasses helps the person wearing it to get a framed image of the world in front of his/her eyes, and a spectrogram of a star enables astronomers to know the chemical composition of the star while ignoring other aspects. For the latter, technology does not merely fulfill humans’ in-
Design for Green

tention to act, but also shapes what humans can do and how to do it. For example, a paper cup implies to its users that it can be thrown away immediately after being used, and a speed bump can prevent drivers from driving too fast even though some drivers have no such a plan beforehand. Whether experience or praxis, technology is not just a nexus between humans and the world; rather, it is a mediator shaping both sides and making them possible. In short, what technologies do is to mediate not only human perception but also action (Figure 2).

![Diagram for technological mediation](image)

Figure 2. Diagram for technological mediation (Verbeek 2006b, 3)

The above classification can help us describe and analyze how technology influences humans. However, the distinction between perception and action can sometimes be blurred. By wearing a pair of glasses, for example, a person’s perception of the world changes, as well as his/her actions differently. A speed bump causes drivers to decelerate and also presents a world in which the roads are not smooth. It seems hard to tell what is mediated exactly in both cases because, perhaps unsurprisingly, the only indicator for both mediations is human behavior. If there is no behavioral change after technology has been used, then it will be difficult to know whether or not the technology-in-use mediated. Even though the mediation of perception can be reported by the person who is using the technology, such a report can be trusted and confirmed only through observable changes in behavior, no matter how slight those changes are. Accordingly, be it the mediation of perception or action, what is most relevant to behavior-steering technology is the mediation of behavior.

Focusing on human behavior can be helpful for identifying and analyzing technological mediation. While a more detailed discussion is presented in Chapter 5, I would like to offer a preliminary point here. As behavior is a reliable indicator for technological mediation, it should be the main target for designers. On the one hand, behavioral change needs to be measured in order to understand the degree of technological mediation; on the other hand, it is also what
needs to be envisaged even though the technology-in-design is not meant to change user behavior. This seems quite straightforward for action-mediating technology, but it is also true for perception-mediating technologies. Technologies that are designed to represent, reveal, or render the world may have unintended impacts on user behavior. For example, clocks were originally designed to measure time intervals rather than displaying the time and were only for very limited use, such as religious activity. However, when clocks became technologically advanced and widely used, unified transportation systems and work schedules became possible; thereby, people’s behavior has largely changed by becoming time-oriented (Mumford 2010). Such a consequence, apparently, is unexpected to the original designers of clocks. Focusing on the mediation of behavior rather than the mediation of perception or action alone can help designers anticipate “actual” changes that may be caused by technological design.

3.2. The Classification According to the Modes of Interaction

Extending the idea of technological mediation, another classification was made by Steven Dorrestijn (2012) with an emphasis on the notion of “subjectivation” as a starting point in considering human-technology relations. Subjectivation is a Foucauldian concept, referring to the question of how to become a human subject. Dorrestijn argues that technology plays a fundamental role in subjectivation because a person’s “subject” is formed gradually and adjusted continually through his/her interaction with technologies. Such interaction can be distinguished into four modes: above-the-head, before-the-eye, to-the-hand, and behind-the-back (Figure 3).

![Figure 3. Diagram for human-technology interaction (Dorrestijn 2012, 64)](image)
Above-the-head is a no-direct-contact interaction between humans and technology, concerning how technology is understood and defined. For example, three types of thinking can be identified in the history of philosophical inquiry into technology. In the early days of philosophy of technology, technology was regarded as a wonderful tool to complete and perfect human beings. Contrary to such an optimistic attitude, the so-called classical philosophy of technology offered a more negative evaluation of technology, accusing it of becoming too autonomous to be controlled by humans (Franssen, Lokhorst, and van de Poel 2018). However, contemporary thinkers tend to avoid presuming an essence for all technologies on the one hand and to investigate the pros and cons of technologies case by case on the other. All these conceptualizing about technology, instead of technology itself, impact how humans interact with technology.

Behind-the-back is another no-direct-contact human-technology interaction. In this mode, technology is regarded as a driving force for human progress (so-called technological determinism). Camera obscura is a good example here. It helped the development of linear perspective in painting art and brought forth the conception that humans are independent from the world and capable of perceiving and manipulating the world externally (Kockelkoren 2003). It should be noted here that the two above modes have common features. Both make some abstractions and concern the long-term relationship between humans and technology. While technology in the mode of above-the-head is actually “Technology-with-a-capital-T”, an essence abstracted from concrete technologies, humans in the mode of behind-the-back are in fact “humans-as-a-whole”, a form of being abstracted from a large number of humans. Such abstractions, concerning different sides of human-technology interaction, are done by looking into a certain period of time in history. In other words, the modes of above-the-head and behind-the-back are made to describe long-term human-technology interaction.

The remaining two modes, in contrast, are more about short-term interaction between humans and technology. The mode of before-the-eye is about “reading” the signs or meaning of technology. For example, drivers who see a standing “No Parking” sign on the roadside tend not to park their cars alongside the road. The message sent by technology can be implicit as well. Paper cups, as mentioned previously, through its material imply to users that it is okay to throw them into the rubbish bin after finishing the drink. In contrast to the mode of before-the-eye, the mode of to-the-hand has little to do with reading. Whether a person understands a technology message is irrelevant to how the person interacts with it. For example, bikes require specific gestures to ride, but most people
can do so without thinking much. People may also interact with a technology unconsciously. Walkers who are passing a bakery, for example, are likely to be attracted by the nice smell of fresh baked bread.

Concerning behavior-steering technology, this classification for human-technology interaction seems too complicated to be applied. The first two modes (above-the-head and behind-the-back) are highly abstract and more about long-term human-technology relationships. In contrast, the latter two modes (before-the-eye and to-the-hand) can help us to analyze or guide the design of behavior-steering technology. Designers can choose, for example, the mode by which a technology-in-design changes user behavior. However, these two modes are insufficient to provide further information about the form (explicit or implicit?) and the effectiveness (strong or weak?) of technology-in-design; it therefore helps little to make a connection between the types of behavior-steering technology and possible ethical concerns.

3.3. The Classification According to Force and Salience

Another four-quadrant classification, which is also an extension of the idea of technological mediation, was made by Hekkert, Tromp and Verbeek (2011). This classification is different from the other two in its aiming at technologies themselves rather than relations or interactions between humans and technologies. According to the force that a technology exerts (strong or weak) and the salience of technological influence (hidden or apparent), four types of technology can be distinguished: decisive, seductive, coercive, and persuasive (Figure 4).

![Diagram for technology's form and effectivity](Hekkert, Tromp, and Verbeek 2011, 12)
A stone put on a specific area to prevent cars from being parked is coercive because its technological force cannot be easily overcome or ignored. In contrast, the warning words on a cigarette package are persuasive because although it sends a clear message about the consequence of smoking, the force of those words is obviously weak. A coffee machine put in the hallway of a company or department is a good example of seductive design. The force of such a coffee machine in gathering colleagues for small talk seems unclear and only takes effect sometimes. A decisive design can be exemplified by the location of a bus stop; it determines how far passengers have to walk, but this decisive force is rarely noticed by them.

This classification, compared to the other two, is directly relevant to behavior-steering technology and more detailed to help designers understand or conceive how technologies-in-design can or will influence their users. These four types can be regarded as strategies for designers to choose. However, as I would like to point out here, this classification lacks something important to fit into the present dissertation. While the distinction of force can be easily related to the form of technology's influence, the distinction of salience is more about how users experience technology. Recalling the examples above, the force of both strong technologies (the no-parking stone and the bus stop) is physical; in contrast, the force of both weak technologies (the cigarette warning and the coffee machine) comes from their meanings. However, such a stable, if not clear-cut, boundary does not exist for the salience of technology because whether a technology's influence is salient often varies with the users' observational skill or critical ability. For students who have taken the course of philosophy of technology, the coffee machine in the hallway is by no means a mystery in creating close relationships between the colleagues; the mediating effect of the coffee machine is easy to see and predict. Similarly, while a Japanese tourist may understand nothing about the warning words when buying a pack of cigarette in a Dutch shop, these words are literally readable and therefore really a warning to Dutch people. In short, the boundary between hidden and apparent technological mediation depends highly on users' experience and therefore can be very floating; this makes salience hard to be a reliable criterion to classify behavior-steering technology.
3.4. A Classification Made for Behavior-Steering Technology

To focus more on behavior and to make a steady distinction between different types of behavior-steering technology, I would like to suggest a new classification for further discussions in the coming chapters. Rather than technology in general, this classification is made specifically for behavior-steering technology. Below I will first distinguish two types of behavior-steering technology by differentiating their mechanisms in influencing users. Then, I will discuss the pros and cons of each type in detail by providing several examples. As we shall see in the coming sections, this classification is simpler but sufficiently helpful to make a connection between the types of behavior-steering technology and the degree of ethical concerns.

If a technology is designed to change user behavior by providing or sending perceptual clues, such as signs, texts, images, colors, or smells to the users, it can be called “informational behavior-steering technology” (hereinafter abbreviated as IBST). In contrast, if a technology is designed to change user behavior by setting physical conditions, it is called “material behavior-steering technology” (hereinafter abbreviated as MBST). It should be noted that this classification is analytic; therefore, an actual behavior-steering technology sometimes may be able to be identified as IBST as well as MBST. Also, an MBST at one time may be identified as IBST at another (see below for details). In short, what type a behavior-steering technology should be classified as depends on how it reaches the goal of changing user behavior at the moment that users interact with it.

For IBST, the mechanism by which it takes effect can be described as “intrasomatic”, and for MBST as “extrasomatic”. These terms are borrowed from Bruno Latour (1992), who argues that skills can be possessed not only by humans but also nonhumans:

It is true that in Paris no driver will respect a sign (for instance, a white or yellow line forbidding parking), nor even a sidewalk (that is a yellow line plus a fifteen centimeter curb); so instead of embodying in the Parisian consciousness an intrasomatic skill, authorities prefer to align yet a third delegate (heavy blocks shaped like truncated pyramids and spaced in such a way that cars cannot sneak through); given the results, only a complete two-meter high continuous Great Wall could do the job, and even this might not make the sidewalk safe, given the very poor sealing efficiency of China’s Great Wall. So the deskilling thesis appears to be the general case: always go from intrasomatic to extrasomatic skills; never rely on undisciplined people, but always on safe, delegated nonhumans. (Latour 1992, 167–68, emphasis added)
Apparently, by the term “somatic”, Latour refers to a description of the boundary, the surface, or the shell of the human body. The adjective “intrasomatic”, therefore, describes what is inside the body or under the skin. Conversely, the adjective “extrasomatic” describes what is outside the body or above the skin. These two descriptions can be applied well to the two mechanisms of behavior-steering technology.

When an IBST steers its users by providing information, including language clues and perceptual data, it operates intrasomatically. For language clues, such as texts, symbols, or verbal sounds, the locus of behavior-steering operation is usually conscious. It requires the person whom an IBST steers to interpret the information s/he gets; otherwise, the information can be meaningless as language clues. For perceptual data, such as images, colors, or smells, the locus of behavior-steering operation is often unconscious. It does not require the person who is being steered to understand or notice the information to which he was exposed. However, the information may still be translated by the person’s unconscious; such a process is often called “intuition”. In contrast, when an MBST steers its users by setting physical conditions to increase or decrease the possibility of a certain behavior, it operates extrasomatically. The person who is influenced has no need to process information consciously or unconsciously. There might be information processing, but that is not what makes an MBST work. Rather than consciousness or unconsciousness, the locus of material behavior-steering operation is the human body. In short, to influence user behavior, IBSTs take intrasomatic routes while MBSTs take extrasomatic ones (Figure 5).

3.5. Informational Behavior-Steering Technology

As already pointed out, an IBST operates at the locus of either consciousness or unconsciousness. This sub-distinction can well reflect the two mainstream ideas
on how to influence human behavior through design: one is “persuasive technology”; the other is “nudge”. Both ideas have attracted the efforts of many designers, engineers, and theorists. To be sure, while persuasive technology has been widely recognized as behavior-steering technology, nudge is rarely put into such a category. However, as we shall see below, the nature of nudge is consistent with behavior-steering technology.

The term “persuasive technology” was coined and defined by B. J. Fogg (1998; 2003) as any computing system, device, or application intentionally designed to change a person’s attitudes or behavior in a predetermined way. Having originated from the research field of human-computer interaction, it is quite natural that persuasive technology takes an intrasomatic route. It aims to change user behavior by providing meaningful texts or images through reminders, such as pop-up windows, text messages, emails, and the like. At present, human health and energy saving are the two main research interests in the field of persuasive technology.

For human health, several computer or mobile-phone applications have been designed to persuade users to smoke less (Oinas-Kukkonen, Pahnila, and Räisänen 2008), self-monitor weight (Chatterjee, Drew, and Maheshwari 2008), eat more healthful food (Carofiglio, De Rosis, and Mazzotta 2007), or maintain a workout routine (Goris, Lacroix, and Saini 2009). Such technological persuasion can be done in a basic way, for example, sending an email to users every morning at seven o’clock with the message “Want to be healthier? Put on your shoes and begin your day with a 20-minute jog!” A more sophisticated way to technologically persuade users is to detect user behavior and then send reminders according to what has been detected. For example, when the software detects that the keyboard has been continuously typed for at least one hour, a window with a message “Stand up and stretch your back!” plus the emoji “ciąg” will pop up on the laptop screen. With the development and popularity of smart phones, persuasive technology is designed to be smarter and more complicated. For example, a smart bracelet or watch connected with a mobile-phone can regularly trace its user’s activities, analyze the data it gets, show the user his/her activity pattern, as well as offer lifestyle suggestions.

For energy saving, “smart meters” are a good example. Most people do not know how much energy they have consumed until they receive their monthly bills, and it is not possible to tell from the bills how much energy each device consumes. To deal with this situation, smart meters have been invented to give users more details about household energy consumption, let them know where
the electricity goes, and show real-time feedback on the energy usage of each
device (Riche, Dodge, and Metoyer 2010). As is widely known, smart meters are
typical persuasive technology, and determining how to improve them is still de-
signers’ main concern (Gamberini et al. 2012; Ham, Midden, and Beute 2009;Holmes 2007).

While persuasive technology operates more at a conscious level, nudge has
much to do with the unconscious. The idea of nudge has been suggested and
popularized by Richard Thaler and Cass Sunstein (2009). It is now a frequently
referred term in some academic areas such as behavioral economics, political
theory, and policy studies. As originally defined, “[a] nudge is any aspect of the
choice architecture that alters people’s behavior in a predictable way without for-
bidding any options or significantly changing their economic incentives”
(Sunstein and Thaler 2009, 6). For example, the arrangement of the position of
food in a college canteen can influence the students’ choice of food. If the college
promotes healthy eating, then the manager should put healthful food at eye level
and move junk food to an unobtrusive corner. Another famous example is the
urinals at Schiphol Airport, Amsterdam. After a housefly image was attached to
every urinal in men’s rooms, the spillage decreased by 80% (Sunstein 2014;
Sunstein and Thaler 2009).

Rather than providing information, nudges operate mainly on the human
unconscious. The theoretical background of the idea is a cognitive model con-
cerning how humans think (Kahneman 2011). It argues that humans have two
ways of thinking: System 1 (Automatic System) and System 2 (Reflective Sys-
tem). System 1 is called automatic because it is a fast intuitional process; we use
it frequently in daily life, such as speaking in our mother language, riding a bike,
or taking a shower. We not only pay little attention while doing them, but also
spend nearly no time on thinking of how to do them. In contrast, System 2 as a
reflective system is slow and discreet; we use it only in certain occasions, such as
calculating, drafting an essay or a tweet, or speaking in a second language. These
actions require our energy, attention, and time; otherwise, they cannot be per-
formed well. Although System 1 as a mental shortcut is very helpful in reducing
our load in making decisions, it creates several “cognitive biases”, such as an
anchoring effect, availability heuristics, and loss aversion, which often lead us to
make mistakes. Unfortunately, the fact is that we tend to use System 1 rather
than System 2 to deal with complicated and long-term issues, including smoking
cessation, organ donation, and the environmental crisis.
Building on this dual-thinking-system model, Thaler and Sunstein (2009) argue that if people have a tendency to rely on system 1 and thereby often make mistakes, the government should help them to make the right decisions through nudges. On the one hand, nudges can be designed to steer people’s actions by reminding them to avoid using the mental shortcut. For example, a countdown timer that pops up when privacy-concerning software is being installed can prevent users from non-stop clicking “next” and hopefully redirect their attention to the article of the software’s privacy policy. On the other hand, nudges can make use of mental shortcuts. The design of the urinals with housefly-images belongs to this kind. It works because men’s attention and accuracy will increase instinctively when there is a target to aim for (Vicente 2004). Another interesting example is the design of birth-control pills. If you take a pill a day for several days, a routine will be formed so that you have no need to keep reminding yourself. This is why the number of birth-control pills in a package is usually 28. While 21 of them do have medical effect, the remaining seven pills are in fact placebos. These seven pills are needless for the purpose of contraception, but such a design helps users to become “automatic” in avoiding missing any of the pills.

In short, while persuasive technology provides its users with language or symbol clues that require interpretation and translation, what nudges give are rarely in need of users’ understanding; they are processed without being noticed. Either way, persuasive technology and nudges exert their influence on human behavior via intrasomatic routes. In taking such routes, IBSTs often turn out to be ineffective in changing user behavior because intrasomatic routes are possible only when users can successfully “digest” what the IBSTs “feed” them.

Let us consider persuasive technology first. Users may not be able to understand the message sent by a persuasive technology without being equipped with relevant knowledge. For example, the amount of consumed energy shown by a smart meter can be meaningless if its users cannot recognize the unit of electricity consumption. The value “17kWh/day” (read: 17 kilowatt-hours per day) on the screen reveals nothing to these users and therefore cannot produce energy-saving behavior. To solve a problem like this, designers often put effort into improving the “readability” of a persuasive technology’s message. If the unit “kWh” is hard to understand, the designers of smart meters can use graphical, rather than numeral, symbols to express the same thing. The image of a dying polar bear, for example, may be influential because it shows the consequence of energy over-consumption and requires less knowledge to understand (Hong, Kim, and Magerko 2010). However, such an image may also fail because some
users have no idea about the connections among electricity production, CO₂ emission, glaciers melting, and the death of polar bears. In other words, whether users are knowledgeable enough is crucial for persuasive technologies to take effect.

This is also the case for nudging. For example, setting a printer’s default to double-sided mode may make some users feel unbearably annoyed if they do not know the fact that paper is made from trees. As a result, these users may discard the default and demand a new setting, and the original nudge may be erased. To be sure, such a nudge is supposed to be unnoticeable and induce no dislike. However, when users have a strong need for one-side printing, the default will become visible and troublesome. In such a situation, giving users the knowledge about the connection between paper and trees may help them tolerate the inconvenience and therefore make the nudge keep working.

The examples above reveal two difficulties that the approach of IBST for sustainability has to face. First, basic knowledge about the issues steering users is indispensable for the IBSTs to take effect. This is, apparently, where environmental education can help. However, the problem with environmental education as mentioned in Chapter 2 can be troubling here. For those who doubt the scientific research about climate change and who are not convinced by ethical arguments, IBSTs are annoyances rather than helpers. In other words, IBSTs are barely effective for the users who ignore or reject environmental education’s teaching. Unfortunately, these users are exactly the ones whom the approach of behavior-steering technology should target. When people’s ability varies, the effect of an IBST is inconsistent. This makes IBST unlikely to produce behavioral change in a large scale. In other words, IBST is quantitatively ineffective. Individual differences, accordingly, is a severe challenge for the designers of IBST (Kaptein, Lacroix, and Saini 2010). It is also no accident that “personalization” has become a developing trend in IBST design (See, for example, Cummins et al. 2008; Brug, Kroeze, and Werkman 2006; Milat, Neville, and O’Hara 2009).

Even though the quantitative flaw can be solved, an IBST may end up losing its impact because its users have already become too familiar with it. For example, a health application that sends eye-catching reminders usually works well in the beginning, but its effect is likely to decrease over time (Bickmore et al. 2007). Another example is the Piano Staircase, built next to the escalator at the Odenplan metro station in Stockholm, Sweden. Designed to look and function like a piano keyboard for people to play, its attempt is to make people take the stairs instead of the escalator. In a recorded video for the first two days after it
was completed, the usage of the Piano Staircase indeed went higher than usual (Volkswagen Group Sverige AB 2009). However, the fun of the staircase was soon reduced and its effectiveness in the long run was seriously questioned (Brombacher et al. 2013). Apparently, IBST is also qualitatively ineffective.

Although there are quantitative as well as qualitative flaws, taking intrasomatic routes offers some advantages. On the one hand, given that adjusting information is usually easier and less expensive than adjusting material, it is more possible to tailor IBSTs, rather than MBSTs, for each user. In receiving the information that one really cares about, a person may be more willing to accept being steered and therefore raise only few, if any, objections. Not only are the designers in the field of persuasive technology going in this direction (Busch, Schrammel, and Tscheligi 2013), but the promoters of nudges also address the importance of tailoring (Egelman et al. 2017). On the other hand, as intrasomatic routes can be easily interrupted, IBSTs are hardly coercive. This feature is in line with the public’s impression about education, which regards education as giving children knowledge for deliberation and letting them make decisions by themselves. As a result, the approach of IBST has more potential to be incorporated into consumer products and to face little resistance on the market. As IBST is more attractive to commercial companies, it is no accident that persuasive technology and nudge dominate the practice of behavior-steering technology.

3.6. Material Behavior-Steering Technology

An MBST exerts its influence mainly through an extrasomatic mechanism. A typical example is a speed bump. It gives a material constraint to the cars passing over it. Drivers have to decelerate; otherwise, the sudden change in height may cause bodily discomfort. It is not necessary for drivers to know the speed limit, see and understand the traffic reminder “Don’t drive too fast!”, or be educated beforehand that driving too fast may endanger others and therefore is immoral. Although according to the idea of “script” a speed bump expresses the instruction “Slow down and protect your body!” (Akrich and Latour 1992); drivers who reduce speed in fact neither see nor hear a single word. Without anything to be read or understood, a speed bump does a great job in improving people’s driving behavior, even though it has no innovation and is not interesting.

What is interesting about this approach is that an MBST may become an IBST after being widely adopted. A fence is a good example here. At the begin-
ning, as a physical barrier to protect a private yard, a fence had to be tall enough to prevent non-owners from entering the yard without permission. Without surprise, no one behaves offensively thanks to the fence. As more and more households have installed their own fences, the symbolic meaning of a fence would emerge gradually: “What is within this enclosure is mine, not yours!” As long as this meaning is recognized by most people, a new fence can be built much lower than before. People today clearly know how to “read” fences, and in such a context these fences work like an IBST rather than an MBST.

Due to its extrasomatic feature, the steering effect of MBST is often stronger and longer than an IBST. An example from Latour (1991) demonstrates this point vividly. The customers of a hotel in Berlin often forgot to return their keys to the front desk when they left the hotel, which resulted in a missing key issue that troubled the manager a lot. To solve the problem, the manager’s first try was to tell (with serious tone) the customers, “Bring back your room key!” when they checked in. However, only a few customers followed what he said. The manager then made another attempt, putting up several notices which said, “Bring back your room key!” on the walls around the hotel and made it noticeable by adding a shining red light to each of them. The situation improved, but not sufficiently. Finally, the manager figured out a creative solution. He attached a heavy but beautiful piece of metal to each key with a key ring. Thus, the problem was largely solved. The size and weight of the metal made the “key-set” hard to put into a pocket, so the customers would rather leave their keys at the front desk. Even though some customers had big pockets to hold it, they usually felt uncomfortable with a heavy object on one side of their bodies (Figure 6).

![Figure 6. The strategies to solve the key-missing problem (Latour 1991, 107)](image-url)
The manager’s first try (Row 2) represents the most direct, simple strategy to tackle moral decline. Oral reminders, advice, instructions, warnings, and admonitions are often used to correct misbehavior. As they function most like education, it seems not surprising that this strategy fails, as already shown in Chapter 2. The written notice as the second try (Row 3) is a typical IBST, which predictably bumps into the difficulties discussed in the previous section. For the customers who never learned German (or English, as the most globalized language) or who do not know how to use Google Translate, the notice was not understandable and was therefore meaningless. Moreover, the shining red light may only be noticeable for customers’ first one or two days in the hotel because they are busy with their sight-seeing plan. Indeed, human attention is a limited cognitive resource (Chabris and Simons 2011); when it is used for one thing, other things naturally retreat into the background. Thanks to the metal (Row 4), the difficulties were well solved. Apparently, it is the material features—its size and weight—which made attaching metal successful; doubtless, the metal should be regarded as MBST. Regardless of where the customers come from, the metal keeps reminding them, albeit silently, to bring the keys back to the front desk.

As shown by the above hotel key example, MBST is effective both quantitatively and qualitatively. To be sure, there may be a few customers who are too sturdy to see the metal as burdensome (there are also dogs which care nothing about the keys, as Figure 6 shows). Compared to the variance in cognitive ability between persons, the differences in physical status are usually small enough to allow MBSTs to work smoothly. The quantitative and qualitative effectiveness attests to MBST being a more competent approach to help with the environmental crisis. Given that the environmental crisis is a long-term, accumulative result of small environmental problems, to mitigate it often requires not only durable but also collective changes in people’s behavior. Such requirements are hard to meet by appealing to the approach of IBST. In other words, if the environmental crisis is what we design and implement behavior-steering technology for, then we should look at MBST rather than IBST.

However, the strongest can be the weakest: The approach of MBST tends to cause more ethical concerns than that of IBST. This becomes an Achilles’ heel for MBST, and I will deal with it in the coming chapters. Here we only need to know that because MBSTs are often regarded as ethically problematic, they are designed and implemented in quite limited ways and occasions. Up to now, most MBSTs are made to avoid mistakes. For example, USB sticks can only be plugged into a laptop at a specific angle because the shape of the port and stick
can only match one another at that angle. If a USB stick can be plugged freely at any angle, then the stick may not work or even damage the laptop’s operation system. This kind of design comes from Japan and has been named “ポカヨケ” (poka-yoke), in which “ポカ” means “mistakes in daily life” and “ヨケ” means “to prevent”. This term is usually translated as “fool-proofing” in English. This idea originates from automobile industry, for preventing workers from making mistakes that may result in life-threatening consequences (Shingo 1986). With the help of fool-proofing designs, workers do not need to be knowledgeable or experienced about the machines they are operating, or pay full attention to every tiny step during the work. Fool-proofing designs are definitely MBSTs. However, fool-proofing design has rarely been connected to the idea of behavior-steering technology, and probably reflects the fact that the acceptance of MBST is much lower than that of IBST.

3.7. Concluding Remarks
Technology, be it a behavior-steering or general one, impacts how humans think and what humans do. It can be classified by distinguishing what it mediates (experience or praxis), how it interacts with humans (the point of contact), or the degree of its force and salience. These classifications equip us with a basic understanding about the multiplicity of the relations between humans and technologies. However, with regard specifically to behavior-steering technology, the questions of how and by what means a behavior-steering technology has influence on human behavior is more relevant for taking advantage from it. The mechanism matters. When a behavior-steering technology exerts its influence by providing users perceptional clues that require interpretation or translation, it can be called informational behavior-steering technology (IBST). By appealing to human consciousness (persuasive technology) or unconsciousness (nudge), an IBST takes the intrasomatic route toward behavior promotion. In contrast, a material behavior-steering technology (MBST) influences its users by setting physical conditions that make some behaviors more (or less) possible than others. Operating at the locus of human body, an MBST takes the extrasomatic route to produce behavioral changes. Different routes result in different degrees of effectiveness. Qualitatively as well as quantitatively, while IBSTs tend to be ineffective, MBSTs usually not only work better but also longer. Therefore, the ineffectiveness of intrasomatic routes makes both persuasive technology and nudge less ethically worrying and therefore popular inside and outside acade-
mia. In contrast, because the environmental crisis is a large scale issue, MBSTs are much more competent and reliable in creating collective behavioral changes. However, to take advantage of MBST, ethical concerns related to it have to be well answered and eased.
4 Ethical Concerns about Behavior-Steering Technology

Behavior-steering technologies take effect by feeding users information that requires conscious interpretation or unconscious translation, or by giving material conditions that increase or decrease the possibility of a certain bodily action. Be it the former (IBST, informational behavior-steering technology), or the latter (MBST, material behavior-steering technology), behavior-steering technology tends to provoke ethical concerns. The central concern, as I will briefly recapitulate in the first section below, is the loss of human freedom. To avoid or ease the worry concerning such a loss, several ethical guidelines and arguments have been developed. However, as MBST is seldom considered to be an option, these guidelines and arguments are made for IBST. Therefore, I will analyze the guidelines for persuasive technology as well as the arguments for nudge and see if they can help with the worries concerning green behavior-steering technology. It will become clear, then, that while MBST is more or less a must to tackle the environmental crisis, applying the developed guidelines and arguments to MBST tends to take it to a dead end. Such a dilemma, I will finally argue, implies a need to revise the current understanding of human behavior.

4.1. The Issue of Human Freedom

By analyzing ethical concerns that may potentially accompany the design and adoption of behavior-steering technology, Philip Brey suggests three concerns regarding which advocates of behavior-steering technology need to take care:

1. What are its negative consequences for the individual freedoms of users, and can such consequences be morally defended?
2. Is there sufficient democratic input during the process of design and implementation of the technology?
3. What implications do the behavior-steering technology have for the distribution of responsibility between the user on the one hand and the technology and its developers on the other? (Brey 2006, 363)
Among the three concerns, the central one is the first, namely, users’ freedom. In fact, both the second and the third concerns are related to the first one and can be regarded as deduced concerns, as explained below.

The requirement of sufficient democratic input for the design and implementation of behavior-steering technology is made to ensure users’ agreement on the technology by which they are, or will be, steered. The democratic process is regarded as a way to legitimize behavior-steering technology because it preserves users’ freedom in choosing what they are willing to do. If behavior-steering technology is to be designed to steer people into a direction against their interests, it is better to be approved by a democratic process, like the process used for passing laws (Nagenborg 2014; Verbeek 2011). In other words, in the second concern, democracy functions like a safeguard; it is an instrument to protect users’ freedom.

Similarly, the third concern for responsibility is intrinsically connected to the issue of freedom. If a user is steered by technology and therefore is not free enough to make decisions independently on what to do, then how can we say that s/he is and should be responsible for her/his behavior? It is always assumed that a person is responsible for what s/he did only when the decision for the action was freely made by the person. If users’ freedom is “cut down” by behavior-steering technology, then how and where do we attribute the responsibility? The idea of behavior-steering technology poses a challenge to the idea of freedom and therefore may threaten the idea of responsibility. This implies that as long as users’ freedom is protected as much as possible, the problem of responsibility can be diminished as easily as possible.

In short, the central concern about the design and implementation of behavior-steering technology is the erosion of human freedom. Most objections to the approach of behavior-steering technology is related to, or focused, on this point. Some warn that behavior-steering technologies are potentially coercive and therefore may undermine the autonomy of their users (Bichler et al. 2007; Hill 2017). Some regard the nature of behavior-steering technology as manipulation, which means their users are not respected (Dunt 2014; Larson 2014). Users may be cheated, misguided, or even forced by behavior-steering technology, and

It should be noted here that although Nagenborg does agree with Verbeek on the necessity of democratic processes, he takes a looser attitude regarding (noncoercive) behavior-steering technologies, viewing them as an alternative to legislation (Nagenborg, 2014). This stance is more like Sunstein’s on nudge (see Section 4.3 and 4.5 for details).
therefore not treated as autonomous beings. Whether coercion or manipulation, behavior-steering technology is regarded as technological control over users (Purpura et al. 2011; Goodwin 2012; Guttmann and Lotan 2011). Therefore, determining how to find a balance between the effect of technology and human freedom should be the most important ethical question for behavior-steering technology (Boks and Pettersen 2008; Kool, Timmer, and Van Est 2015). Some worry that those who rely on the effect of behavior-steering technologies are not really moral because they just “perform” moral behavior and learn nothing (Kumar 2016; Cotterill et al. 2013). Yet some insist that regardless of whether the directions into which users are steered are right, the action of steering in itself is wrong because human freedom is not something that can be traded (Ehrenfeld 2012; Wilkinson 2013).

In order to prevent worse case scenarios or to respond to objections, scholars and designers have made efforts on guiding or defending behavior-steering technology. The guidelines for persuasive technology have been developed and the arguments for nudge have been constructed. What are these guidelines and arguments? Are they useful for the behavior-steering technology that is aimed at people’s environmental behavior? Moreover, as MBST tends to be regarded as a serious threat to human freedom, it is rarely considered or adopted, and therefore lacks similar guidance or defense. Can the guidelines and arguments for IBST be applied to MBST? If not, why do we still need MBST?

4.2. Ethical Guidelines for Persuasive Technology

Let me start with the guidelines for persuasive technology. In advocating the idea of persuasive technology, Fogg had already been aware of the potential threat of persuasive technology to human freedom. So he made a distinction between technological persuasion and coercion: “Coercion implies force; while it may change behaviors, it is not the same as persuasion, which implies voluntary change—in behavior, attitude, or both” (Fogg 2003, 15).

Although non-coerciveness has been built into the definition of persuasive technology, there are still doubts concerning the practices of designing and implementing persuasive technology. For example, users may be persuaded by technology to do something with which they do not agree. In such a case, users’ freedom would be undermined because they are doing something that they would not do if they could make the decision independently. To prevent such a
situation, the “Golden Rule” for persuasive technology has been suggested as follows:

The Golden Rule of Persuasion: The creators of a persuasive technology should never seek to persuade a person or persons of something they themselves would not consent to be persuaded to do. (Berdichevsky and Neuenschwander 1999, 52)

If the power imbalance between persuaders and the persuaded is not easy to avoid, at least the golden rule ensures that designers, in principle, are not allowed to steer users into the direction that they do not like. What is behind this rule is that it is better for designers and users to live within the same value system, that is, they rank values in a very similar way. If a designer thinks doing A is important but some users do not, then the persuasive-technology-in-design may steer those users into the direction they do not want. The designer indeed follows the golden rule in persuading users to do what s/he would consent to be persuaded to do, but it turns out that the persuasive technology can still be threatening. In other words, the golden rule works only when there is a common view on the ranking of values. Such a common view, however, is hard to find in a pluralistic world or society, and this would become a challenge for persuasive technology.

Due to the informational feature of persuasive technology, as the adjective “persuasive” implies, the ethics for human persuasion sometimes have been adopted to guide technological persuasion. For example, Andrea Spahn (2012) suggests that “discourse ethics” for human-human communication should be also applied to human-technology interaction. If cheating is unethical for interpersonal communication, then it is also unethical to give users untrue information through persuasive technology. To be sure, a persuasive technology may be effective by “twisting the truth” in feeding false information, but such effect can only hold in the short run. Once people find the “dishonesty” of a given persuasive technology, they will stop using it and tell others to do likewise. Accordingly,

12 Discourse ethics is based on an ideal concept of communication, aiming for the truth of claims; as long as a consensus is reached through rational intersubjective conversation, such a consensus must be true, namely, universal agreed upon (Habermas 1985). However, as we will see in Chapter 8, the emphasis on the role of rationality in searching consensus has been seriously criticized.
“[i]n the long run persuasive technology has to be truthful or else it won’t be trusted” (Spahn 2012, 640).

Applying discourse ethics also helps deal with the challenge left by the golden rule of persuasive technology. Spahn set consent as a requirement as well in his ethical guidelines for persuasive technology: “G1: Persuasion should be based on prior (real or counterfactual) consent” (Spahn 2012, 643). Sphan’s approach may be able to solve the problem of the unreachable-ness of consent in a plural world because while the golden rule presupposes a common view on the ranking of values,

[d]iscourse ethics, however, is aiming at ideal rational consensus and is thus not seeking for what people actually agree upon, as they might indeed agree on very immoral ideas. Rather, the focus is on what a rational subject should agree upon, if he is only taking the better arguments (and not his irrational desire to smoke or to value other races lower than his own race) as decisive reason for action. (Spahn 2012, 644)

In this suggestion, whether users can think rationally is not relevant; instead, the key question is whether or not the consensus is rational. However, this requires criteria for what can or cannot count as rational, and this has been discussed for ages as a philosophical theme (which obviously cannot be solved here). In addition, this suggestion may implicitly bring back the issue of power imbalance. One could claim that decisions on the goals of persuasion should be given to experts such as designers and specialists on ethics because they are generally trained to think more rationally than ordinary people. As a result, users’ freedom of choice on which way to go can still be undermined.

To protect users’ freedom from being eroded by the power imbalance between persuaders and the persuaded, another guideline is suggested: “G3: Persuasion should grant as much autonomy as possible to the user” (Spahn 2012, 645). This guideline can be done by leaving room for users to say no or by building plural options into a persuasive technology. Furthermore, a user’s autonomy can be preserved if the persuasive technology is used to persuade him/her by himself/herself rather than by others. In other words, “[c]ases of self-persuasion are most likely cases in which the user already shares the value in question and uses the persuasive only to overcome a ‘weakness of the will’”

See Wong (2013) for a very similar stance; the article argues that paternalism engendered by technology is not something that philosophers and ethicists can and should avoid.
However, this seems inapplicable to the environmental crisis. As discussed in Chapter 2, solving the environmental crisis requires collective changes in behavior, but up to now there are still those who are not willing to change. Some of them are climate change denialists, and some do not value nature even though they acknowledge the fact of climate change. They are hardly moved by scientific evidence or ethical arguments and therefore should be the main target that behavior-steering technologies are supposed to steer. If we take self-persuasion as a principle to design and implement behavior-steering technology for sustainability, we are leaving those who are in need of change behind and tend to fail in promoting significant change for a sustainable future.

Another guideline concerning users’ freedom is, “G2: Ideally the aim of persuasion should be to end the persuasion” (Spahn 2012, 644). This guideline differentiates education from manipulation; persuasion should be educative rather than manipulative. While the aim of education is to create autonomous people who can teach themselves afterward without teachers, manipulation can only increase the dependency of the manipulated on the manipulators. Therefore, persuasive technologies should be designed to help users improve themselves and eventually behave morally “from the heart”. Is this guideline applicable to green behavior-steering technologies? In fact, this distinction between education and manipulation, and the worry about the dependency on external factors have long been two major concerns that surround the idea of nudge. At this point, therefore, I shall turn to the ethical concerns about nudge, and the answers may also be the ones for persuasive technology here.

4.3. Defense for Nudge

Since the term was popularized by Thaler and Sunstein (2009), the idea of nudge has been controversial and under attack not only at the theoretical level, but also at the practical level. In contrast to persuasive technology, nudge is regarded as more ethically problematic because it takes advantage of the human unconscious. As we saw in Chapter 3, the theory underpinning nudge is the distinction between automatic and reflective thinking systems, and the nudges often operate on the former. This feature, unfortunately, makes the idea of nudge often misunderstood as manipulation and therefore a threat to human freedom.

Given that the approach of nudge appeals to the automatic process of human thinking, it has been argued that nudge “leverages individual cognitive biases” (Hill 2017, 2), and that advocating nudges shows no respect for human rational-
ity and is not different from “closing our eyes to institutions that steer behavior so that they can take advantage of it” (Gigerenzer 2015, 365). Moreover, compared to persuasive technology, which provides readable information to users, the effectiveness of nudges often based on users’ unawareness, and therefore the designs of nudges have to be hidden or undisclosed. Such a lack of transparency can be an identifiable feature of manipulation (Goodin 1980). Being effective or ethical seems a dilemma inherent to the design of nudge. At this point, a nudge works like a “cheating” persuasive technology and may become ineffective because of the loss of users’ trust. “Once people recognize that their cafeteria choices are being influenced by attentional biases, they may be disposed to range their eyes more widely as they decide what to eat for lunch” (Kumar 2016, 869).

To respond to the worries concerning manipulative nudges, Sunstein often points opponents to the fact that there are always manipulations. It is impossible to find a place without manipulation:

Choice architecture is inevitable. The social environment influences choices, and it is not possible to dispense with a social environment. This point holds whether the social environment is a product of self-conscious designers or of some kind of invisible-hand mechanism. (Sunstein 2014, 118)

Not only the social environment influences people, but also “[n]ature can, in a sense, manipulate people; cold weather and snow, for example, can affect people without sufficiently triggering deliberation” (Sunstein 2016, 218). Besides, given that people’s automatic system often results in cognitive bias, what nudges do is to correct those biases rather than worsen them. In other words, “the word ‘counteracting’ is better than ‘exploiting’. Nudges can counteract biases (such as unrealistic optimism) without exploiting anything” (Sunstein 2014, 59).

Some opponents have argued that, similar to persuasive technology, nudge is not educative. People may be nudged to do the right things, but they are not doing it “from the heart” because during the process they learn little or nothing ( Cotterill et al. 2013; Hertwig and Grüne-Yanoff 2017; Kumar 2016). However, nudges can be educative if they are designed to ask people to stop for a while and consider the choice to be made. As Sunstein suggests, “[f]or those who emphasize the value of learning, it might seem best for choice architects to call for ac-
tive choosing rather than default rules. (Sunstein 2014, 95, italics in original)” The product “Keymoment” demonstrates this suggestion nicely. It is designed to require its users to make a choice deliberately between biking and driving
Design for Green (Diefenbach et al. 2014). As a key hanger, Keymoment shows the car key and the bike key side by side. If the user takes the bike key, nothing happens. If the user’s choice is the car key, it will “throw” the bike key to the floor. “You can pick it up, hang it back and still take the car—or reflect on what might be good for your health and the environment.” (Diefenbach et al. 2016, 195) The designers describe it jokingly as a “pleasurable troublemaker” and call the strategy “an aesthetic of friction”, which means inspiring reflection by breaking up routines (Diefenbach, Hassenzahl, and Laschke 2015).

Those who have only mild objections to nudge, attempt to complete the idea of nudge with the approach of “think”, “bump”, or “boost”. The first two approaches admit that human thinking system is flawed and hold that interventions do not necessarily target the automatic part. Instead, interventions can be aimed at the reflective part, which means that choice architecture should be designed to encourage people to deliberate in the process of decision-making (Cotterill et al. 2013; Kumar 2016). For the approach of boost, human thinking is not really flawed because it is “good enough” for people to respond quickly to urgent matters and increase the chance of survival. Interventions, therefore, can be designed to match the features of human thinking: if people are naturally poor at numbers, change “the representation of statistical information ... from probabilities to natural frequencies, from relative to absolute risks, or from numerical to graphical representations” (Grüne-Yanoff and Hertwig 2016, 157). However, for Sunstein (2015a), all the interventions above count as nudges too because those who are intervened cannot “choose not to choose” and therefore the freedom of choice is still limited. In fact, when nudges are designed in the way of think, bump, or boost, they work much more like persuasive technologies rather than nudges. Accordingly, those difficulties and ethical objections that persuasive technology has been faced with apply to such nudges as well.

With regard to freedom of choice, Sunstein also offers some guidelines, which are also restrictions, for designing and implementing nudges. He makes two distinctions for behavioral interventions: the first is the distinction between hard and soft interventions, and the second between interventions to means and to ends. These two distinctions compose a four-quadrant schema (Figure 7). As suggested by Sunstein (2014), the design of nudges is better to stay in the upper-left area, namely, the quadrant of soft-means. Similar to one of the guidelines for persuasive technology (G1), there should be consent on the ends toward which the users are steered. In other words, acceptable nudges only help users to reach their goals and have no influence on how they choose the goals. In this
way, freedom of the user can be largely preserved. This is why, apparently, most nudges (and persuasive technologies) are designed to improve people’s health, wealth, and happiness. After all, people seldom deny that these values are on the top of their wish lists.

![Figure 7. Nudge is a soft intervention to people's means](image)

However, people sometimes want to relax and do something that may not be good for themselves. Everyone knows how nice it feels to be a couch potato after completing a challenging task or to treat oneself to a big meal after getting a promotion. Any nudges should not entirely prevent people from relaxing; otherwise, it is in fact a “shove” and therefore can invade people’s freedom. In other words, the costs imposed to choosers are better to be small; therefore, a nudge can be “soft” enough for choosers to opt out. In short, what counts as a nudge or what is acceptable as a nudge is an intervention only to means and in a soft way.

Again, as pointed out in the section on persuasive technology, people have different goals in a pluralistic world. Even though most people agree on the importance of health, wealth, and happiness, there can be divergent answers for different people to what counts as health, wealth, or happiness. Sunstein also thinks that the way to intervene effectively but softly is to respect heterogeneity “by avoiding one-size-fits-all approaches and by attempting more personalized approaches” (Sunstein 2014, 99). This suggestion is, apparently, consistent with the trend in the development of persuasive technology. However, this trend can induce the concern about privacy. For example, it is argued that to know how to design personalized nudges, choice architects need to collect a great amount of personal information, some of which the users probably want to keep private (Kapsner and Sandfuchs 2015). The dilemma between being effective and being ethical, unsurprisingly, returns again.
Stronger objections to the idea of nudge point not to nudges as a form of intervention but to the action of intervention itself. Sunstein (2014) observed that these objections are mostly rooted in Mill’s Harm Principle. To put it simply, Mill argues that only the person himself/herself knows what is best for him/her, so there should be no room for others to make decisions for him/her, unless his/her decision would harm others (Mill 1978/1859). In other words, only when there is, or will be, harm to others, is external intervention allowed and justified. This principle, therefore, is also called the Liberty Principle. It poses a fundamental challenge to the design and implementation of behavior-steering technologies. To deal with this challenge, it might be helpful to start with Sunstein’s replies to the Harm Principle.

Sunstein (2014) distinguishes the thin version of the Harm Principle from the thick one: the thin version holds that because no one knows better what is good for the person than the person himself/herself, there should be no intervention to any person. For Sunstein, this argument however does not hold because, as we have seen already, people tend to make wrong decisions, which means they do not always know the best option for themselves. An ideal human (*homo economicus*) can figure out what is most beneficial for himself/herself, but people in the real world do not act like that; they are *homo sapiens*. This is why Sunstein calls the Harm Principle an “epistemic argument”, implying that it only works in an abstract world. Sunstein then argues that the defenders of the thin version are not so different from him and his allies because the aim of both sides is to improve people’s welfare, and these defenders just think that the freedom of choice offers a way toward achieving this aim. If the freedom of choice cannot ensure the way, then there is no reason to keep this freedom completely untouched. In other words, for the thin version, human freedom is expendable if the returns of welfare are large enough. This is also the reason that Sunstein always demands cost-benefit analyses for each case and stays avoids abstractions (Sunstein 2015a).

The most difficult challenge comes from the thick version of the Harm Principle. Unlike the thin version, which appears as consequentialism, the thick version looks more like a deontologist argument:

> The thick version of the autonomy position stresses not that freedom of choice is part of welfare but that it is an end in itself and thus decisive—or at least a very weighty matter, to be overridden only for the most compelling reasons. (Sunstein 2014, 127)
Ethical Concerns about Behavior-Steering Technology

The thick version does not only argue for the freedom of choice, but also defends autonomy in implementing that freedom. For the defenders of the thick version, any intervention, soft or hard, to means or to ends, is ethically problematic and therefore unacceptable. To deal with it, Sunstein offers two replies. First, it is not possible for a person to make decisions for everything; otherwise, his/her energy and time will be exhausted very soon before making important decisions. Autonomy can be guaranteed only when there is a “background” composed of many things that we take for granted, and “if active choosing were required for everything, our autonomy would quickly evaporate” (Sunstein 2014, 131). Second, the strong attachment to autonomy and the undifferentiated objections to interventions are mostly intuitive, which means they probably are the products of the automatic thinking system rather than the reflective one. For Sunstein, “in many contexts, the objection from autonomy may be a heuristic” and “what we really should care about is welfare” (Sunstein 2014, 135).

4.4. In Need of Material Behavior-Steering Technology

As pointed out in Chapter 3, while MBSTs in general are more qualitatively and quantitatively effective than IBSTs, they tend to create more serious ethical concerns about the loss of users’ freedom. The ethical guidelines or arguments developed for IBST, therefore, are unlikely to work for MBST.

On the one hand, given that MBSTs take effect through their material feature and thus are literally “hard” interventions, the argument that the freedom of choice can be maintained by reducing the cost of opting out is not applicable to them. To be sure, for small MBSTs, the argument can still apply. For example, one can argue that a speed bump is just a nudge because it is usually easy to find another way. But when an MBST appears in the form of architecture, such as the green building in Taiwan (Chapter 1), the cost will become significantly higher. Such an MBST can change behavior to a greater degree than a small MBST can, and is more useful for tackling the environmental crisis, but it leaves little room for human freedom.

On the other hand, one challenge concerning the environmental crisis is that not everyone agrees upon the value of nature. Even though most people agree, they rarely prioritize it. In other words, to improve people’s environmental behavior, interventions targeting merely means are far from enough; it is necessary to steer people’s ends. Although the guidelines for IBST hold that IBSTs should not decide ends for users, in principle an ends-intervening IBST is not as worri-
some thanks to the low cost of skipping it. This is, again, the reason that an IBST is less helpful for environmental problems. In short, the problem of ends-intervention becomes more salient and serious when it comes to an MBST.

Are we allowed to design and implement MBSTs for sustainability even though they may invade people’s freedom to choose their ends? Returning to the Harm Principle might help. We may argue that nature-unfriendly behavior can harm others because it would worsen the environment in which others live. In so doing, ends-intervening MBST can be justified. However, this is not an easy path to take. Given that climate change results from the accumulation of CO₂ that has been emitted from every corner of the world for many decades, evaluating how much a person’s CO₂ emission contributes to the increase in temperature is a complicated task; this is why most people have difficulty in relating their behavior to climate change (Kollmuss and Agyeman 2002). Worse, it will become much harder if we try to identify the causal relationship between who takes part in creating climate change and who is harmed by it. This is an intra-generational problem. Once we consider future generations who may or may not be harmed by climate change, the identifiability problem will become extremely complicated. The temporal and spatial dispersion makes climate change a “perfect moral storm” (Gardiner 2011). The inability, if not impossibility, to answer the questions: “Who is harmed?” and “Who is harming?” create a dead end to the appeal to the Harm Principle for justifying the design and implementation of green behavior-steering technologies.

One may believe that all the difficulties would be gone, or at least become much easier to solve, if we only look for help from IBST. MBSTs may be not necessary because IBSTs can be made very effective, qualitatively and quantitatively, if they are personalized. Indeed, for persuasive technology and nudge, personalization is often the best way to strengthen their effectiveness. But as pointed out earlier, personalization requires data collection and therefore can induce another ethical concern, privacy. Moreover, personalization works mainly because the messages sent by an IBST are in accord with what the users already care about (Hamari, Koivisto, and Pakkanen 2014); this may strengthen their previous preferences. For example, for those who care mainly about money, the message “Less driving, more saving!” is more powerful than “Less driving, more health!”, and for those who care a lot about health, the message “Less driving, more health!” is more persuasive than “Less driving, less CO₂ emission!” In other words, the personalization of IBST appeals inevitably to individual interests. But this is inconsistent with the nature of pro-environmental behavior, be-
cause “[i]n most cases, pro-environmental behavior does not maximize individual interests, but mainly benefits other people or the environment” (De Groot and Steg 2009, 61) As a result, personalized green IBSTs may not really improve environmental behavior in the long-run. They only repeat what many governmental campaigns have been doing and therefore may get the same outcome:

These campaigns often fail to promote sustainable behavior because they are designed to motivate egoistic considerations only. By doing so, it is likely that egoistic values are being prioritized, while altruistic and biospheric values are perceived as less important in that context. (De Groot and Steg 2009, 64)

For those who complain that MBSTs can only create “performers” of pro-environmental behavior, who do not learn about the value of nature and the virtue of unselfishness, it should be noted that personalized IBSTs may not do a better job.

Regarding the worries about transparency, IBST is more ethically problematic than MBST. On the one hand, some IBSTs can only work if the mechanisms are kept secret. Users may retract their trust to such IBSTs once they find that there is something hidden. For these users, any IBST working in this way is not ethically acceptable. On the other hand, with the development of ICT (information and communications technology), personalized IBSTs often build themselves on some algorithms. Even though engineers make the algorithms public, they are not understandable to most users. A smart building with ambient intelligence detecting and analyzing users’ walking patterns can remind them to take the stairs, as does a building in which the stairs are highlighted and the elevator is hidden. In the former, users do not see any algorithms or even detectors, but in the latter what is steering them is visible. In short, if the problem of transparency is our major concern, then we should choose to design and implement MBSTs rather IBSTs.

It is now clear that we are faced with a dilemma: while MBST is indispensable in tackling the environmental crisis, the guidelines and arguments we already have for IBST cannot be successfully extended to its design and implementation. Is it possible to develop guidance and defense for MBST in such a difficult situation? To solve the dilemma, I suggest engaging directly with the concept of human freedom.
4.5. Human Freedom under Material Behavior-Steering Technology

For the opponents, MBST is inevitably an attack against human freedom. A possible reply to this objection is to put freedom into a concrete context and take a closer look at what the freedom is in a specific scenario. The tale of Odysseus (also known as Ulysses) and the Sirens, for example, can demonstrate this point. In his journey sailing home (the Island of Ithaca), Odysseus, a hero in the Trojan War, had to pass by an island where the monsters Sirens stay. The Sirens, who sing with golden-sweet voices and beautiful melodies, can attract men who are listening and drown them. To safely pass by Sirens, Odysseus stopped his shipmates’ ears with bee wax and had himself lashed to the mast. By so doing, the shipmates could not hear Sirens singing and therefore weren’t affected, and Odysseus could not be moved to drown although he heard the songs. All the men survived as a result. Was the freedom of Odysseus and his mates compromised?

An argument made by Michael Nagenborg (2014) may help us to answer the question:

[W]e might have a choice to do X or Y. While X helps us to achieve an important goal, doing Y is so much more fun. Actually, Y is very tempting. Hence, any self-imposed measure that makes it more likely that we do X may be considered as an enhancement of our freedom to do X. (Nagenborg 2014, 46)

With this argument, we could say that in the story of Odysseus and the Sirens, X is going back to the home island and Y is listening to the beautiful song. Therefore, for Odysseus and his shipmates, the freedom to approach the Sirens was restricted, while the freedom to approach their home island was not. The goal of Odysseus and his mates was to pass by rather than pass away, and of course the Sirens were very tempting. But the wax and the rope, playing the role of MBSTs, enhanced their freedom to return home safely. In other words, whether freedom is compromised depends on what freedom we would like to focus on, and in some cases, such as the one of Odysseus and Sirens, we could even say that freedom is increased rather than decreased.

Such an argument, actually, is not far from Sunstein’s argument for nudges. In the case of Odysseus, tying oneself with rope is hard intervention, but it is done only to intervene with the means. The end was decided by Odysseus himself without any interference. Although the shipmates, whose ears were filled with bee wax, probably did not truly decide by themselves but took Odysseus’
command instead, it is reasonably plausible to assume that none of them wished like to give up his life. Therefore, for Odysseus as well as his shipmates, what were steered were only the means rather than the ends. As in the similarity between the two arguments, the flaw of Sunstein’s argument applies here as well. In other words, the problem of ends-intervention by MBST has not yet been solved.

Another reply to the objection that human freedom is compromised by the design or use of MBST is distinguishing freedom into different kinds. It has been widely recognized, according to Isaiah Berlin (1979), that there are two kinds of freedom: negative and positive freedom. Negative freedom is about the situation in which a person makes a choice. When there is no constraint preventing or pushing the person, s/he has complete negative freedom. Positive freedom, in contrast, refers to a person’s ability to decide and reach his/her goal. An example offered by Brey (2006) explains this distinction well: While a king enjoys negative freedom in his kingdom because of no restriction for him to do what he wants, he may not have as much positive freedom because of his father, “to whom he is psychologically dependent to a large extent: all his choices and actions in life are in fact mere attempts to please his father rather than the reasoned, autonomous choices of a free agent” (Brey 2006, 359). This conception of positive freedom is equivalent to autonomy, referring to whether a person can choose his own ends. As Brey (2006) points out,

\[\text{[a]utonomy becomes threatened when essential parts of our lives are conditioned by behavior-steering technologies, to such an extent that they do not merely interfere with our actions, but go as far as to shape and condition our plans and goals. (Brey 2006, 360)}\]

In other words, as long as a behavior-steering technology only influences users’ means, it can be defended because their ends are not altered. Under the influence of such a behavior-steering technology, freedom may be compromised, while autonomy is however preserved.

With this line of argument, one can even argue that autonomy, namely positive freedom, can be improved by behavior-steering technologies. This is exactly how proponents of nudge defend the idea. For example, Sunstein (2014) uses a metaphor: that a nudge functions like GPS (Global Positioning System); it only helps drivers to navigate to the destinations they have planned to go, rather than choosing their destinations. It is also argued, similarly, that if nudges are designed to facilitate one’s pursuit of one’s own goals, then the objection can be
Design for Green

weakened because autonomy is understood as “the capacity for an individual to
determine and pursue her own conception of the good according to her own
will” (Mills 2015, 497). This defense may prove helpful for the MBST that is de-
signed and implemented to steer the users who know and prioritize the value of
nature; however, it again does not really help us answer the question of
ends-intervening MBST. As pointed out previously, one reason for putting be-
behavior-steering technology into our toolkit is the need to move those who do not
care about or even intentionally behave against nature, and such a task cannot be
done without strong intervention regarding their ends. Even if we should not
leave those people behind, the objection to ends-intervening MBST remains to
be solved.

Yet another reply to the objection is to recognize the fact that there is always
intervention, whether intended or not. As shown by the theory of technological
mediation, human perception and action are always mediated by technologies
(Verbeek 2005). Technology plays the role of mediator between humans and the
world, and therefore has existential impacts on humans. There is no freedom
without technology, but only freedom with technology. Given this fact, ethical
concerns can be shifted from the question of how to protect human freedom
from technology to the inquiry of what relations humans would like to have with
technologies (Verbeek 2006a). Such a fact has also been observed in the context
of nudge:

It is essential to see that some form of choice architecture is inevitable.
Human beings cannot wish it away, however much they are committed to
freedom and autonomy. Any store has a design; some products are seen
first, and others are not. Any menu places the options at various locations.
Television stations come with different numbers, and strikingly, lower
numbers are better, even when the costs of switching are vanishingly low;
people are more likely to choose a station numbered 2 or 3 than one
numbered 150 or 200. Any website has a design, which will affect what
and whether people will choose. Streets, street signs, computers, cell
phones, and ballots offer choice architecture of their own. (Sunstein
2015a, 512)

By recognizing the fact that technological intervention, be it technological me-
diation or choice architecture, is pervasive, one can further argue that behav-
ior-steering technology can and should be designed to “recover” people’s
autonomy. For example, many people become smokers largely because of pre-
vious manipulation such as advertising and social fashion. Governmental offi-
cials can then claim that their intervention is a way to make a balance between
encouraging and discouraging smoking. If the officials do nothing, people are left alone to cigarette industry’s manipulations (Sunstein 2016). In such a context, then, intervention is more about balancing rather than guiding.

Such an argument can be applied to green MBST. For example, most people do not choose driving at the moment of leaving the house, but they have already been made to prefer driving in the first place. The built traffic infrastructure made people learn that driving is convenient and biking is tiring. Moreover, they are likely to prefer driving and will therefore vote for those senatorial candidates who vote for car-centered urban planning. The planned city, afterwards, will again strengthen people’s original preference for driving. This is exactly what we are faced with today. We have been living for centuries in an environment that our ancestors built without knowing its damage to nature. As a result, most of us are shaped to behave as if we know nothing about such damages. From this perspective, we could argue that green MBSTs in fact help us to get back our autonomy. However, this arguments is quite similar to the argument of Nagenborg (2014) mentioned above; it appeals to “freedom-to” rather than freedom in general. While it is plausible to assume that Odysseus and his shipmates prefer the “freedom-to-stay-alive” (so that they can go home) in the tale, the assumption that people would prefer the “freedom-to-behave-pro-environmentally” is contrary to the fact. In other words, although MBST can be understood as a way to reach a balanced autonomy, it is however ends-intervening.

To be sure, the arguments above still help, at least for some cases where the ends of people are not so disputed, but they have difficulties in dealing with green MBST for sustainability. As long as the discussions are framed around human freedom, be it the negative or positive one, the questions and objections concerning the intervention to ends will come to the forefront again and again. It turns out that we still find little room for the design and implementation of MBST to improve environmental behavior. Focusing on human freedom, therefore, seems to become an obstacle standing in our way toward realizing a sustainable future. I would like to suggest a strategy to deal with it, namely, to circumvent it. The question is how. Getting a better understanding of the nature of human behavior can be first step.

### 4.6. Concluding Remarks

The approach of behavior-steering technology often induces ethical concerns. Among them, the worry about the erosion of human freedom is central, and
from which other concerns can be deduced and derived. To prevent the design and implementation of persuasive technology going too far, the Golden Rule has been referred to and discursive ethics has been applied in order to develop ethical guidelines. Similarly, nudge has been strictly defined as a soft intervention to people's means, while the Harm Principle also gets deconstructed in order to detach nudging from the tag of manipulation. However, such guidance and defense for IBSTs can hardly be applied to MBSTs that are designed for sustainability. On the one hand, while human freedom can be preserved due to the low cost of opting out of IBSTs intervention, an MBST often leaves little room for users to go their own way due to its physically concrete nature. On the other hand, as environmental problems are temporally as well as spatially dispersed, and thereby tend to incapacitate current human-centered moral systems, a behavior-steering technology for sustainability cannot avoid being designed to be ends-intervening. Avoiding ends-intervention with those who have not learned from environmental education and those who have not been convinced by environmental ethics will miss the opportunity to solve the environmental crisis. Green MBSTs, therefore, are indispensable. However, as technological ends-intervention, MBSTs pose a serious challenge to human freedom. To make effective use of MBSTs for sustainability, we may need to circumvent the challenges by adjusting our current understanding of the nature of human behavior.
5 Technology of Behavior:  
B. F. Skinner’s Thoughts and Defense

Understanding the nature of human behavior is crucial to explore the possibilities and limitations of behavior-steering technology. The theory developed by the psychologist B. F. Skinner, a pioneer in studying the behavior of organisms in a scientific way, can offer us an entry point to further discuss the dilemma of designing and implementing behavior-steering technologies. In this chapter, I will first introduce Skinner's study, recapitulating his experimentation and theory on behavior. I will then focus on his most two famous books, *Walden Two* (1976a/1948) and *Beyond Freedom and Dignity* (1971); both represent his thoughts and defense concerning behavioral modification. In the former, Skinner turned his discovery of the principles of behavior into a blueprint for behavioral engineering; in the latter, he constructed various arguments to answer the worries and objections that had been raised after the former was published. By looking into the two books in detail, it will become clear that Skinner’s views can underpin the significance and indispensability of behavior-steering technology and that his defense helps to untangle the dilemma caused by MBST. However, Skinner’s thoughts are not flawless. Concerning the practice of the design and implantation of behavior-steering technologies, as I will finally argue, leaves two important questions for us to answer.

5.1. A Science of Behavior

B. F. Skinner (1904-1990), a famous behaviorist, is widely recognized as the most influential psychologist of the 20th century (Haggbloom et al. 2002). By conducting various experiments to train pigeons, Skinner explored the nature of animal behavior and developed a systematic theory to explain how behavior can be learned and modified. Rather than limiting his theory to animals inside laboratories, Skinner extended it to humans and suggested that human behavior can and should be modified in the same way. Although most people had no doubt regarding the validity of his experiments and theory on animals, his ambition to take theory and practice to the outside world provoked worries and objections. While he was chosen as the person of *Time Magazine*’s cover story in September
1971, more than half of the articles on him and his ideas took a hostile stance, criticizing him as a threat not only to American society, but also to humankind. Today, Skinner’s ideas are still unwelcomed, becoming just a few paragraphs or merely a “box” in the textbooks for undergraduates in psychology (Rutherford 2009). However, considering that our living environment is mostly human-made and with growing interest in influencing people’s behavior through design, it is time to revisit Skinner’s views to see how they can help with the issue of behavior-steering technology.

Let me begin with a typical Skinnerian experiment. To train a pigeon, a researcher puts a pigeon into a box (usually called a Skinner box), which has a small window and a hole on its wall. The researcher then presents the pigeon with the sign “+” through the window; as soon as the pigeon pecks the “+”, it receives a crumb of bread through the hole. After doing so several times, the pigeon will peck the “+” every time as long as there is a “+” presented. This behavior remains steady no matter where the position of window is and even though the pigeon only gets a crumb every three or four pecks. Up to this point, the pigeon has learned how to respond to the sign “+”. Through similar procedures, a pigeon can be trained to perform various behaviors, for example, turning in circles when seeing the sign “@” or bobbing its head when there is a red light shining. Skinner himself even successfully trained pigeons to play Ping-Pong and, more amazingly, guide a missile (Skinner 1979).

Through such experiments, Skinner (1966/1938) concluded that behavior (how an organism responds to the environment or a thing presented in the environment) is shaped by its consequence (how the environment responds to the behavior). For generalization, Skinner developed his terminology to describe what happened during the experiments. The process of learning was called “reinforcement”, the crumb of bread “reinforcer”, and the whole settings (Skinner box and food provider) “contingency of reinforcement”. While he admitted that animal behavior was relatively simple and the real-world environment was more complicated, Skinner insisted that the principle can be applied to human behavior. If the researcher used the word “PECK” instead of the sign “+” and “TURN” instead of “@” in the experiments above, as Skinner did, what we would see is the pigeon that can “read” the words and “knows” the meaning of each. For

---

14 When the same pigeon acquired both the behavior of “pecking when seeing ‘+’” and the behavior of “turning when seeing ‘@’”, it meant that the pigeon learned to discriminate between different signs.
Skinner, any learned behavior, be it animal or human, is the product of the organism’s environment, no matter how purposive or intentional it looks. If it is fully sufficient to explain animal behavior with the effect of environment, there is also no need to look for an “inner man”, be it a soul, an ego, or a mind, as an explanation to human behavior.

For Skinner, attributing human behavior to something observable, measurable, and manipulable is truly scientific, and this should be the main, if not the only, task of psychology. To be sure, his philosophy of science is not without flaws, according to several studies (see, for example, Sagal 1981). The importance is, however, that his discovery is not merely a theory to explain, but also a way to change, human behavior. For Skinner, the fact that behavior is a function of the contingency of reinforcement implies that a change in contingency can induce a change in human behavior. This is where Skinner went outside the ivory tower and got attention in the mass media. His ambition to put the theory into practice and to change human behavior to create a better world is embodied in his utopian novel *Walden Two*, first published in 1948.  

5.2. How to Change Human Behavior: Walden Two

Among Skinner’s books, *Walden Two* is the most popular one. It had sold over 2.5 million copies by 1990 (Bjork 1997), and it has more than 4,000 reviews on the website Goodreads, “four times as many as Skinner’s next most reviewed book, *Beyond Freedom and Dignity*” (Sundberg 2015, 43). The title of the book is an allusion to Henry Thoreau’s *Walden*; the “Two” implies that the place is wonderful too, worth the name as a second Walden, and most importantly, made for a group of people, at least two.16 Walden Two is a small community with about a thousand people, constructed by following Skinner’s theory of behavior. To emphasize the possibility and feasibility, it was located in America instead of

15 According to the new “Forward” for the second edition of *Walden Two*, Skinner had no plan to bring his theory into practice when writing his first theoretical book on behavior: “In *Behavior of Organisms*, I had refused to apply my results outside the laboratory” (Skinner 1976a/1948, v-vi).

16 “The difference is that Walden (One)—if you will permit me to call it that for clarity’s sake—was a utopia for one. Thoreau was no hermit (he could walk into Concord—to the post office or the lyceum—whenever he felt like it), but he never came to grips with the problems which arise when people must interact with each other.” (Skinner 1978, 191)
an isolated island, and the whole story was also set to take place in the mid-20th century, rather than a remote future.

The story begins with a plan to visit Walden Two by a group of outsiders: the psychologist Burris (the narrator of the story), the philosopher Castle, and two young couples: Roger and Barbara, Steve and Mary. They have a four-day trip around the community, and their tour guide is its founder and main planner, Frazier (representing of Skinner), who was the classmate of Burris when studying psychology in college. During the visit, Frazier shows them the physical and non-physical designs of Walden Two, but the philosopher Castle keeps interrupting him by raising ethical objections. Castle, apparently, plays the role of an opponent in the book, drawing forth defensive arguments from Frazier's mouth. In contrast, the psychologist Burris is less hostile and takes a pragmatic attitude toward Frazier's project. At the beginning, all six visitors feel more or less doubtful about the success of Walden Two, but after numerous debates between Castle and Frazier during the visit, three of them turn out to be convinced by Frazier and decide to stay as new members of the community, including Burris. One of the two couples has not decided yet but obviously has been moved. The only refuser, who strongly opposes the idea and practice from beginning to end, is Castle.

Frazier’s secret is what he calls “behavioral engineering”, a combination of the knowledge and skill in modifying human behavior and is strictly based on experimentation. For Frazier, behavioral engineering, like all other disciplines of engineering, is concentrated on problem-solving rather than phenomenon-explaining, and experimentation is the best approach, if not only, to test solutions. However, this is not what Castle thinks. On an occasion of afternoon tea, when Frazier showed a new type of teacup to the visitors, demonstrating that it becomes more useful than ordinary ones because the design of it results from a series of trial-and-error, Castle responded sarcastically:

“That's all very interesting”, said Castle, “but I hope you aren’t going to attribute the success of your community to trivial technical achievements of that sort. After all, a slight improvement in a tea service won’t shake the world.”

“We shake the world in other ways”, said Frazier, without a smile. “The actual achievement is beside the point. The main thing is, we encourage our people to view every habit and custom with an eye to possible improvement. A constantly experimental attitude toward everything—that's all we need.” (Skinner 1976a/1948, 25)
More crucially, how can we modify the behavior of the residents? For Frazier, modifying adults’ behavior is much harder than modifying children’s because adults already have their own history of the contingencies of reinforcement. Therefore, Frazier puts efforts on children, and attempts to complete ethical training “by the age of six” (Skinner 1976a/1948, 98). A vivid example of behavior modification shows how Frazier’s idea works: How to teach children to be self-controlled in front of the temptation of lollipops?

“First of all, the children are urged to examine their own behavior while looking at the lollipops. This helps them to recognize the need for self-control. Then the lollipops are concealed, and the children are asked to notice any gain in happiness or any reduction in tension. Then strong distraction is arranged, say, an interesting game. Later the children are reminded of the candy and encouraged to examine their reaction. The value of the distraction is generally obvious ... When the experiment is repeated a day or so later, the children all run with the lollipops to their lockers ...” (Skinner 1976a/1948, 98)

What is noticeable here is the method that children have learned to deal with the temptation of lollipops. The children were not told to suppress their desire for the lollipops; instead, they were asked to inspect the relationship between their feelings and the surroundings, and to notice the arrangement that helps them to “forget” the lollipops. In other words, what they indeed learned is to change the environment outside their body, not to tame the invisible desire within. This is exactly the basic principle for behavioral modification: in order to change behavior, change the environment rather than the mind. The children, having been exposed to the contingency of reinforcement that was arranged by the behavioral engineer, learned how to arrange a contingency of reinforcement for themselves once there is similar temptation in the future. They became self-controlled, not because they were able to control themselves, but because they were able to control the environment.

Frazier also adds that such a reinforcing effect can be transferred from the arranged environment to the people who live in the community. When the children who had learned to resist temptations and regarded self-control as a good behavior (what a person ought to do in front of temptations) grew up and become the main members of the community, they would create peer pressure to any old or new members who just surrender to temptations by uttering something like “It's bad!” or “You should not!”. In contrast, those who successfully resist temp-
Design for Green

...tations received verbal encouragement like “It’s good!” or “You are doing right!”.

In other words,

[the control of the physical and social environment, of which Frazier had made so much, was progressively relaxed—or, to be more exact, the control was transferred from the authorities to the child himself and to the other members of his group. (Skinner 1976a/1948, 107)]

However, this does not mean that such training can be done once for all. The example of sheep management in the story can illustrate this point. On the first day of the visit, Frazier showed the visitors that a flock of sheep can be easily kept together by using electric fence only for the first generation. The lambs that were born after the electric fence was replaced by ordinary string and never experience an electric shock, still behave as their elders. Frazier called this the “force of tradition” (Skinner 1976a/1948, 15-16). However, on the last day, one of the sheep escaped through the ordinary fence and caused a little chaos in the community. Apparently, the electric fence is still needed, even though it does not have to be electrified all the time. In other words, the environment is indispensable, because it is the primary force that shapes the behavior.

“People are not sheep!”, one may protest. Indeed, many objections like this have been raised against applying the method used to train animals on humans. In fact, Skinner does not stand on the other side. There are important distinctions between positive reinforcement, negative reinforcement, and punishment. When a behavioral engineer arranges a contingency of reinforcement to increase the likelihood of a certain behavior by providing the subject with something that he/she likes, then he/she is undergoing positive reinforcement. In such a process, the subject performs the behavior to get the thing, and the thing is called positive reinforcer. In contrast, if a contingency of reinforcement is arranged to increase the likelihood of a certain behavior through removing something that the subject hates, it is the situation of negative reinforcement. In such a process, the subject performs the behavior to avoid the thing, and the thing is called negative reinforcer. Different from the two types of reinforcement above, punishment is made to decrease the likelihood of a certain behavior. In the situation of punishment, the subject stops performing the behavior to avoid unpleasant response, be it an electric shock or the loss of food. Having these distinctions in mind, we can understand now how Frazier answered to the escape of the sheep:

“It doesn’t work, even with sheep, you see”, he (Frazier) said.

“What doesn’t?” [said Burris.]
“Punishment. Negative reinforcement. The threat of pain. It’s a primitive principle of control. So long as we keep the fence electrified, we have no trouble—provided the needs of the sheep are satisfied. But if we relent, trouble is bound to arise sooner or later.” (Skinner 1976a/1948, 283)

Because there is something aversive to the subject in both the situations of punishment and negative reinforcement, the subject tends to escape from them, and the effects on behavior change would diminish. For Skinner, it is not so problematic to train animal through negative reinforcement or punishment; although both are only effective in the short run, people are “not raising for the good of the sheep” (Skinner 1976a/1948, 283). An indirect message made by Skinner, unsurprisingly, is that only positive reinforcement should be applied to humans. It is a mistake, although quite common, to think that Skinner’s theory of behavioral modification has a great connection to punishment.

As a technology-based utopia, Walden Two may remind readers about another technology-based place, Brave New World. While Walden Two is a utopian novel, Brave New World (Huxley 2006/1932) is entirely a dystopian one. If both fictional places are constructed by controlling human behavior through psychological technology, what are the differences between them? Skinner was clearly aware of such a question during his writing, so he made a quick discussion between Frazier and Castle for readers. After noticing that the members of Walden Two are willing to do unpleasant jobs and think that all kinds of work are equally desirable, Castle tried to hint the likeness between Walden Two and Brave New World:

“I suppose you put phonographs in your dormitories which repeat ‘I like to work in sewers. Sewers are lots of fun.’”, said Castle.
“No, Walden Two isn’t that kind of brave new world”, said Frazier. “We don’t propagandize. That’s a basic principle ...” (Skinner 1976a/1948, 46, italics in original)

Although a clear explanation does not follow, we can find a related answer in another discussion. Burris once suggested that Walden Two is quite like communist Russia in its effort to shape people’s behavior, so he raised a question about the failure of Russia. Frazier agreed that “[a]s originally conceived, it was a good try” (Skinner 1976a/1948, 258), but he then pointed out several mistakes the Russian government had made. One of them “is the decline in the experimental spirit” (Skinner 1976a/1948, 258), and another is that “Russia has over propagandized, both to its own people and to the outside world” (Skinner
1976a/1948, 259). According to Frazier, these two reasons are related. Because propaganda “has made it impossible to evaluate their success. We don’t know how much of the current vigor of Russian communism is due to a strong, satisfying way of life, and how much to indoctrination” (Skinner 1976a/1948, 259).

In other words, while experimentation is to check whether an arrangement works, propaganda may create false signals, and therefore disturb the check. Having been propagandized, people may think that they like doing something, but in fact they would not do the same thing if there were no propaganda. For Frazier, the collapse of communist Russia was inevitable because eventually the government did not and could not create effective contingencies of reinforcement. In the same vein, Brave New World is inevitably a dystopia. No propaganda, therefore, is a must in Frazier’s project and what differentiates Walden Two from Brave New World.

5.3. Against Worries: Beyond Freedom and Dignity

Beyond Freedom and Dignity was published in 1971, 23 years later after Walden Two. By authoring this book, Skinner responded in a systematic way to the worries and objections against his idea in Walden Two. He changed the name for the knowledge and skill of behavioral modification from “behavioral engineering” to “technology of behavior”\(^{17}\) and offered various arguments for the theory and practice of such a technology.

For Skinner, the key to establish and develop the technology of behavior is to dispense with the assumption of an inner man. Such an inner man usually takes the shape of soul, ego, or mind, but none of them are truly reliable as an explanation to human behavior. As it “has been created in the image of the outer” (Skinner 1971, 21), what the inner man does is exactly the same as what the person we seen does, and therefore it explains nothing about the behavior. Attributing a person’s behavior to the inner man is easy but in fact lazy, and it can create a barrier for further study on human behavior: “Unable to understand how or why the person we see behaves as he does, we attribute his behavior to a person we cannot see, whose behavior we cannot explain either but about whom we are not inclined to ask questions” (Skinner 1971, 19).

\(^{17}\) In the rest of dissertation, I use “behavioral engineering” and “technology of behavior” interchangeably.
In other words, for Skinner, the future development of psychology should probe into things outside rather than inside the human body. At this point, Skinner took physics and biology as models, arguing that their advance began from the disposal of something invisible and inside:

We can follow the path taken by physics and biology by turning directly to the relation between behavior and the environment and neglecting supposed mediating states of mind. Physics did not advance by looking more closely at the jubilance of a falling body, or biology by looking at the nature of vital spirits, and we do not need to try to discover what personalities, states of mind, feelings, traits of character, plans, purposes, intentions, or the other perquisites of autonomous man really are in order to get on with a scientific analysis of behavior. (Skinner 1971, 20)

Although recognizing Skinner’s effort on the scientification of psychology, several famous intellectuals, such as Carl Rogers (1956) and Noam Chomsky (1959), have attacked Skinner’s viewpoints, criticizing him for neglecting the fact that every human is born with something already in-built, be it a disposition or a linguistic structure, and inevitably driven by it. However, Skinner has never denied that there may be something that are in-built already; though they are still the product of the environment (Skinner 1971, 101–9). For Skinner, humans’ natural tendencies to behave in a certain way are the result of the evolutionary history of the species; although natural selection is a very slow and long process, it is also a kind of reinforcement. For example, being careful (the tendency to perform the behavior of looking around and hearing frequently) helps to prey on small animals (positive reinforcers) and avoid from big animals (negative reinforcers), early man who had such a tendency survived more easily and thus could pass such a tendency to the next generation. In other words, the natural environment is also a kind of contingency of reinforcement and has created whatever humans are born with today. However, it should be noted that although Skinner acknowledges the power of the natural environment, he believes that social environment, including cultural practices and built surroundings, has more immediate, direct, and stronger effect in shaping human behavior, especially in the age of technology.  

Skinner takes a similar point of view on the mental status of humans. He has never denied the existence of mental phenomena, such as purposes, intentions,

\[\text{18}\] Accordingly, it is unfair to say that Skinner’s blueprint cannot succeed because it neglects human nature (for example, Swirski, 2012).
or impulses; rather, he regards them as behavior as well. In other words, what he really rejects is to take them as *explanans* for human behavior. For him, any mental status of humans is no more an *explanandum* than behavior itself. This is the reason that Skinner calls himself a “radical behaviorist” in his latter book, distinguishing his behaviorism from his predecessors’ (Skinner 1974). Moreover, for the purpose of modifying behavior, Skinner even suggests that it is fully sufficient to see a person as a “black box” as long as we can grasp the relation between inputs (the contingencies of reinforcement) and outputs (the person’s behavior).

Attributing human behavior to the environment instead of an inner man implies that there is no such thing as freedom because humans are not in control of themselves, but in control of environment. Indeed, there is no room for freedom in his theory of behavior, and Skinner has made this point clearly through Frazier in *Walden Two*:

“*My answer is simple enough*”, said Frazier. “I deny that freedom exists at all. I must deny it—or my program would be absurd. You can’t have a science about a subject matter which hops capriciously about.” (Skinner 1976a/1948, 241-42)

To answer the defenders of freedom, Skinner conducted literature review on the advocacy of the importance of freedom, and argued that it “often emphasizes the aversive conditions under which people live, perhaps by contrasting them with conditions in a freer world” (Skinner 1971, 35). In other words, “[w]hat we may call the ‘literature of freedom’ has been designed to induce people to escape from or attack those who act to control them aversively” (Skinner 1971, 34) As a result, the concept of freedom developed from the literature has very little to say about non-aversive control, under which people feel free and do what they want but are still under control. The term “freedom”, therefore, is suspicious for Skinner. On one hand, “We have no vocabulary of freedom in dealing with what we want to do ... The question never arises” (Skinner 1976a/1948, 247). On the other hand, striving in the name of freedom may increase the misuse of positive reinforcement, which means that the controller provides the reinforcers that are positive in the short run but has deferred negative consequences. The defenders of freedom will have no reason to protest against such a control until the nega-

---

19 There are also “methodological behaviorism” and “analytical or logical behaviorism”, which are both covered by radical behaviorism (Graham 2016).
tive consequences shows up. Even though the negative consequences eventually become clear, it is hard to connect them to the original reinforcer because during the period the same people are exposed to other contingencies of reinforcement.

Moreover, as human behavior is the product of its environment, it is not possible to eliminate control. For Skinner, in addition to nature’s control, behavioral control happens everywhere in society. Unfortunately, “in one form or another intentional aversive control is the pattern of most social coordination—in ethics, religion, government, economics, education, psychotherapy, and family life” (Skinner 1971, 33). Rejecting control in the name of freedom is no better than leaving people to those punishment and negative reinforcement. In other words, the question is not how to get rid of control, but how to control; and the answer, apparently, is through positive reinforcement.

Positive reinforcement, however, is objected by another kind of literature. Such a literature, called the “literature of dignity” by Skinner, is concerned with preserving due credit for the person who performs good behavior because “[w]e are not inclined to give a person credit for achievements which are in fact due to forces over which he has no control” (Skinner 1971, 48). For example, the children who arrive at school on time due to the reminder or warning from their parents usually get no commendation, while those who are punctual without being urged every morning do. Similarly, “[w]e do not give a writer much credit for a potboiler, or an artist for a picture obviously painted to sell in the current fashion” (Skinner 1971, 50). We do not praise them because their behavior is not fully “voluntary”, just like “we do not give [a person] credit for coughing, sneezing, or vomiting even though the result may be valuable” (Skinner 1971, 49). However, the defenders of human dignity neglect the fact that praise and commendation are also reinforcers. When people get praised or commended for doing something, they are likely to repeat it. In other words, they are still under the control, although the control is less obvious.20

For Skinner, the reason that he was attacked for the advocation of positive reinforcement can be explained by his theory of behavior as well:

What we may call the struggle for dignity has many features in common with the struggle for freedom. The removal of a positive reinforcer is

20 Probably because praise and commendation are rarely regarded as a form of control, people are not interested in identifying the controller, and we do not have any word like “praiser” or “commender”.

71
aversive, and when people are deprived of credit or admiration or the chance to be commended or admired, they respond in appropriate ways. They escape from those who deprive them or attack in order to weaken their effectiveness. (Skinner 1971, 57)

Skinner also proposed that the defenders of dignity are inconsistent. While people refuse to accept the technology of behavior in order to keep credit for humans, they seldom raise objections to physical and biological technology that have reduced opportunities for people to be praised. For example, “medical science has reduced the need to suffer in silence and the chance to be admired for doing so; fire proof buildings leave no room for brave firemen, or safe ships for brave sailors, or safe aeroplanes for brave pilots” (Skinner 1971, 59). If such advances in technology are not objectionable, there is also no reason to resist the advance of technology of behavior, namely, the theory and practice of positive reinforcement. Moreover, in rejecting positive reinforcement, the possibility of effectively improving behavior would be missed. Any valuable behavior for which people would receive praise or commendation, therefore, would be badly postponed.

In Skinner’s line of argument, the literature of freedom and that of dignity are paired with one another. The coupling of them has created a weird situation. On the one hand, the defenders of human freedom throw the baby out with the bath water: in resisting aversive control to protect human freedom, they reject all forms of control, and therefore, neglect the forces that have pointed humans to the behaviors they like and want to perform. On the other hand, although the defenders of human dignity do not oppose being positively reinforced by the reinforcers such as praise and commendation, they reject any further positive reinforcers that may help them to perform admirable behavior just because such reinforcers seem to reduce the value of that behavior. The only acceptable force in both situations, therefore, is the one that are used to remove the behavior they do not want or valueless—namely, the force of punishment. Although none of the defenders of freedom or dignity are in support of the measures of punishment (because of they are aversive), they are ironically “largely responsible for the fact that [those measures] are still with us” (Skinner 1971, 64).

There is also a similar inconsistence in people’s attitude toward the theory of behavior: “Although people object when a scientific analysis traces their behavior to external conditions and thus deprives them of credit and the chance to be admired, they seldom object when the same analysis absolves them of blame.” (Skinner 1971, 78)
5.4. Skinnerian Approach to Environmental Behavior

The basic reason that Skinner’s thought is applicable to the approach of behavior-steering technology lies in its direct focus on the relationship between human behavior and physical conditions, without involving anything like mind or cognition in charge of knowing. As discussed earlier in Chapter 2, the reason that behavior-steering technology is indispensable for improving environmental behavior is the need to move those who have not been moved by the teaching of environmental education or the argumentation of environmental ethics. In other words, as a complement to the approach of education, in which the target is an “inner man”, the approach of behavior-steering technology does not need such a thing. To be noted, we do not need to go as far as Skinner in denying the existence of an inner man, who is assumed to be the initiator of behavior; otherwise, we have to deny the approach of education. What we do rather is to pretend that the inner man does not exist, because it is fairly sufficient for the approach of behavior-steering technology to see humans as just a black box.

Another reason to take a Skinnerian view is that it helps us to understand why the likelihood of changes toward pro-environmental behavior is usually low. As Skinner has stated, behavior is shaped by its consequences. For example, if you go out without wearing sunscreen on a hot day and therefore get sunburned, on another hot day you are unlikely to go out again or not use sunscreen. Or as you felt pain in your foot after kicking a rock, you probably will not kick another rock next time. However, that is not the case for anti-environmental behavior. As pointed out in Chapter 4, the environmental crisis is the result of the accumulation of numerous anti-environmental behaviors; it takes a very long time until negative consequences emerge. Worse, even if negative consequences show up, they may not be in the proximity of those who have been behaving anti-environmentally. In other words, the negative consequences of anti-environmental behavior are not only deferred but also dislocated, and therefore hard to be negatively reinforcing to those who performed the behavior. In the same vein, those who have behaved pro-environmentally rarely enjoy any positive consequences and therefore are not positively reinforced. Such behavior, according to Skinner, will eventually fade.

The way to deal with behavior that has only non-immediate consequences is, as we usually do, to provide non-primary consequences. For example, driving fast may create harm to the driver, but it only happens accidentally, so most drivers are likely to drive fast frequently. To improve the behavior of driving, what we usually do is to describe to drivers the scene of car accidents (negative
reinforcer), praise those who drive slowly as a moral driver (positive reinforcer), or legislate to fine those who drive fast (punishment). However, none of these is sufficiently effective because they are merely non-primary, which means none of them is the response made immediately by the environment itself. For the driver who has been told about a car accident, as long as s/he has never had a crash, s/he may still drive fast. For the driver who has been praised for driving slowly, unless s/he can be praised every time or has experienced a successful avoidance of a car crash, s/he may eventually discover the fun of driving fast. For the driver who has only been fined once among twenty times of fast driving, unless s/he crashes due to the fast driving, she is unlikely to become a slow driver.

In contrast to the methods above, the road integrated with a speed bump can provide a primary physical consequence to the fast drivers, the uncomfortable jolt. Such a road is a contingency of reinforcement, in which the jolt is a negative reinforcer, and the behavior of driving slowly can reduce its occurrence. To be sure, this example counts as negative reinforcement, which is not recommended by Skinner, and we should try to design roads in a more positively reinforcing form in the future. However, it is sufficient to illustrate the point here: MBST is essential not only because its effect takes a physical form, but also because it can be a part of the environment or is the environment in itself. In addition, the advantage of behavior-steering technology is apparent. For the road with a speed bump, it becomes unnecessary to spend time telling stories about car accidents, give the commendation repeatedly to the driver, or hire a police officer for a speed trap. As Skinner has pointed out, “[o]ne of the advantages in being dependent on things rather than on other people is that the time and energy of other people are saved” (Skinner 1971, 89). Moreover, “[a]nother important advantage of being dependent on things is that the contingencies which involve things are more precise and shape more useful behavior than contingencies arranged by other people” (Skinner 1971, 90). At this point, Skinner seems quite in advance of Latour (1992) in discovering the reliability and dependability of things.

While Skinner’s thoughts are highly relevant and supportive to the approach of behavior-steering technology, somehow he has been barely referred to in such a context. Neither the literature on persuasive technology nor the literature of nudge mentions his name. Outside both, the only article I have found is “De moralisering van de apparaten” (The moralization of devices), an article authored by the Dutch philosopher Hans Achterhuis (1995). In the article, Skinner was referred (along with Michel Foucault) to argue that technology can
be a power in shaping human behavior and therefore should be designed to help people behave morally. Rather than moralizing people, according to the article, what should be done is to moralize technology, which means a moral task can be built into the devices in our daily lives. A turnstile in the train station, for example, makes people buy the tickets and therefore prevents them from stealing a ride. Unfortunately, “Achterhuis’ approach was heavily criticized, especially because people were afraid that it would threaten human freedom” (Verbeek 2013, 90). The worry about the loss of human freedom varies in degree, as discussed in Chapter 4, depending largely on the type of behavior-steering technology. For IBST and means-intervening MBST, the worry can be eased through the guidelines and arguments made for persuasive technology and nudge. The problem that has not been answered yet concerns ends-intervening MBST.

One possible solution is to establish a democratic decision-making procedure, such as Constructive Technology Assessment, an approach recommended by Verbeek (2011, 112–13), or the advanced version of it, which is called Ethical-Constructive Technology Assessment (Kiran, Oudshoorn, and Verbeek 2015). By gathering all the relevant actors to deliberate and anticipate the possible consequences of technology-in-design, a democratic procedure looks for participants’ agreement on not only “for what” but also “how” they are going to be steered. As soon as agreement is achieved, the fear of the loss of freedom is dispelled. However, such a procedure may have difficulties with the environmental crisis, again because there are still deniers of climate change and we do not yet have an overwhelmingly convincing argument for the value of nature. In other words, the problem of ends-intervening MBST can hardly be solved by appealing to democracy.22

The problem probably can be solved by appealing to Skinner’s suggestion of positive reinforcement. The reason that negative reinforcement and punishment tend to cause worry about, and the fight for, freedom is that they provide something aversive to the person under control. People under such control do feel being controlled because they are not doing what they prefer doing, and therefore are likely to escape or protest. In contrast, under the contingencies of positive reinforcement, people “are doing what they want to do, not what they are forced to do. That’s the source of the tremendous power of positive reinforcement—there’s no restraint and no revolt” (Skinner 1976a/1948, 246).

22 The tension between environmental protection and liberal democracy will be further discussed in Chapter 8.
In other words, as long as a MBST is designed to provide positive reinforcers or as a contingency of positive reinforcement itself, the people who are steered by it are not really being steered because there are no conflicts between their ends and the ends that have been built into it. The problem of ends-intervening MBST, accordingly, evaporates. While positive reinforcement is still a kind of control, “[t]he curious thing is that in that case the question of freedom never arises” (Skinner 1976a/1948, 247, italics in original). It should be noted here that Skinner does not agree to the use of short-term positive reinforcement. For Skinner, such reinforcement is not really a positive one because in the long run its consequences are negative. People may feel good for a while after being reinforced by such reinforcement but will experience harmful consequences and eventually lose the good feeling.

It is said that even though behavior is completely determined, it is better that a man “feel free” or “believe that he is free”. If this means that it is better to be controlled in ways which have no aversive consequences, we may agree, but if it means that it is better to be controlled in ways against which no one revolts, it fails to take account of the possibility of deferred aversive consequences. (Skinner 1971, 44)

Just like we do not see drug takers or gamblers as free even though they may feel free when taking drugs or gambling and decided to do it on their own, people are not free under short-term positive reinforcement. In other words, real positive reinforcement closes the gap between “feeling free” and “being free”, and this is what Skinner has been proposing and advocating.

From such a strict Skinnerian perspective, if democracy is called for protecting human freedom in ensuring that designers do not abuse their power, it may be an unnecessary practice for the approach of behavior-steering technology. Even though the technocracy of behavior-steering technology may arise, there is nothing to worry about or oppose as long as the method of positive reinforcement is applied. This is exactly how Frazier replied to Castle’s doubt on the government of Walden Two, which consists of behavioral engineers as “Planners”:

Insofar as the Planners rule at all, they do so through positive reinforcement. They don’t use or threaten to use force. They have no machinery for that. In order to extend their power they would have to provide more and more satisfying conditions. A curious sort of despotism, Mr. Castle. (Skinner 1976a/1948, 256)
In short, Skinner’s insights help the approach of behavior-steering technology in many ways. On the one hand, it supports the view that human behavior can be changed without appealing to knowing; on the other hand, it provides further reasons for the necessity of MBST. Most importantly, it points us to the technique of positive reinforcement, by which the problem of ends-intervening MBST can be reduced and the danger of technocracy avoided. The worry that human freedom will be damaged under the effects of behavior-steering technology, accordingly, is dissipated.23

5.5. Unsolved Questions

Skinner’s thoughts are not without flaws. First of all, appealing to positive reinforcement alone is unlikely to be effective because being sustainable is more about the removal of anti-environmental behavior than the creation of pro-environmental behavior. Reinforcement, by definition, is to increase the likelihood of behavior, not to decrease it. In other words, it has less power to deal with the behavior that has been learned. This is why, in Walden Two, Frazier targets mainly at children. When he was asked by Castle about how to manage the current civilization that “is running way like a frightened horse” (Skinner 1976a/1948, 81), he could only reply that “[l]et her run till she drops from exhaustion” (Skinner 1976a/1948, 81) and that “[m]eanwhile let’s see what we can do to her lovely colt” (Skinner 1976a/1948, 81). However, his answer cannot be applied to the problem of the environmental crisis. Time is pressing, and the “dropping from exhaustion” of unsustainable lifestyle can be catastrophic. The problem cannot be solved in time if changes in environmental behavior begin merely from the next generation. In other words, changes in environmental behavior have to happen to our contemporaries, and it cannot be done without removing, or decreasing at least, anti-environmental behavior. At this point, a question pops up: If aversive measures are still needed, how can we deal with the

23 It is worth mentioning here that Skinnerian approach has nothing to do with physical or nonphysical “brainwashing”. In his commemorative talk for the honorary doctorate at Keio University, Japan, 1979, Skinner said: “To many people it is known as ‘behavior modification’, but the term has been widely misunderstood. I do not mean the modification of behavior with drugs or implanted electrodes, or Pavlovian conditioning with electric shock or nausea-producing drugs. The term was invented to refer to behavior changed through what the layman calls ‘reward’ or what, in the experimental analysis of behavior, we call ‘positive reinforcement’.” (Skinner 1991, 100)
problem of freedom that can only be solved by the technique of positive reinforcement?

Even though people raise no objections to being pro-environmental, there is still another key point. The fact that there are no objections to something does not mean the thing is good or right. In other words, while people still feel free, thanks to non-aversive control, can we say that the direction into which they are being steered is good or right? Can we justify the end, a pro-environmental society, by saying that it is good or right? We may find that Skinner helps little for such questions if we look into his view on values. For Skinner, there is no difference between saying “apples are good” and “being honest is good”, in the sense that both are reinforcing. But why they are reinforcing? It is because they are both the product of an evolutionary history. Our prehistoric ancestors ate everything in the beginning; however, those who tended to eat apples lived longer and had more children. In other words, apples have contributed to the survival of human species.

Parallel to the relationship between apples and the survival of human species (as who we are now), being honest has contributed to the survival of our culture (as it is now). Imagine a culture in which people often lie. The people there tend not to trust each other because being cheated can bring negative consequences. Such a culture will eventually collapse because collaboration is extremely hard. In contrast, a culture in which people tend to be honest has a greater chance to survive because the people can work together to solve the problems they are faced with. In such a culture, being honest is reinforcing and is called good. Just like the practices concerning apples have been repeated and developed, such as eating apples, asking why apples are good (a study on apples), and growing apple trees, the practices concerning being honest have been repeated and developed as well. We tell the truth as much as we can, ask why being honest is good (a study on honesty, that is, ethics), and teach children to be honest. While calling the good of apples “nutritious”, we call the good of being honest “moral”; apples are good because they are nutritious, and being honest is good because it is moral.

Such an explanation of good implies a severe problem: The moral value of a (new) cultural practice emerges only if the culture carrying the practice survives long enough. We cannot know whether “being pro-environmental” is good or not because we just started performing the behavior and inquiring the meaning of “being pro-environmental” about a century ago. We do not really know whether a “pro-environmental” culture will eventually survive or not, although,
paradoxically enough, we use the term “survival” to argue for a pro-environmental society. In other words, it is extremely difficult, for now at least, to justify the design and implementation of green MBST by saying that a pro-environmental society is good and therefore should be the goal which we should pursue. Moreover, just because “being pro-environmental” has not evolved to the degree that “being honest” has evolved, most people currently regard it as something unessential and having nothing to do with morality. As a result, a consensus on moving toward a pro-environmental society is still elusive.

Although Skinner’s concepts give room to the design and implementation of green MBST, they leave important problems for us to solve. If the removal of anti-environmental behavior is essential to tackling the environmental crisis but cannot be done through positive reinforcement, what can we do about the problem of human freedom? Can people still feel, or be, free under the aversive control of green MBST? Even if the problem can be solved, what can we do if the destination to which people are steered cannot be justified? Should we stop designing and implanting MBST, after all, because there is hardly a consensus on moving toward a pro-environmental society? The answers to these questions, although quite implicit, lie in Skinner’s experimental approach to behavioral engineering.

5.6. Concluding Remarks

B. F. Skinner’s views on behavior help to strengthen and improve the approach of behavior-steering technology that is developing. On the one hand, his Walden Two shows us the possibility and plausibility to modify human behavior through behavioral engineering, in which technological designs play a crucial role. Rather than the inner man, through whom education attempts to create behavioral change, it is behavior itself that should be the focus and worked on. His Beyond Freedom and Dignity provides us with clues to untangle the ethical dilemma concerning human freedom. Given that an organism’s behavior is highly the product of its environment, it would be pointless to obsess over the question of how to get rid of control. Both the literature on freedom and that of dignity take a hostile attitude to non-self-control, in fact they are incapable of dealing with non-aversive control and, ironically, have fostered the use of harmful punishment. The question of how to exercise control, therefore, is essential and relevant. Rather than punishment and negative reinforcement, the technique of positive reinforcement can not only largely reduce the objections and worries
about the loss of human freedom, but also shorten the distance between “feeling freedom” and “being freedom”. However, with regard to the environmental crisis, two issues are left. As solving environmental problems often needs the removal of anti-environmental behavior more than the creation of pro-environmental behavior, which cannot be done by positive reinforcement, the problem of aversive control tends to resurface. Moreover, parallel to the theory of evolution, Skinner’s view on value can hardly help to justify pro-environmental behavior as good or right, for which green behavior-steering technologies are supposed to be designed. The solutions to these issues may be found in his experimental approach to behavioral modification.
6 Behavioral Modification in Small Scale

Against the background of the global social recession in the middle of the 20th century, *Walden Two* shows that a better world is possible through behavioral engineering; it has inspired the enthusiasts and activists of social reform to build their own Skinnerian communities. Among the efforts, Twin Oaks in America and Los Horcones in Mexico are most widely noticed and have been often mentioned in mass media. In this chapter, I will firstly introduce the two communities and discuss why one of them has become hard to maintain but the other keeps thriving. By looking into their practices, we will see that what is crucial to social reform is behavioral change rather than political reformation. Then I will turn to question of “why community?”, arguing that the scale is crucial to Skinner’s blueprint on both the theoretical and practical levels. Behavioral modification can only start with a small sized group and a small number of adjustments. Finally, I will juxtapose Skinner’s approach with the philosopher Karl Popper’s “piecemeal engineering” to not only show the parallels between them, but also argue that the epistemological argument for the latter can help us to answer the questions that has been left in the last chapter.

6.1. Twin Oaks

Twin Oaks was initiated and planned mainly by Kathleen Kinkade, who was inspired by *Walden Two*. In the beginning, she, together with some other social activists, founded Walden House within a building in Washington, D.C. With newsletters, they attracted those who might be interested in their project. As the number of members increased, they expanded the project, moved to Virginia, and eventually started their community, Twin Oaks, in 1967. As described in Kinkade’s *A Walden Two Experiment* (1974), a record for the early years of Twin Oaks, the community was intended to imitate the structure of Walden Two. For example, all decisions concerning the communal life of Twin Oaks had to be made through a planner-manager system, and the residents had to reach the working hours that had been assigned according to labor-credit calculation.

Such an imitation, unfortunately, did not work well. The planner-manager system, for example, was attacked soon by those who wanted to take power and become “communards”, the governors of Twin Oaks. The closeness of adminis-
Design for Green

tative meetings was also dissatisfying to the majority. As a result, the system had to be changed into a more open model: majority votes for simple issues and deliberative discussions for complicated problems. However, the residents asked for more. More changes, therefore, were made for achieving a power balance between the communards and the residents. However, becoming a fully democratic system does not lead the Twin Oaksers to a more satisfying life, even though the worry concerning totalitarianism was gone. “[T]he biggest problem faced by Twin Oaks today in regard to its political system is not power struggles but political apathy … due to the fact that the decision-making process at Twin Oaks is slow, tedious, and unrewarding for those involved in it.” (Kuhlmann 2005, 100)

Twin Oaks can hardly be regarded as a successful try. There are only about 100 adults and 17 children living there (Greenfield 2015), and the residents have been saying that they sometimes feel “trapped” (Pinchbeck 1997). Moreover, the problem of lack of money for savings remains, and every year about 20 percent of the membership leaves (Clay 2017). Such failure is probably due to their insufficient attachment to Skinner’s theory of behavior:

[B]ehaviorism—which is, after all, the basic idea underlying Walden Two—was not all that important to those who founded Twin Oaks … [B]ehaviorism was not a major stumbling block for the communards, yet to them, Walden Two presented mainly a good starting point for redesigning society, a fresh approach that held the promise of solving at least some of society’s problems. Taking Walden Two as a manual for action, the communards’ main interest lay in trying out the suggested labor-credit and planner-manager systems. Their approach to Walden Two was thus characterized by extracting the proposed communal structure, with the more general idea of applying behavioral principles to the design of society being secondary. (Kuhlmann 2005, 87, emphasis added)

In fact, Skinner has already noticed such a mistake in his “Foreword” for A Walden Two Experiment:

Is the result a Walden Two? Not yet, says Kat, and she is right. Is it a Walden Two experiment? Certainly it is not much like the experiment described in the book. The life portrayed in Walden Two was the goal of Twin Oaks, but it was not approached through the application of scientific principles. (Kinkade 1974, x)

Such an unlikeness between Twin Oaks and Walden Two even becomes annoying to the residents there. “Walden Two, they felt, was too removed from what
they were actually doing and thus gave outsiders a wrong impression. Twin Oaks in particular tired of having visitors expecting (or fearing) to find Walden Two.” (Kuhlmann 2005, 121)

Skinner’s theory on behavior, which has been largely neglected by Twin Oaks, in contrast plays a fundamental role in another experimental community: Los Horcones.

6.2. Los Horcones

Los Horcones, which has the slogan “Social change is achieved when each individual makes a personal change” on its entrance, was set up formally in Northern Mexico in 1973. Like Twin Oaks, it has a predecessor as well. In 1971, a school-like institution named “Center for Children with Behavioral Deficits” had been founded by Mexican psychologist Juan Robinson and his wife, Mireya Bustamente, who had also been trained in psychology. They were inspired not only by Walden Two but also by all Skinner’s scientific work on behavior. Working with two teachers whom they had hired for the Center, they initiated and planned Los Horcones as a community to apply Skinner’s theories. In contrast to the funders of Twin Oaks, they did not read Walden Two as a manual, but only as one of many possibilities to realize Skinner’s thoughts. Such a starting point, which was different from that of Twin Oaks, led the community to a different development.

The most distinct practice of Los Horcones probably is its educational program. The children in Los Horcones receive something very similar to the “self-control training” program of Walden Two. When the teachers find them performing wrong behavior,

[t]hey do not try to uncover the reasons for the children’s unusual behavior patterns, instead providing an environment that is as simple, predictable, and unthreatening as possible while at the same time giving the autistic child assignments of a very basic nature. Once the assignment is fulfilled, the child is reinforced by praise, hugs, sweets, or whatever other reinforcers are available. If the child does not fulfill the assignment, he or she is not punished. The undesirable behavior is simply ignored. (Kuhlmann 2005, 144)

Apparently, the technique of positive reinforcement is adopted in Los Horcones. Moreover, “Children at Los Horcones are not ignorant subjects of subtle manipulation. They are taught exactly what the community’s objectives and meth-
ods are” (Kuhlmann 2005, 148). In other words, they are not merely the “victims” of the technique; they are taught with the theories underpinning how they have been trained. As described in one of Los Horcones’ newsletters Walhdos,

The children who live at Los Horcones study behaviorology (behavior analysis and behaviorism) since they are very young. This helps them to better understand how and to what extent their environment influences the way in which they and other people behave, and how they can modify their environment so that it influences their behavior in the desired direction. (Los Horcones 1988, 10)

The focus on the effect of environment also reflects the architectural design of Los Horcones. A lot of buildings there are designed for communal use, such as the dining hall, TV room, laundry, and yogurt and cheese factory. Through these shared spaces, not only the children, but also the adults in Los Horcones learn how to interact and live with one another within the community.

The political system of Los Horcones, called “Personocracy”, was developed and tested for the several years, and has been formally implemented since 1982. It is a decision-making mechanism by looking for agreement from residents in the community. But it is different from
democracy, which is government by the majority, and from totalitarianism which is government by few ... As it is described, personocracy maximizes the reinforcing consequences of pro-social behavior and is non-hierarchical and cooperative. (Rutherford 2009, 142)

One may raise a question about this system: If it is more open than that of Twin Oaks, why does it not suffer the problem of political apathy? The answer to the question lies in their emphasis on education as a part of personocracy. On the one hand, the coordinator of each issue has to be taught by the former coordinator with relevant knowledge and skill; on the other hand, the residents are already taught “to effectively participate in decision making to achieve to a common benefit” (as cited in Rutherford 2009, 143). As a result, the residents actively take part in decision-making processes, and consensus about communal affairs is usually easy to reach.

To be sure, the paternalistic culture in Mexico may influence the achievement of Los Horcones, especially in the dimension of politics (Kuhlmann 2005, 169–70). However, what is really crucial is that they embrace Skinner’s principles. “[F]irst, to guarantee the survival of the community by teaching its members the required behavior, and second, to constantly increase the amount of
Behavioral Modification in Small Scale

reinforcement available for all” (Los Horcones 1991, 255). Again, the main feature of Los Horcones that differentiates it from Twin Oaks is its sticking to what is fundamental to Walden Two, as confirmed clearly by one of its founders:

I am very committed to a science of behavior. I believe that only by applying this science can we design a better community ... A science of behavior, or perhaps better stated as an appropriate application of this science, can enormously contribute to our shaping of an ideal society and solving global problems. (Robinson 1996, 152)

This quotation reveals an important aspect of Skinner’s thought: the interrelationship between science and technology. Science, for Skinner, is not about stand-by observation of natural phenomena or processes; rather, it concerns hands-on experimentation, by which prediction and control on experimental subjects can be made more effectively and precisely. This is a Baconian view of science, in which “the search for causes, although sometimes aided by naturalistic observation, is best pursued through the experimental method—that is, the production and reproduction of effects” (Smith 1992, 218). Skinner himself also admitted in his autobiography that he was deeply influenced by Bacon’s work at his young age (Skinner 1976b) and Walden Two was highly inspired by and a respect for Bacon’s utopian book The New Atlantis (Skinner 1983). In other words, “a science of behavior” and “an application of this science” cannot be separated from one another in Skinner’s thoughts: the science of behavior and the technology of behavior are the two sides of the same coin, and this coin can be invaluable for a better future.

By taking the coin, Los Horcones has been and still is vigorous and growing. The members keep publishing scientific journal papers, in the name of the community, to share their experience in behavioral modification with the world. They also have been invited to join or give talks at psychology conferences. However, curiously enough, Skinner has never paid any attention to this sticky-to-behaviorism community. We are not even sure whether Skinner knew of its existence. Los Horcones, contrary to Twin Oaks, probably can be seen as a real Skinnerian community24 because it is “not a community based on Skinner’s

24 Interestingly, a self-introduction video made in 2008 by Los Horcones begins with the scenes of two animals: a pigeon (the symbol of Skinner’s work) and sheep (the livestock of Walden Two). See: https://www.youtube.com/watch?v=hVRUtVEGE1M (accessed May 18, 2018).
novel ... but on the science on which that novel is based” (Los Horcones 1986, 129).

6.3. Community as a Unit for Behavior Modification

It is no accident that Skinner chose a community rather than a city or a society as the unit to create his behavioral utopia. As he claims that humans are largely shaped by social environment (even though it can be traced back to physical environment), Skinner was also influenced by the general rejection of megacities in the first half of the 20th century. It is highly possible that the form of Walden Two mirrors the work of architect Frank Lloyd Wright (1869-1959), because during the gestation of the novel he was “[o]ne of the immediately resources for architectural models for a community planner and social utopia in the U.S. Midwest in the 1930s” (Breshears and Devonis 2015, 60).

Wright was a modernist who insisted on the idea that “forms follow functions” and regarded the growth in the size of city as a major cause of many of the social ills in America. He believed that small communities were more suitable for people living together and interacting with each other. To realize his idea, Wright designed a small estate in Wisconsin, named “Taliesin”, where he and his colleagues lived and worked together for over 30 years. Many open, small spaces in Taliesin were designed intentionally for both regular and unexpected exchanges between residents, the curving walkways for slowing down residents' pace and making them “feel” surroundings, and the specialized rooms for various communal activities. All these physical settings can find their counterparts in Skinner’s Walden Two:

We filed up a narrow staircase and found ourselves at one end of a broad corridor called the “Walk”. This ran the full length of the building, curving slightly as the building followed the contour of the hill. The sun had not yet set and the evening promised to be fine, but there was a good deal going on indoors. The Walk was dotted with strollers, who seemed to be there for the sake of greeting others like themselves or to settle their dinners ... As we joined in this procession Frazier called our attention to various common rooms, arranged on either side of the corridor. On our right were reading rooms, libraries, and small lounges with chairs and tables grouped for conversation or games. These rooms looked out upon

---

For the emergence of the atmosphere against mega city and typical practices in architecture, see Fishman (1982).
Behavioral Modification in Small Scale

the Walden Two landscape from which we had seen the building during the afternoon. They were all occupied. (Skinner 1976a/1948, 34)

Skinner never mentioned this background against which the size of Walden Two were conceived, but he indeed talked about the disadvantage of being big in the preface of the second edition of *Walden Two*:

> The bigness of a large city is troublesome precisely because we meet so many people whom we shall never see again and whose commendation or censure is therefore meaningless. The problem cannot really be solved by delegating censure to a police force and the law courts. (Skinner 1976a/1948, xi)

For Skinner, the people with whom we interact in our daily lives are *potentially* reinforcing to us because their responses (even just a smile) are the immediate consequences of our behavior. However, they can be really reinforcing *only if* we interact with them repeatedly and frequently. When the consequences of our behavior are absent, unobvious, or unsteady to some degree, we will hardly be shaped. This is why lying to a person that we will not meet again is much easier than lying to a close friend or family members. Unfortunately, in a city, most people are “not-meet-again” persons to us. In contrast, in a small village or town, we can easily predict people’s responses to our behavior according to our experience in interacting with them in the past. As a result, a moral solidarity between all communal members can be easily built. For Skinner,

> [a]lthough a small community does not bring out “human nature in all its essential goodness” (small towns have never supported that romantic dream), it makes it possible to arrange more effective “contingencies of reinforcement” according to the principles of an applied behavior analysis. (Skinner 1976a/1948, x)

The “contingencies of reinforcement” in the quotation can be made social as well as physical. In a city there are too many roads to set up speed bumps, and we can only make some of them reinforce the behavior of driving slowly. While drivers avoid the negative consequence by driving slowly on these roads, there are far more roads on which driving slowly does not lead to positive consequences. As a result, the drivers in such a city are scarcely reinforced to drive slowly by the roads-with-speed bump, and the morality of driving slowly can hardly emerge and become settled.

This is why Skinner prefers a community, instead of a city or society, as the unit for behavioral engineering. The best thing about a community is its size;
nothing else that makes it a community. Skinner stressed the importance of size again in a follow-up to *Walden Two*, which was published in 1985 and titled “News from Nowhere, 1984”. In the story, a new visitor named “Blair” (by the name, Skinner referred intentionally to George Orwell, the author of *1984*) discusses with Frazier about why Walden Two can achieve its goal. According to Frazier, the small size of Walden Two is no less important than the technology of behavior underpinning it:

> Communities ... have always been *multum in parvo* ... They are *miniature* states. They must be small if they are to be experimental ... The trick is to *stay* small. Walden Two works because it is small. Cities need police forces just because they are big, because face-to-face control of decent personal behavior is impossible. (Skinner 1985, 10–11, italics in original)

Note the sentence “They must be small if they are to be experimental”; this will lead us to the second reason for staying small.

The only thing that Walden Two needs, accordingly to Frazier, is experimentation. This is the most distinctive feature of Walden Two, and what makes the novel special. “Where *Walden Two* departs from earlier works of utopian fiction is in Skinner’s focus on experimentation—not being ‘married’ to an idea or ‘way of doing’” (Frank 2017, 7). In other words, for Skinner, applying the technology of behavior to change human behavior is not different from experimenting. However, as a scientist who had been too familiar with the techniques of doing experiments, Skinner did not talk much about how experiments can be conducted well. Fortunately, most of us today seemingly have basic knowledge about it.

In general, the first step to conduct an experiment, be it a physical, biological, or psychological one, is to reduce the number of subjects we are going to study and limit the factors that are not relevant but may influence the results of the experiment. If there are too many subjects in an experiment, it can become cumbersome to carry out, not only because time and money are limited, but also because the impact of the interactions between the subjects may be entirely unknown to us. Similarly, if there are too many factors in play, the effort on conducting an experiment may turn out to be in vain because the causal relations between inputs and outputs are likely to become unidentifiable. In short, “simplification” is the nature of experimentation, and staying small probably is the best policy for it. With regard to behavioral modification, simplification means
that we should test a behavioral design, be it an artifact, a law, or a new way to praise, in a small community.

There is another advantage of staying small for behavioral modification. Applying a new design directly to most people without knowing its pros and cons is risky. It can be too late when harmful side effects appear, not only because most people have been harmed, but also because retracing it may become very hard. Such a risk can be avoided if we experiment with a small community first. Besides, even if no undesirable effects follow, it may be a waste of time and money once we find that the new design does not work well. Staying small, again, helps to avoid such a situation. Accordingly, it is no accident that Skinner insisted that behavioral engineering has to begin with a small number of people and put size and experimentation in the same context:

Frazier’s critics will protest. What can we conclude from a successful community of a thousand people? ... Frazier might answer by calling Walden Two a pilot experiment. Industries do not invest in large plants until they have tried a new process on a smaller scale. If we want to find out how people can live together without quarreling, can produce the goods they need without working too hard, or can raise and educate their children more efficiently, let us start with units of manageable size before moving on to larger problems. (Skinner 1976a/1948, viii-ix)

6.4. Behavioral Modification as Piecemeal Engineering

Starting with small-scale experiments on human behavior, as we have already seen, is an essential requirement for Skinner to apply the technology of behavior. Such an approach to social reform resonates the idea of “piecemeal engineering”, a term coined by the philosopher Karl Popper (2013a/1945). While Skinner can provide us with practical reasons for staying small, Popper may offer theoretical support to the approach. Moreover, as we shall see, linking Skinner’s technology of behavior to Popper’s piecemeal engineering helps to answer the questions that have been left in Chapter 5.

In his two-volume book, *The Open Society and Its Enemies*, published first in 1945, Popper proposed “piecemeal engineering” as a strategy for social reform, in opposition to what he called “utopian engineering”. Popper (2013a/1945) argued that the utopian approach to social reform is based on two problematic doctrines concerning how human society should be developed. The first is historicism, and the other holism. Historicism, according to Popper, consists of the belief that there is a perfect form for society (Plato, for example) and the as-
sumption that there is a law underlying human history (Marx, for example). Both of them can and should be discovered as long as humans can penetrate, by virtue of rationality, what is in front of their eyes. Once the form and the law become clear, human society will have a faultless roadmap to follow and an absolute destination to head for. As the historicists tend to see human society as a whole, in which every part is connected to one another, they are in line with the idea of holism and unavoidably in favor of an overall reconstruction for human society. Popper seriously warned that such a utopia-orientated method for social reform is bound to be violent and oppressive because individuals are inescapably forced to sacrifice for a highest purpose. In short, utopian engineering, for Popper, may be a non-stop train transporting humans toward a totalitarian society.

In contrast to utopian engineering, piecemeal engineering is a slower but safer method. Popper made this point vivid in his latter book, *The Poverty of Historicism*:

The characteristic approach of the piecemeal engineer is this. Even though he may perhaps cherish some ideals which concern society ‘as a whole’—its general welfare, perhaps—he does not believe in the method of re-designing it as a whole. Whatever his ends, he tries to achieve them by small adjustments and readjustments which can be continually improved upon ... Accordingly, he will make his way, step by step, carefully comparing the results achieved, and always on the lookout for the unavoidable unwanted consequences of any reform; and he will avoid undertaking reforms of a complexity and scope which make it impossible for him to disentangle causes and effects, and to know what he is really doing. (Popper 2013b/1957, 6)

For Popper, this is a much more moderate approach to social reform and, from the perspective of politics, more compatible with the notion of democracy. Although there is no term referring to experimentation in the quotation above, it is not hard to tell that the major task assigned by Popper to the piecemeal engineers is to conduct experiments on social reform.

The approach of piecemeal engineering can be interpreted in two ways. On the one hand, it is about the number of variables we would like to change. If we change too many things at the same time, “it may become extremely difficult to see which changes are bringing about which result, and, thus we may lose track of the cause-effect chain” (Irzik 1985, 2). On the other hand, it is about the scale on which we conduct experiment. If we directly go for large-scale adjustments, we are running the risk of unexpected and irreversible side effects. In contrast, “since small adjustments will most likely produce small unwanted conse-
Behavioral Modification in Small Scale

quences, it is easier to back and correct our mistakes” (Irzik 1985, 2). These two interpretations, apparently, correspond to the basic principles of experimentation.

The resemblance between the Popper’s approach and Skinner’s to social reform becomes obvious. It is not possible to remove the practice of experimentation from Popper’s piecemeal engineering; nor is it possible to take away the principle of being piecemeal from Skinner’s behavioral engineering. As Skinner himself stated: “Perhaps we cannot now design a successful culture as a whole, but we can design better practices in a piecemeal fashion” (Skinner 1971, 153). A review paper on Skinner, published 40 years after Beyond Freedom and Dignity, also rediscovered the piecemeal fashion in Skinner’s thought and attempted to draw psychologists’ attention to it (Leigland 2011). Popper, though opposing Marxism, did not really take a hostile attitude toward early Soviet Russia. The reason, perhaps surprisingly, is that he regarded their experiments, at least the early ones, as examples of piecemeal engineering rather than the utopian kind (Hacohen 2002, 396–97).

For Popper, the utopian approach to social reform not only entails the danger of totalitarianism, but also has an epistemological limitation that makes it highly implausible. The root of this limitation can be found in Popper’s philosophy of science, that is, falsificationism. According to Popper (2002b/1959), we can never be sure whether or not a scientific theory is true (in the sense of ultimately correct) by looking for evidence that supports the theory. What we can only know, instead, is whether a scientific theory is untrue because any evidence that is inconsistent with the predictions made through the theory can disprove it. In other words, a scientific theory can never be verified; rather, it can be falsified. Any scientific theory, therefore, can be false or temporarily correct; it might be true, but we can never know. Given that, how can science progress if the certainty is not on the side of truth? We do make science better and better, as Popper (2002a/1963) has argued, not because we have found untimely correct theories, but because we keep cutting off false theories and putting new ones into the trial of falsification. The growth of scientific knowledge, therefore, depends on removing what is false rather than adding what seems correct.

Such falsificationism, for Popper, does not only work for science but also for human society. Just as he denied the possibility that we can make sure of true scientific theories, Popper also denied the possibility that we can find a perfect form of society. Historicism claims about the law of history are unscientific because they cannot be falsified. If the progress of science is non-directional, so is
the progress of society. There is no such thing as destiny for society to proceed. This is why he named his book *The Open Society and Its Enemies*. Society is to be open by its nature because it is not possible to know its direction beforehand, and it *ought* to be open because any direction to which we guide everyone may turn out to be wrong (Popper 2013a/1945).

Piecemeal engineering, therefore, is more about removing mistakes than creating goods: “The piecemeal engineer will, accordingly, adopt the method of searching for, and fighting against, the greatest and most urgent evils of society, rather than searching for, and fighting for, its greatest ultimate good” (Popper 2013a/1945, 148). This is the only reasonable way to develop our society, according to Popper, since piecemeal engineering is an “introduction of scientific method into politics” (Popper 2013a/1945, 153). Just as we cannot know which theory is true but can identify which theory is false, we cannot know what is good but can identify what is bad for society.

Popper’s epistemological arguments for piecemeal engineering, then, help us to answer the questions left in the last chapter. Since in Skinner’s approach moral values can only emerge and become solid *after* the behavior related to it has been practiced widely and long enough due to the survival of the culture carrying the behavior, we have difficulty in justifying the design and implantation of green behavior-steering technologies by saying that being pro-environmental is good and should be the direction into which everyone should be steered. Such a difficulty, from the perspective of Popper, is a sort of misplacement. The main task for social reform, again, is to remove what has been found problematic rather than moving everyone toward something that is supposedly good. In other words, we do not need to know, and we cannot actually know, whether being pro-environmental is good, nor do we need to task ourselves with designing and implementing behavior-steering technology for it. Rather, we identify anti-environmental behavior and work on removing it from our daily lives. The design and implementation of green behavior-steering technologies, therefore, is for steering people *away from* being anti-environmental, *not into* being pro-environmental.

By connecting to Popper, we can also solve the problem about human freedom under aversive control. While removing anti-environmental behavior is more urgent and relevant in tackling the environmental crisis, the technique of positive reinforcement becomes inapplicable and therefore may trigger people’s concern about freedom. However, according to Popper, the direction for people to go is still open. This gives us an opportunity to make positive reinforcers as a
complement to our aversive behavior-steering technology. For example, while we design ways to remove the behavior of driving by increasing the inconvenience, we do not limit people to any of following options: biking, walking, taking public transportation, and staying at home. Moreover, we should make these options enjoyable, pleasant, or satisfying. By so doing, the final consequences can still be positive, and thereby all the contingencies of reinforcement are still positively reinforcing. Although positive reinforcement is not the only technique used to modify people’s environmental behavior, the overall effect is not far from being the only positive reinforcement in operation. The problem about human freedom under aversive control, accordingly, can be largely reduced.

The connection between Skinner and Popper seems odd at first glance, especially because Popper has been famous for his stance against utopian engineering. However, the distance between them is not as far as we thought. On the surface, Skinner’s proposal is utopian, but in fact it is quite open due to its experimental nature. Insisting on the policy of trial-and-error, Skinner neither rejects changes and corrections, nor sets up a destination for society to reach. For Skinner, utopia is about experimentation rather than perfection:

It is not surprising that, so far as the real world is concerned, the word utopian means unworkable. History seems to offer support; various utopian designs have been proposed for nearly twenty-five years, and most attempts to set them up have been ignominious failures … If planned economies, benevolent dictatorships, perfectionistic societies, and other utopian ventures have failed, we must remember that unplanned, undicted, and unperfected cultures have failed too. A failure is not always a mistake; it may simply be the best one can do under the circumstances. The real mistake is to stop trying. (Skinner 1971, 153)

While Skinner’s behavioral engineering is not so utopian, Popper’s piecemeal engineering is not so anti-utopian. If utopia refers to a place in which some basic principles govern its development, a continually piecemeal-engineered society can be seen as utopia as well. Just like Skinner proposes an application of his science of behavior as the principle of behavioral modification, Popper also suggests an application of his philosophy of science as a principle for social reform. In other words, Skinner and Popper are not so contradictory to each other in how to (re-)shape the future of humanity: both are committed to experimentation, openness, and progress.

Combining the approaches of Skinner and Popper leads us to a practicable and achievable program to take advantage of the power of behavior-steering
Design for Green

technology. We can call it “piecemeal-behavioral engineering”. The hyphen here means the equal position between the two adjectives. The term therefore can be regarded as “piecemeal” behavioral engineering, referring to behavioral modification in a piecemeal fashion. It can also be understood as “behavioral” piecemeal engineering, indicating carrying out social reform through behavioral change. For both, the locus is still human behavior, where behavior-steering technology helps most.

6.5. Concluding Remarks

While social reform has often been made through political transformation, the comparison between two real-world communities, Twin Oaks and Los Horcones, shows the crucial importance of behavioral modification in bringing about changes. This is where Skinner’s behavioral engineering helps; central to it is to regard modifying behavior as the result of experiments. Size matters, therefore; too big can be unmanageable, while too small may be meaningless. A community of around one thousand people is a suitable size for testing and identifying which design works on the one hand, but also helps to decrease the risk of making mistakes and avoid irreversibility on the other hand. Such a feature echoes the idea of piecemeal engineering, which Popper defends as a scientific and moderate way to reform human society. Against utopian engineering that attempts with large scale changes, piecemeal engineering aims to deal with social issues in a small scale. Moreover, due to the theory of falsification, targeting and fixing problems is not only more epistemologically possible, but also more politically and ethically feasible than creating a perfect social structure. The spirit of experimentation makes a combination of Skinner and Popper, the approach of piecemeal-behavioral engineering, possible, whereby social reform can proceed through behavioral modification in a piecemeal fashion. In this approach, clinging to the question of how to justify “being pro-environmental” is a misplacement, and the goal of behavior-steering technology is more about how to prevent people from “being anti-environmental”. In addition, while anti-environmental behavior is what needs to be removed, it does not mean the conformance of the public; rather, it gives an opportunity to complement behavioral removal with positively reinforcing designs, by which the whole procedure can still count as positive reinforcement and the problem of freedom tends to not appear.
7 Design Recommendations and an Application: Village Homes

Behavior-steering technology can be very helpful for tackling the environmental crisis. However, due to its power, the worries concerning human freedom tend to arise, which may seriously restrict the opportunity to take advantage of it. Taking a piecemeal-behavioral engineering approach to the design and implementation of behavior-steering technology can help to reduce such worries and leave space for making good use of it. To further elaborate the approach I have developed, I will first formulate a list of design recommendations, which embodies the points that have been mentioned, discussed, or argued in previous chapters. I will then show the feasibility and achievability of these recommendations by referring to a famous green community in American, named Village Homes. By going into the details of its design, we shall see what role an MBST can play in greenizing a community, how it works, and why it is crucial.

7.1. Design Recommendations

For designers who would like to make contributions to sustainability, the approach of behavior-steering technology is worth considering. However, such an approach tends to provoke ethical concerns about user freedom. The good news is that these worries can be eased or avoided. The list of design recommendations below may help designers to employ the power of behavior-steering technologies without causing anxiety or receiving critiques.

1. **Design not only nature-friendly products but also behavior-steering artifacts**

   Most products are made for human use, which means how people live is a more fundamental root of the environmental crisis. As pointed out in Chapter 2, “greenizing” products is far from enough; we need to greenize people as well. By changing people’s environmental behavior, a sustainable way of life can be made possible. Nature-friendly design and behavior-steering design are not mutually exclusive; rather, they should complement one another. Nothing could be better than a product that achieves the two goals at the same time.
2. **Taking care of the behavioral influence of non-behavior-steering technology**
   If the product-in-design only aims to be nature-friendly, its impact on users’ environmental behavior should be taken into account. Will it cause anti-environmental behavior? Will it counteract the effects of other behavior-steering technology? Among other problems, the rebound effect is probably the most noticeable one. A so-called energy-saving appliance may give its users a false impression that they are saving energy by using it and therefore can lead to a zero or negative result. A nature-friendly device does not need to improve users’ environmental behavior; however, it should not worsen it.

3. **Material behavior-steering technology first, then informational technology**
   After the behavior to be changed has been targeted, consider MBST first. Due to differences in mechanism, as discussed in Chapter 3, MBST is usually more effective than IBST, both qualitatively and quantitatively. In other words, when thinking about how to change users’ behavior, try taking an extrasomatic route first. While personalized IBST may work, it tends to create the problems of privacy and transparency, so it should be the last option. Simply put, the more concrete, the better.

4. **Think of low-tech rather than high-tech designs**
   Echoing Recommendation 3, “mundane artifacts” in people’s daily lives, such as chairs, houses, roads, and the like, are concrete and solid, and have no less impact on what people do than electronic devices or advanced technologies. But as such technologies exert their influences silently, we tend to pay little attention to them. While the notion of innovation is often associated with high-tech, we might already forget that a creative redesign of existing technologies can be counted as innovation as well. Moreover, as pointed out by countless studies on technology, new technologies often bring unexpected side effects and may even put people at risk. The environmental crisis is definitely a consequence of this kind. Prioritizing low-tech designs helps to avoid such a problem. Again, this is where MBST prevails.

5. **Community as a suitable size**
   Due to the environmental problems, designing for separate individuals may have little or no effect. But the number of the people to be steered cannot be too big. As shown in Chapter 4 and Chapter 6, considering
Design Recommendations and an Application: Village Homes

MBST's physical constraints and the experimental nature of behavioral engineering, a community with about one thousand people is probably the most adequate. Designing and implementing behavior-steering technologies in a piecemeal fashion not only helps to ensure that a design works, but also reduces the risk of unexpected negative results.

6. Keep experimenting

Experimentation is essential to any attempt for social reform, as argued in Chapter 5 and Chapter 6. Applying a new design to modify people’s behavior should be an open process that allows further adjustments or cancellation. Fixing or stopping what has been going wrong is easy, thanks to Recommendation 5. Only when the effect of a behavior-steering technology has been confirmed as positive, can it be applied to a larger population. However, such an application still needs to follow the spirit of piecemeal engineering, which means that the growth of targeted population should be gradual rather than rapid.

7. Aiming at anti-environmental behavior but complementing with positive reinforcers

Due to epistemological limitations, we cannot be sure that being pro-environmental is ultimately good. But to identify which behaviors are problematic and should be reduced is easier. Accordingly, we need to design to remove, or at least decrease, such behavior rather than to create pro-environmental one. This is the main task of piecemeal-behavioral engineers. Furthermore, to reduce the worry about people’s freedom, we can make other options enjoyable, pleasant, or satisfying. In doing so, the whole contingency of reinforcement can be positive, and the avoidance of aversive control becomes possible.

This list, containing the seven recommendations above, is a synthesis of the arguments made in previous chapters. It can, hopefully, help with the design and implementation of green behavior-steering technologies. Moreover, given that the environmental crisis is the hardest issue to which behavior-steering technology applies, the list developed to deal with it should also be able to be used for easier issues. In other words, some of the recommendations above are applicable to behavior-steering technology in general.
To show the usefulness of the list, I will introduce a real-world community named “Village Homes” and explain its success with each recommendation in the list. To be sure, this community was constructed long before I started the present study, and it is by no means possible that they were inspired by any of these design recommendations. However, this does not mean that the work they have done cannot reveal significant principles that one may discover some years later. It might also be possible that the designers were aware of the meanings of their design but did not express them in a theoretical, systematic way.

7.2. Village Homes: An Overview

Village Homes (Figure 8), a community built in 1982 and located near Davis, California, has been widely recognized as the greenest community in the US; it has become an exemplar for architecture and landscape designs (Francis 2002). Some of its design can be identified as green in a traditional sense; namely, they are made for using less energy or benefiting plants. In contrast, some other designs are intentionally designed to alter the residents’ behavior, and as I will argue later, one of them is fundamental and crucial to the whole community as a complex of all the green designs.

Figure 8. Aerial view of Village Homes in the mid-1980s (Francis 2002, 26)

Back in the early 1970s, Michael and Judy Corbett, the designers of Village Homes, were studying at UC Davis (University of California, Davis), a university
famous for its researches and courses concerning environmental sustainability. They were affected by the atmosphere of UC Davis and then had a plan to establish “a better place to live” (Corbett 1981). They decided to realize their idea on the scale of a community containing about 700 residents (244 housing units).

Since 1995, every new resident joining the communal life of Village Homes would receive a copy of *The New Homeowner Guide*, in which we can find the most important designs for the community:

A number of design features help Village Homes residents live in an energy-efficient and aesthetically pleasing community. All streets are oriented east-west and all lots are oriented north-south. The orientation helps the houses with passive solar designs and makes full use of the sun's energy. Street widths are all narrow with curving cul-de-sacs less than 25 feet wide minimizing the amount of pavement exposed to the sun in the long, hot summers. The curving lines of the roads also give them the look of village lanes, and the few cars that venture into the cul-de-sacs usually travel slowly. The common areas also contain Village Homes’ innovative natural drainage system, a network of creek beds, swales, and pond areas that allow rainwater to be absorbed into the ground rather than carried away through storm drains. Besides helping to store moisture in the soil, this system provides a visually interesting backdrop for landscape design. (Village Homeowners Association 1995, 1)

Reducing the usage of energy was the top goal for the designers. Davis has a Mediterranean climate with very warm days and cool nights, and the temperature outside often goes over 35 degrees Celsius in summers. The challenge, therefore, is to keep indoor temperature comfortable without using air-conditioners. Most green designs mentioned in the quotation above are sophisticatedly constructed to respond to this challenge. Even the natural drainage system has a positive effect on temperature reduction. As the soil has abundant moisture, the temperature of the land in summers can still remain comfortable because the heat can be carried away by moisture evaporation. This is why cities are often much warmer than their outskirts. Most cities are mainly made with concrete, a very impermeable material, so the heat stays and circulates in the city and therefore produces the phenomena called “heat island effect”.

What is crucial, for the whole community, is its traffic design. It aims directly at interfering with the behavior of moving and commuting. In order to promote green transportation, the designers created the walk-bike paths connecting every household and communal area. These paths are not alongside the street; instead,
they are located behind the houses in between every two rows of houses (as are the common areas; see Figure 9). If we compare Village Homes with its neighboring communities, we will find that the number and density of walk-bike paths in Village Homes are higher. The residents are encouraged to use their bikes and feet rather than cars. Moreover, the streets in Village Homes here look different from the regular streets in America. The streets are narrow, curved, and the end closed. The general width of streets in Village Homes are only 7 meters, about half the national average. Moreover, as these streets are curvy, it usually takes longer for the residents to enter the main roads connected to the city center of Davis if they choose to drive. To be sure, there are also cul-de-sacs in the nearby communities, but they are neither narrow nor curvy.

These narrow-curvy-cul-de-sacs in Village Homes are intentionally designed to inconvenience car-driving. Driving on them requires carefulness and slowness, causing boredom. As a result, the usage of cars and the consumption of gasoline in Village Homes are significantly lower than those in other communities (Table 2). Moreover, the side of the walk-bike paths becomes the front of the houses because most residents use the walk-bike paths much more often and tend to regard the streets only as their back alleys (Corbett & Corbett, 1999, p.37). This design works not only for residents but also for outsiders. As the other communities have no such design, outsiders tend to drive through them because those

Figure 9. Traffic Arrangement of Village Homes (Francis 2003, x)
ordinary streets are in fact shortcuts to the main roads. But this rarely happens in Village Homes.

Table 2. Car usage and gasoline consumption in Village Homes and other communities (Data source: Lenz 1990)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Conventional Neighborhood</th>
<th>Village Homes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average annual miles per car</td>
<td>13,400</td>
<td>11,300</td>
</tr>
<tr>
<td>Average gas mileage of vehicles</td>
<td>23.5 mpg</td>
<td>27 mpg</td>
</tr>
<tr>
<td>Gasoline consumption per car per year</td>
<td>577 gallons</td>
<td>422 gallons</td>
</tr>
<tr>
<td>Gasoline consumption per household per year</td>
<td>1,171 gallons</td>
<td>753 gallons</td>
</tr>
</tbody>
</table>

Such a traffic design, however, is unwelcome because the idea underlying it is in opposition to so-called New Urbanism. New-urbanists also think that traffic jams are a serious problem for American society, but the solution they propose is a much larger, well-connected network that can smooth the traffic flow (Southworth and Ben-Joseph 2004). They rarely take a critical attitude toward the importance of cars to American life, so their designs tend to remain car-centered. The view taken by Village Homes’ designers is entirely contrary: The traffic problems result from too many cars, not too few roads. Therefore, making people drive less is more necessary than making them drive comfortable. This idea is proven to be right by the case of Village Homes:

The design of this highly successful community breaks many of the rules popularized by the proponents of New Urbanism. First of all, it eschews the grid and provides access to houses via long, narrow cul-de-sacs—those “lollipops” of 1950s suburbia much hated by proponents of New Urbanism. The green-shaded, narrow, dead-end streets save money on infrastructure, use less land, reduce urban runoff, keep the neighborhood cooler in summer, and create a quiet and safe public area where neighbors meet and children play. (Cooper Marcus 2000, 128)

7.3. A Nature-friendly as Well as Behavior-Steering Community

The achievement of Village Home is not only based on its nature-friendly designs, but also by those behavior-steering arrangements (Recommendation 1). In
fact, the change of residents’ behavior is fundamental to its success, probably much more than those nature-friendly designs. If the behavior of driving were not targeted, the whole story would be entirely different.

Imagine that the streets in Village Homes are like ordinary ones, not in the shape of narrow-curvy-cul-de-sac. What would happen? Nothing! The residents drive frequently as usual, doing what people do in other communities. They are unlikely to start walking or biking even though there are bike-walk paths just behind their houses. For them, driving is still more convenient than walking or biking, so there is no reason to give up what they are already accustomed to. As the residents rarely walk or bike, they seldom meet one another and stop to chat at the communal areas in the community. Consequently, these communal areas may never become meaningful places to them, so there is no need to take care of the plants there. The green landscape eventually disappears. Moreover, outsiders are happy to drive through the community because they find these streets can be shortcuts to the main roads. Cars, not only residents’ but also outsiders’, come and go in the community, emitting a great amount of CO₂ and producing a local greenhouse effect. As a result, the temperature in the community stays high, and the residents have to turn on their air conditioners if they choose to stay at home. They may decide to go to other cooler places such as a department store, but again, they need their cars. All in all, the energy consumption of the community would stay at a high level, no different from other communities. In other words, Village Homes cannot be green without its residents driving less. If there were no narrow-curvy-cul-de-sacs, the community would not succeed, no matter how nature-friendly the other designs are.

Of course, the designers could choose a softer way to reduce the frequency of car-driving, but if they did so, the behavior of driving might not change much (Recommendation 3). They could, for example, persuade the residents by stating the fact that driving emits a lot of CO₂ or putting the slogan “Driving harms the Earth” everywhere in the community. They could also make persuasive technology to influence the residents, for instance, a warning device for high amounts of CO₂ emission, and install it into every resident’s car. Fortunately, they did not take any actions. As we already know, IBST is less effective, qualitatively and quantitatively, than MBST in changing people’s behavior. If what were applied are IBSTs, the community may still fail to be green because the resident’s driving behavior would not change. Taking the streets to steer the residents is a good choice: narrow-curvy-cul-de-sacs are no doubt MBSTs.
Even though the design of narrow-curvy-cul-de-sacs is quite low-tech, more about redesigning than inventing, it has no less effect than other high-tech designs made for the same purpose (Recommendation 4). For example, there is already a technology called “Actibump” in use. It is a dynamic speed management system; when it detects an oncoming car that is faster than the speed limit, “a hatch, integrated into the road surface, is lowered a few centimeters creating an inverted speed hump giving a physical reminder to the speeding driver” (Edeva AB 2014, 2). As its name implies, it is an “active speed bump”, which can make high speed driving uncomfortable and thereby may indirectly decrease the likelihood of poor driving behavior. No doubt, we can use it for a “Village Home Two”, but apparently it will cost much more than the streets in Village Home. In this sense, the design of narrow-curvy-cul-de-sacs is no less innovative than Actibump, and therefore worth being called an innovation.

Some nature-friendly designs in Village Homes might have a negative impact on the residents’ behavior, but the designers did not overlook them (Recommendation 2). For example, the location of the natural drainage channels in fact was carefully chosen. The designers could put them just alongside the streets as usual, but in so doing, they might lose the chance to decrease the likelihood of driving. Why? One reason for the residents not driving is that when they walk or bike on the walk-bike paths they can enjoy the scenes of every house’s backyard and common area, on which trees and flowers are flourishing. Such scenes are made possible by the drainage channels alongside the paths, by which the soil can be full of moisture and therefore fertile for plants to grow. If the drainage system were not designed in this way, the beautiful scenes might never be there and therefore the residents’ willing of choosing walking or biking would be largely reduced.

The most amazing fact about the narrow-curvy-cul-de-sacs is that such a design is exactly what makes all the other designs work well. Imagine a community built in the same climate condition and with the same goal. But its designer just decided to make its streets identical to those in other communities: wide, straight, and not dead-ended. Will this community be as green as Village Homes? Unlikely! Because the streets are wider, the community would have less permeable land to contain water, and the trees alongside the streets may not be big enough to shade the streets. Consequently, the temperature around the

Design for Green

streets would be high. Besides, the rainwater that cannot penetrate the streets’
surface can be too much for natural drainage system to handle, so the designers
would have to adopt the traditional sewerage system, which would further de-
crease the water content of the land. Moreover, the side-by-side houses alongside
the straight streets would become barriers for air circulation and exchange, and
such arrangement may put one house’s solar panels under the shadow of an-
other house’s roof. Eventually, the heat island effect would emerge in the com-
munity, and the attempt to be green would end up with high energy
consumption.

The key to the achievement of Village Home is its attempt to change the
residents’ behavior. And the method it chose is low-tech MBST instead of IBST.
Moreover, the effect of the nature-friendly designs on behavior has been consid-
ered and avoided. What has been “greenized” in Village Homes are not only
things but also humans. Village Home, if viewed as a whole, is a beautiful de-
sign that reaches two goals: to be nature-friendly as well as behavior-steering, at
the same time.

7.4. Small-Scale Engineering on Driving Behavior

It is not clear whether the designers had the idea of piecemeal engineering in
mind when planning Village Homes, but the final product they created embod-
ies it (Recommendation 5). In fact, they consciously chose to design for a com-
unity rather than individuals or a huge population. For them, on the one hand,
a singular house seems hopeless to mitigate environmental problems, but on the
other hand, a city is too much over a “human-scale”27 (Corbett & Corbett, 1999).
Scaling their plan at a community level is not only important for the community
to be green but also what makes the plan come true.

In the beginning of implementing the plan, almost no one was willing to
support them. No bank in Davis was willing give them a loan; governmental de-
partments did not approve any page of their proposal. Village Homes was re-
garded as infeasible and even impossible because its design violated numerous
codes and conventional practice at that time. Although eventually they obtained
money from one bank and got approved by officials, we can easily imagine that if

---

27 Recall the trend against mega cities mentioned in Section 6.3. The Corbetts were part of this
trend and highly influenced by the architect and city planner Ebenezer Howard (Corbett &
Corbett, 1999).
the size of Village Homes were scaled larger, it would be much hard to get sup-
ported and much easier to get rejected. We may learn from this another advan-
tage of staying small: In practice, a small-scale project is more tolerable for
existing institutions and therefore more likely to be carried out. With regard to
the sense of community, one of the designers even wished the community to be
smaller than it was: “We would have had a much stronger sense of community if
there were about 500 of us” (Owens, 1993, p. 20).

There seems no practice related to Recommendation 6 because the designers
neither tested their designs before launching the project nor had ambitions to
expand the community. However, we may regard the sales of the houses in Vil-
lage Homes as a test to see whether such an experiment is successful or not. If
the houses are popular, then the experiment goes well, and vice versa. The fact
is, “properties at Village Homes command a substantially higher price, $10 to
$25 more per square foot, than those of surrounding subdivisions, and homes
sell faster when they come onto the market” (Staff et al. 1998, 12). Moreover, not
only has the price of the houses been climbing higher and higher, but the resi-
dents at Village Homes are also more satisfied and happier than those in adja-
cent communities (Lenz 1990). These are the indicators showing how popular
Village Homes is. Accordingly, the current version of the design is no doubt
successful. If someday the community is going to be expanded or a larger ver-
ion of it is going to be constructed, Recommendation 6 should be applied.

The design of narrow-curvy-cul-de-sacs nicely illustrates the approach of
piecemeal-behavioral engineering; it reduces the behavior of driving but has
been complemented with other options (Recommendation 7). As discussed in
the last section of Chapter 6, epistemologically it is not possible to know what is
ultimately good or right, but it is highly possible to identify what is problematic
and cannot continue, and the main purpose for designing and implementing
behavior-steering technology is to remove those mistakes. In the case of Village
Homes, the designers targeted driving as an anti-environmental behavior and
designed to reduce it. They did not define walking or biking or staying at home
as what the resident ought to do; instead, these behaviors are regarded as options
open to every resident. The design of narrow-curvy-cul-de-sacs works because the
consequences of driving on them are aversive. By not driving or driving less the
residents can avoid such consequences. However, if the other options are closed
to them or too awful to choose, there would be objections and protests, and the
price of the houses may decline eventually. Fortunately, by several nature-
friendly designs mentioned above, the designers have made these options en-
joyable, pleasant, and even attractive. Besides the beautiful views and the chance to chat with neighbors, sometime there are also fruits on or beside the walk-bike paths to pick. And because the temperature is usually comfortable both inside and outside the houses, thanks to the co-effect of narrow-curvy-cul-de-sacs and the other designs, staying at home is as favorable as walking or biking. As a result, the net consequences are positive, and the problem of losing the freedom to drive is unlikely to arise. Indeed, while there have been complaints, only few though, about the solar heating system, the unclear boundary between the private and communal areas, or the smell of open drainage channels, no one has complained about the narrow-curvy-cul-de-sacs (Owens 1993).

7.5. Concluding Remarks

For practicability, concepts, perspectives, and arguments have to be turned into design recommendations. To put these recommendations into one sentence: Design and experiment with green MBST at the level of community to reduce anti-environmental behavior. It reflects and represents the core principles of piecemeal-behavioral engineering. Moreover, the list of design recommendations is not only a practical suggestion but also a tool to analyze existing designs. The case of Village Homes illustrates this point well. Through the lens of piecemeal-behavioral engineering, Village Homes is a notable application of the approach, embodying all the design recommendations. Central to it is the design of narrow-curvy-cul-de-sacs, by which such a green community is made possible. By making the streets hard and boring to drive, the community successfully reduces the behavior of driving; however, there are designs to positively reinforce the behavior of walking and biking. As a result, the net consequences of the whole traffic design are positively reinforcing, and there are seldom, if any, complaints or worries about the loss of freedom from the residents. Moreover, these narrow-curvy-cul-de-sacs are what “activate” other green designs. Most of the nature-friendly designs in Village Homes would become redundant or useless without the presence of the current street design. This fact proves how fundamental and indispensable MBSTs are in changing human behavior and helping with the environmental crisis. The approach of piecemeal-behavioral engineering, accordingly, provides a theoretical basis to take advantage of behavior-steering technologies without inducing ethical concerns about human freedom.
In order to make good use of behavior-steering technology for tackling the environmental crisis, this research thus far has focused on constructing theoretical defenses and practical suggestions to ease related ethical concerns. This approach is fairly sufficient for launching projects or immediately taking action. However, this approach has its political implications. This chapter aims to take the whole discussion to the level of politics. First, I will point out that the challenge of the environmental crisis is politically relevant and requires a shift toward another political system. Our liberal democracy has difficulties in mobilizing collective changes in behavior, but things are unlikely to get better if the alternative is communitarian democracy. This is not because, as we shall see, we prioritize the individual over the collective, but because the image of humans we have been holding is problematic. I will argue, then, an alternative model of democracy that distances itself from this image is needed. I will finally suggest that one potential candidate for such a model is, “agonistic democracy”, an approach conceived by the political philosopher Chantal Mouffe. It will become clear that agonistic democracy leaves much room for the design and implementation of behavior-steering technologies and therefore has competence in dealing with environmental problems.

8.1. More than Liberal Democracy

Due to its features, such as large scale, uncertainty, dispersiveness, and unaccountability, the environmental crisis poses a serious challenge to us. Its root is our current way of life, and tackling it requires us to act in more or less the same way. In our liberal democracy, any external interference to individual freedom is not allowed unless the interference is already agreed by the individual, or the reasons for the interference can be justified. But for the environmental crisis, as pointed out earlier, both criteria are hard to meet. Neither has an agreement been achieved, nor are the reasons to be pro-environmental justifiable. The environmental crisis becomes a political morass that we cannot help but sink into.
The approach having been developed through previous chapters, hopefully, offers a way out. On the one hand, it reduces the worry about freedom by employing Skinner’s technique of positive reinforcement; on the other hand, it avoids the task of justification by referring to Popper’s idea of piecemeal engineering. This approach is a strategy to make technological intervention, the design and implementation of behavior-steering technologies, adoptable and acceptable. To be sure, it is by no means a perfect solution to the environmental crisis, but is probably the best we can do within the current political structure. Although it might be mild to such an extent, it is strong in the sense that it allows technological interventions to affect people’s behavior.

A more radical change needs to deal with the politics. Indeed, there is no other problem like the environmental crisis requiring a massive, joint effort to solve. The environmental crisis, therefore, points us to the deepest question concerning human society: How, after all, do humans live together? When each individual has his/her own goal, how is it possible for people to be united as a whole in working on the same thing? Can such individuals be, say, “synchronized” for something that is probably beneficial to the collective but perhaps detrimental to some of them? In other words, how do we deal with the tension between the individual and the collective? This is where the problem of freedom comes from and where our liberal democracy sometimes gets stuck.

We may need to go beyond liberal democracy if we would like to solve the problem more efficiently and substantially. No doubt, such a move needs a huge amount of work, and the first step we may take is to reflect on the basis of liberal democracy: individualism. Individualism has been taking the central position in political and social philosophy since the Age of Enlightenment. For individualists, the only reason that people are willing to live within a collective (a country, for example) is the protection and convenience it offers. The task of a government, therefore, is no more than ensuring protection and convenience for each person. In prioritizing the individual rather than the collective, individualism creates a barrier for unconnected persons to form a solidary collective. And such a barrier takes the shape of tension between the individual and the collective.

Although John Rawls’ idea of the “veil of ignorance” has been regarded as a successful invention to relieve the tension, there are still critics who are not satisfied with, or even doubtful about, his individualist stance. These critics reject individualism as the foundation of democracy; rather, they take a “communitarian” approach to democracy. To be sure, there are many versions of communitarian democracy, but they do share the same criticism of individualism (Bell
To highlight the defects of the individualist approach to democracy, I will take Michael Sandel (perhaps the most known critic of Rawls) as an example, showing his arguments against individualism below.

In reviewing Rawls’ *A Theory of Justice* (1999/1971), Sandel (1998) argues that Rawls’ theory cannot stand because its assumption is problematic—humans are not stand-alone and self-sufficient, and a person’s identity as well as personality are largely shaped by his relationship with the community in which s/he grows and lives. A collective cannot be formed through the individuals who can deliberate as if they are behind the veil of ignorance; it is rather the other way around, that is, being within a community is a precondition for being an individual. Only the life accordant with the values shared by the collective is worth living because we are however “as members of this family or community or nation or people, as bearers of this history, as sons or daughters of that revolution, as citizens of this republic” (Sandel 1998, 179). For Sandel, an “unencumbered self” does not exist and cannot be the starting point to form a collective. Such a communitarian view, therefore, prioritizes the collective rather than the individual in democracy.

For communitarians, what flourishes or maintains the community has higher, or at least the same, priority than individual interests. “For the liberal, justice and freedom are necessary for politics; for the communitarian, politics is what makes justice and freedom possible” (Katz 1997, 79). Accordingly, in contrast to a liberal government, the main task of a communitarian government is to ensure the common good of the community and to help its members to move in the right direction. In prioritizing the collective, communitarian democracy seemingly makes more room for external interventions and therefore has more chance to tackle the environmental crisis in rapid and dramatic ways.

However, appealing to communitarian democracy is tricky. While communitarian democracy highlights the common good, it leaves the content of common good mostly empty. The only value that has been predefined as the common good is the “community” per se. That is to say, communitarian democracy still needs a certain procedure to decide what else constitutes the common good. If the common good only means what most people think is good, then a communitarian democracy is likely to slip into a “tyranny of the majority.” To prevent such a slippery consequence, the strategy that communitarians often

---

28 Besides Sandel, Charles Taylor and Michael Walzer are also famous for their communitarian tendencies.
take is to encourage constant communications between the citizens, by which mutual understanding and tolerance can be increased and sustained. In other words,

[asserting the need to balance community and individual rights has the potential to limit the voice of the diverse or marginalized populations. Consequently, it requires a more demanding type of citizen discussion, one that does not encourage partisan groups to dominate an adversarial decision process but advocates for inclusion of all voices in a process of discernment and consensus building. (Heying 1999, 44)]

But in so doing, communitarian democracy becomes similar to its counterpart, and the difficulties we have encountered in liberal democracy will show up again. Although the individual is not placed prior to the collective in communitarian democracy, we still have trouble in justifying the value of being pro-environmental as a common good and thereby taking aggressive action for it. There is no doubt that the idea of piecemeal engineering can be applied here, but then what are the benefits in taking communitarian approach?

As liberal democracy gets stuck on the issue of the environmental crisis, so does communitarian democracy. This is probably why the historians of science Naomi Oreskes and Erik Conway (2014) regard the catastrophe of climate change as “the collapse of Western civilization”, and think that the only government that can make a great change for human revival is China’s. For them, democracies are hopeless at solving environmental problems, and the coming of quasi-autocracies is an unavoidable outcome of the environmental crisis. Shall we give up democracy in order to deal with environmental problems efficiently? If we would like to answer “no”, then what do we do to refine democracy?

What to give up is, as I will argue below, not democracy itself, but the agreement- or consensus-oriented version of it. Liberal and communitarian democracies (below I will call them “classical democracy”) fail not because they rely upon “demos”, but because the image of humans they have held is problematic.

8.2. Image of Humans

In classical democracy, as mentioned above, the individual is placed prior to the collective. It is unquestionable for classical democracy that the individual is and should be an autonomous being having “the capacity to be one’s own person, to live one’s life according to reasons and motives that are taken as one’s own and not the product of manipulative or distorting external forces” (Christman
If everyone is free in such a way, how is it possible for all to agree upon the same goal and therefore live together without any loss of freedom? The secret for classical democracy to ensure that the tension between the individual and the collective can be solved sooner or later is the connection between autonomy and self-rule.

The meaning of autonomy and that of self-rule are interrelated and sometimes exchangeable. To be a rule-imposed self requires the independence of one’s choice from another’s manipulation on the one hand and the capacity to rule oneself on the other hand (Arneson 1991). From a Kantian perspective, the absence of external influencers is not the key for a person to be autonomous; what makes a person autonomous is the capacity to make decisions as if there are no external influencers. This also applies to “internal” influencers such as desire or emotion: a Kantian autonomous person does not act impulsively. Such a capacity, by which a person distances himself/herself from influencers, is founded on the usage of practical reason. As long as a person has rationality, understood here as the usage of practical reason, s/he is autonomous. This is why, for Kant, a person subordinated to any universal moral law is still free. Just because a universal moral law is the product of rationality, it can be universal, and any rational person obeying it does not lose any freedom. In other words, the conclusion made by a person who uses his/her practical reason is the same as the conclusions made by others who are also rational.

This is how agreement or consensus can be achieved. Rawls’ veil of ignorance is “visualization” for such a birth of a universal law. By sending everyone back to the veil of ignorance, Rawls mobilizes, as an imaginary exercise though, people’s capability of using practical reason to detach themselves from all the influencers, and by doing so achieves a universal moral principle or value, namely justice. Such an imagined example then becomes an ideal model for a pluralist society to find common “oughts” that every individual is willing to follow and by which oppression can be eliminated. The tension between the individual and the collective is then gone; everyone acts in the same way but is still free, thanks to rationality.

---

29 To be sure, there are at least four interpretations for the term “autonomy” (Feinberg 1989), but in general they share the same meaning as the one quoted.

30 This is probably why, not only for the theory of technological mediation, but also that of nudge, even though humans are highly influenced by external factors, their autonomy is not really harmed. In the theory of technological mediation, what humans think and do are not deter-
However, real-world humans rarely reach conclusions in such a way. Most of the time, people’s rationality is “bounded”, which means that “the capacity of the human mind for formulating and solving complex problem is very small compared with the size of problem, whose solution is required for objectively rational behavior in the real-world” (Simon 1957, 198). People create and hold their own simplified models of the world, by which they make decisions; moreover, these “imperfect models of the world are not simply a product of our own psychological limitations, but are a consequence of the socio-cultural and organizational settings in which [they] are located” (Jones, Pykett, and Whitehead 2013, 5). More and more scientific studies have supported such a view of humans in recent decades. For example, people’s thinking processes are mostly automatic rather than deliberative (Kahneman 2011); behaviors are executed first by the right side of the brain and then explanations for the behaviors follow thanks to the left side (Gazzaniga 2012); thinking is nearly an illusion because it is more about fragments popping up, combining, or splitting randomly (Chater 2018). Besides, as mentioned in Chapter 3, people’s decisions are highly influenced by choice architecture (Sunstein and Thaler 2009), and human perception and action are always technologically mediated (Verbeek 2011). In other words, the image of the human comes closer and closer to the one conceived by Skinner.

For Skinner, rationality is something like the mind; it is an *explanandum* rather than *explanans*. Human behavior, including thinking of course, is the product of its environment, and there is no such thing as rationality, morality, or beauty in the environment itself. Humans are made to be “reactional” rather than rational. This does not mean that humans are irrational; rather, humans are essentially “non-rational”, i.e., having little or nothing to do with rationality. Such a Skinnerian image of humans poses a serious challenge to classical democracy: If there is no such thing as rationality that can be a guarantee for agreement or consensus, the tension between the individual and the collective cannot be relieved anymore without oppression.

minded because “technological mediation does not take away the rational character of mediated actions and decisions” (Verbeek 2011, 61). Similarly, the theory of nudge does not deny that people have the ability to be rational (although being rational can be time- and energy-consuming) and therefore is inherently libertarian, leaving autonomy nearly untouched (Evans et al. 2017). This is also why, unsurprisingly, both theories appeal to classical democratic procedures as a way to design or assess the influences of technologies.
Skinner might be wrong, to be sure. But the recent developments exemplified in the above studies show that the Kantian image of humans is becoming less reliable than previously considered. Even though human beings can be rational ultimately, it is not realistic to assume that real humans have unlimited time and energy to be so. Continuing with such an ideal image of humans, therefore, is unlikely to offer us an opportunity and possibility to tackle the problem of environmental crisis in a more collective way. We need, while not willing to give up the spirit of democracy, to come up with an alternative model of democracy founded on a non-Kantian image of humans. It embraces the assumption that humans as mediated, nudged, or reinforced beings, are always in relations to humans as well as to non-humans. As such, a non-Kantian image of humans is usually regarded as “worse” than the Kantian one, so democracy based on it will have “downward capability”, meaning that it still works even though humans are indeed capable of rationality.

What would such a democracy look like? How does it deal with tension? Moreover, can it help with the environmental crisis? Does it allow the design and implementation of behavior-steering technology? Constructing a new model for democracy can never be an easy task. What I would like to do here is to offer a potential candidate for further development and elaboration.

8.3. Chantal Mouffe’s Agonistic Democracy

A potential framework for constructing a new model of democracy can be found in the work of Chantal Mouffe, the political philosopher famous since the late 1980s for the co-authored book Hegemony and Socialist Strategy: Towards a Radical Democratic Politics (2014/1985) with Ernesto Laclau. She has strong critiques of deliberative democracy and argues for an alternative called “agonistic democracy”. To introduce her ideas, we need to start with her objection to the possibility of consensus.

To solve the tension between the individual and the collective, a “deliberative turn” in democracy has been proposed, promoted, and is expected to be the best solution (Dryzek 2002). But Mouffe (2005; 2006; 2009; 2013) strongly contests that non-oppressive consensus is no more than an delusion and cannot be the

31 We may also need alternative models for education and the economy, given that the current models are both based on the belief in the Kantian image of humans.
goal for democracy. The reason, as Mouffe provides, is the impossibility of all-inclusiveness. A “we” cannot exist if there is no “they”. Only after a boundary has been drawn, can a collective emerge and be identified. In other words, the “outside” is a condition for the “inside”, and this “constitutive outside” is impossible to eradicate. Inclusion, therefore, always comes with exclusion, and consensuses with dissensus. Conflicts are always there, whether noticed or not.

This is where Mouffe makes a distinction between “politics” and “the political”:

By “the political”, I refer to the ontological dimension of antagonism, and by “politics” I mean the ensemble of practices and institutions whose aim is to organize human coexistence. These practices, however, always operate within a terrain of conflictuality informed by “the political”. (Mouffe 2013, xii)

The political, therefore, “must be conceived as a dimension that is inherent to every human society and that determines our very ontological condition” (Mouffe 2006, 3), by which pluralism can flourish and be ensured. For Mouffe, the main mistake of deliberative democrats is their overlooking of the political, and believing in the reachability of inclusive consensus. To make deliberative democracy work in reaching consensus, it is unavoidable for liberal theorists to appeal, explicitly or implicitly, to participants’ capability of being rational.

At this point, Mouffe shares the same criticism with communitarians toward Rawls. Moreover, she regards Habermas as similar to Rawls because his communicative rationality is merely a substitution for instrumental rationality. In short, despite of their differences, the

point of convergence between the two versions of deliberative democracy is their common insistence on the possibility of grounding authority and legitimacy on some forms of public reasoning and their shared belief in a form of rationality which is not merely instrumental but has a normative dimension: the “reasonable” for Rawls, “communicative rationality” for Habermas. (Mouffe 2009, 85–86)

As a result, on the one hand, anything that cannot be discussed by using reason tends to be expelled to the private sphere, and on the other hand, “irrationality” becomes the main excuse to ignore or disprove competitive viewpoints (Mouffe 2006). In other words, while deliberative democracy has been promoted for the purpose of inclusiveness, its expectation on consensus can only be reached through the action of exclusion.
Would it be better if we go back to so-called aggregative model of democracy? Aggregative democracy, regarded as the other major model of democracy in contemporary politics (Young 2002, 18–26), aims for the determination of what social policy to adopt by soliciting people’s preferences and aggregating them through voting. Leaving aside the deficits that have been attacked by deliberative democrats, the aggregative model is problematic as well because in such a model “[i]ndividuals are portrayed as rational beings, driven by the maximization of their own interests and as acting in the political world in a basically instrumental way” (Mouffe 2005, 12–13).

For Mouffe, there are always factors irrelevant to rationality influencing humans in any decision-making processes, and the most important one of them is “passions”. No matter where passions come from (maybe blood ties, religions, cultural traditions, or something else), it is impossible to ignore the effect of passions on people’s decision on whether or not to join a collective:

By putting the accent either on the rational calculation of interests (aggregative model), or on moral deliberation (deliberative model), current democratic political theory is unable to acknowledge the role of ‘passions’ as one of the main moving forces in the field of politics and finds itself disarmed when faced with its diverse manifestations. (Mouffe 2005, 24)

Given that the political, the ineluctability of antagonism, is ubiquitous, any social order without exclusion is made possible only by oppressing and silencing some insiders. Such a social order is a “hegemony”, the term borrowed from Antonio Gramsci and slightly revised by Mouffe and Laclau, meaning “a contingent articulation of power relations that lacks an ultimate rational ground” (Mouffe 2013, 131). This understanding of the nature of social order implies that power is the very condition for, and inherent to, any unity. Moreover, taking an affirmative stance to Foucault, Mouffe considers any social order that justifies itself by claiming that its foundation is objective and therefore permanent is however temporary, because

[w]hat is at a given moment considered as the “natural” order—jointly with the “common sense” which accompanies it—is the result of sedimented practices; it is never the manifestation of a deeper objectivity exterior to the practices that bring it into being. (Mouffe 2005, 18)

Therefore, democracy should not aim at forming a hegemony-free collective by eliminating power, but rather be open to the growth of any counter-hegemony (which are also the practices of power). The method for such growth is “to
envisage democratic citizenship from a different perspective, one that puts the emphasis on the types of practices and not the forms of argumentation” (Mouffe 2009, 96, italics in original).

However, the relationship between hegemony and counter-hegemonies or the relationship between one type of counter-hegemony and another, should not take the form of “only one can survive”, i.e., either you or we die. Taking the political seriously as the core of any human association, Mouffe suggests that the main task of democracy is to accommodate and establish the discriminate of the “we/they” relation in an adequate way that, on the one hand, “rejects the very possibility of a non-exclusive public sphere of rational argument where a non-coercive consensus could be attained” (Mouffe 2009, 33), and on the other hand, “take[s] a form that does not destroy the political association” (Mouffe 2005, 20). The key here is to transform the we/they relation from antagonism to agonism:

While antagonism is a we/they relation in which the two sides are enemies who do not share any common ground, agonism is a we/they relation where the conflicting parties, although acknowledging that there is no rational solution to their conflict, nevertheless recognize the legitimacy of their opponents. They are “adversaries” not enemies. This means that, while in conflict, they see themselves as belonging to the same political association, as sharing a common symbolic space within which the conflict takes place. (Mouffe 2005, 20)

Such a democracy, accordingly, is named by Mouffe “agonistic democracy”, or sometimes called, due to its foundation on the political, “radical democracy”.

For agonistic democracy, whether or not humans are rational is irrelevant or at least unimportant, and the factors which have been illegitimized in deliberative and aggregative democracies can find their place. Given that the discrimination of we and they is largely driven by passions; agonistic democracy does not expel or eliminate them from the sphere of politics:

In my view, the advantage of such a model is that by recognizing the role of passions in the creation of collective identities, it provides a better understanding of the dynamics of democratic politics, one that acknowledges the need for offering different forms of collective identification around clearly defined alternatives. (Mouffe 2013, 139)

Reinterpreting democracy as an agnostic way to deal with ongoing divergences in human society is crucial for contemporary politics. The malfunction and de-
The current models of democracy at the global level in recent years have been foreseen by Mouffe:

A well-functioning democracy calls for a confrontation of democratic political positions. If this is missing, there is always the danger that this democratic confrontation will be replaced by a confrontation between non-negotiable moral values or essentialist forms of identifications. (Mouffe 2013, 7)

To restore public confidence in democracy and prevent the rise of dictatorship, the agonistic model of democracy, the model rooted in the political, accepting the existence of hegemony and not prioritizing rationality, is worth being taken seriously and developed further. To be sure, Mouffe suggests agonistic democracy at a very conceptual level and rarely provides details about institutions or procedures by which such a model can be implemented. However, Mouffe’s conceptual work offers us an opportunity to deal with the environmental crisis and accommodate behavior-steering technology in politics.

8.4. The Political Framework for Behavior-Steering Technology

The strength of Mouffe’s agonistics democracy lies in its radicalness and openness. It is helpful in many ways. First of all, it does not base itself on the assumption that humans are rational beings; rather, a “deconstructive approach reveals that the vocabulary of Kantian universalist morality, in which the universality of moral imperatives is justified by their rational form, is profoundly inadequate for thinking about ethics and politics” (Mouffe 2009, 135). Moreover, “[t]he reformulation of the democratic project in terms of radical democracy requires giving up the abstract Enlightenment universalism of an undifferentiated human nature” (Mouffe 2006, 13). Accordingly, agonistic democracy, in its giving up of the Kantian image of humans, leaves room for non-rational humans.

Moreover, as agonistic democracy does not depend exclusively on rationally, it may allow various factors to come into play. As pointed out, passions, the factor that has been illegitimized and excluded from, and by, classical democracy, is fully allowed in agonistic democracy. Similarly, technologies, which have been regarded as a threat to democracy and excluded from the realm of politics, in principle can and should be accepted as well. This means that when deciding the shape of a collective—how to live together—through any procedure in agonistic democracy, technologies have to be counted as an active factor. In other words,
for agonistic democracy, mediating technologies, nudges, reinforcers, or behavior-steering technologies are things that can help with the practices of democracy, rather than regarded as troublemakers. Not relying upon human rationality and allowing non-rational factors make agonistic democracy a political framework that is compatible with the non-Kantian image of humans.

Agonistic democracy also gives us a perspective to understand the meaning of the design and implementation of behavior-steering technologies. As we already know, agonistic democracy does not seek agreement or consensus as justification for political practices; rather, it welcomes disagreements, contestations, and confrontations between hegemony and counter-hegemonies, and these conflicts proceed not merely by argumentation, but also by practices. This echoes my suggestion in Chapter 2 that human doing rather knowing is what needs to focus upon. And no doubt, designing and implementing behavior-steering technology are such practices. That is to say, in agonistic democracy, designing or implementing behavior-steering technology should be regarded as a political action, which may reproduce hegemony or create and strengthen counter-hegemonies.

The meaning of speed bumps, for example, can be reinterpreted from this perspective. Today, most roads are made smooth, which reflects the ideology and practice that “people have freedom to drive as fast as possible unless there is an upper limit”. This is the hegemony that dominates the driving behavior and traffic design. People have been, and are, shaped by these roads to behave in such a way, and thereby the hegemony can be consolidated and maintained. Speed bumps, on the contrary, are designed for counter-hegemony, in which the ideology and practice go in opposite directions: “people should drive as slowly as possible unless there is a lower limit”. Drivers, then, are shaped by these roads-with-speed bumps to behave in a way that creates and strengthens the counter-hegemony. In agnostic democracy, installing speed bumps is welcomed because it challenges the current hegemony and materializes the nature of politics (the political). In other words, the design of speed bumps is not justified by arguing that driving fast is wrong or that driving slowly is right; instead, it is justified only because it is a counter-hegemonic practice.

Such a line of thinking is acceptable only within the framework of agonistic democracy. This is where the environmental crisis gets a great chance to be solved. Recall Mouffe’s diagnosis of deliberative democracy: for the sake of reaching consensus, deliberative democracy often excludes issues by transferring
them from the public to private sphere. By so doing, it moderates the tension between individual freedom and collective norms:

The absence of a single substantive common good in modern democratic societies and the separation between the realm of morality and the realm of politics have, no doubt, signified an incontestable gain in individual freedom. But the consequences for politics have been very damaging. All normative concerns have increasingly been relegated to the field of private morality, to the domain of ‘values’, and politics has been stripped of its ethical components. (Mouffe 2006, 65)

This is exactly what happens to environmental issues. In classical democracy, driving, biking, or walking is merely an option for personal lifestyle; it is a choice in the private sphere, having nothing to do with the public. Convenience, health, and sustainability are all values; how to rank them is, again, a private issue. Unfortunately, CO₂ goes literally the other way around. No matter which choice you make, its impact will appear in the public sphere. But, again, classical democracy has no competence in dealing with it.

From the perspective of agnatic democracy, any social order, in other words, a specific ranking of values, is a hegemony that has been produced and maintained by hegemonic practices. The current anti-environmental order, in which convenience and comfortable are ranked higher than energy-saving and zero-waste, should be considered as hegemony, shaped and consolidated by the practices that steer people toward being nature-unfriendly. The designs of current buildings, transportation infrastructures, or electronic devices embody such practices.

Fortunately, “[e]very hegemonic order is susceptible of being challenged by counter-hegemonic practices, i.e. practices which will attempt to disarticulate the existing order so as to install another form of hegemony” (Mouffe 2005, 18). Accordingly, to design green behavior-steering technologies such as narrow-curve-cul-de-sacs is to counter the anti-environmental hegemony. Most importantly, such an action is not backed up by any ethical claim such as “unsustainability is wrong”; it is justified by its nature: a counter-hegemonic practice. At this point, society in Mouffe’s theory seems more open than Popper’s open society. While identifying what is wrong is crucial in the practice of piecemeal engineering, the questions about right or wrong and good or bad are all suspended in the framework of agonistic democracy.

Such a suspension offers us more opportunities to make good use of behavior-steering technology because there is no need to limit the purpose of behavior-
Design for Green

steering technology in removing anti-environmental behavior. Recall the argument in Chapter 6: The reason that we only design for removing anti-environmental behavior is, according to Popper, that epistemologically we cannot know what is right. But from a Mouffeian perspective, whether or not we can know what is right (or what is wrong) is beside the point because they are defined by hegemony. Accordingly, if we adopt the agonistic model of democracy, the restrictive principle that only designs for removing anti-environmental behavior, will become redundant. Agonistic democracy allows us to steer people in a pro-environmental direction not only because such an action is exactly a counter-hegemonic practice, but also because the purpose of counter-hegemony is to replace the current hegemony. In other words, it is not because being pro-environmental is good or right that we design green behavior-steering technologies; quite the opposite, it is because we design green behavior-steering technologies so that being pro-environmental can become good or right. At this point, Mouffe echoes Skinner. If values emerge from the contingencies of reinforcement, according to Skinner (Chapter 5), then what we need to do is exactly to arrange specific contingencies of reinforcement for specific values.

Should we worry that being pro-environmental may become a hegemony one day? No! Just as agonistic democracy allows the current anti-environmental hegemony to be challenged by the pro-environmental counter-hegemony, it will also allow the value of being pro-environmental to be challenged once the value becomes a hegemony. Moreover, according to “the political”, it is not possible to eradicate anti-environmental practices because there are always people who support this value, and the conflicts between the two sides will always continue. In other words, challengers may become challengees, and vice versa. This is how agonistic democracy works. Thereby, agonistic democracy may give a new

---

32 The idea of the political is descriptive; however, it has a normative implication. Counter-hegemony is not only ineradicable but should also be allowed to be so. It might be too much to say that we should proactively encourage the practices of counter-hegemony, but it is plausible to say that we should not inhibit or forbid such practices. This is probably why, although Mouffe’s idea has great potential to enrich pro-environmental practices, it is still difficult to find articulation between the framework of agonistic democracy and environmental issues; after all, agonistic democracy leaves room for anti-environmental practices. Mouffe herself rarely mentions environmental problem too; the only related, albeit weakly, work I could find is a chapter in her 2013 book Agonistics. In that chapter, environmental problems are regarded as the symptom of capitalism (the current hegemony) and may be solved by counter-capitalistic practices, such as artistic activities (Mouffe 2013, 85–105).
meaning to sustainability: Neither the value of being pro-environmental nor that of being anti-environmental can take the position of top value forever; our society is always sustained in-between. Sustainability, in other words, is not about finding a point of balance between nature and humans; rather, it refers to a permanent oscillation between pro- and anti-environmental-ness.

Last but not least, communities, the locus for practicing piecemeal-behavioral engineering, also make more sense in the framework of agonistic democracy. Each community is exactly a “we” against others, just like Walden Two, Los Horcones, or Village Homes is the “we” that counters the mainstream practice in its time. A community, therefore, is not only the unit to apply behavior-steering technology, but also an embodiment of the political. Staying small is still crucial to the design and implementation of behavior-steering technologies, but the reason is no longer to avoid totalitarianism because such a reason only applies in the framework of classical democracy. However, the other reason still stands: the nature of behavioral modification is experimental. We always need to know which behavior-steering technology works and which does not; with a small population, we can fix problems once we find something going wrong. Piecemeal-behavioral engineering, therefore, is still the best approach to the design and implementation of behavior-steering technologies in an agonistic democracy.

8.5. Concluding Remarks

The radical challenge that the environmental crisis poses to us is political, concerning how to deal with the tension between the individual and the collective. While piecemeal-behavioral engineering is the most feasible and acceptable approach for behavior-steering technology within our current liberal democracy, a shift in political framework is however necessary if we are aiming for more substantial change. Neither deliberative nor communitarian democracy can solve the problem because they look for inclusive consensus to justify collective behavioral change, which is possible only if people can be rational. However, it is not the case in the real world; humans are found to be non-Kantian, and closer to the Skinnerian image of humans. As the Kantian image of humans becomes unreliable, a political framework neither basing itself on such an image, nor aiming for inclusive consensus, has to be developed. The agonistic democracy, as suggested by Mouffe, is a potential candidate. On the one hand, by distancing itself from the necessity of rationality and acknowledging the inevitability of in-
fluencers, agonistic democracy is competent to accommodate the non-Kantian image of humans and therefore allows external factors such as technology to come into play. On the other hand, by taking “the political” as the ontological foundation, it not only turns ethical questions into political confrontations between hegemony and counter-hegemonies, but also gives importance to practices rather than argumentation. Agonistic democracy, therefore, makes room for behavior-steering technologies to create, produce, strengthen, or change values and their ranking. The design and implementation of green behavior-steering technologies is no longer a threat to human freedom and democracy; rather, they are counter-hegemonic practices in opposition to current anti-environmental hegemony in which most people in fact have little freedom to act the other way around. Moreover, the approach of piecemeal-behavioral engineering still applies, embodying the spirit of agonistic democracy as well as enabling the development of pro-environmental counter-hegemony.
9 Conclusion: Activating Technologies

This dissertation, as we have seen, aimed to synthesize several theories and ideas constructed in difference research fields into an approach for analyzing and practicing green behavior-steering technology. What, then, are the implications of such an approach? After showing the necessity of behavior-steering technology in tackling the environmental crisis, arguing for the advantage and usefulness of MBST, reducing ethical concerns about human freedom, and further exploring a political framework to accommodate the practice of behavior-steering technology, in this concluding chapter I will discuss what contributions this research may make to related fields. I will start with a brief review of previous chapters to clarify the connections between them and to highlight the present dissertation as a lengthy defense of green behavior-steering technology as ethically adoptable and politically practicable. After that, I will explore the implications of my approach for three closely related fields as part of the background of this dissertation. The first is sustainability advocacy, the second is design in general, and the third is technology studies. Three concepts: “educational environment”, “design to discourage” and “politicalizing technology” will be proposed for future work.

9.1. Recapitulation

Starting with the concern for the environmental crisis, this dissertation argues for behavior-steering technology as a feasible and acceptable approach to improve people’s environmental behavior. After the introduction, in Chapter 2, I argue that traditional strategies, represented by environmental education, aiming to induce behavioral change tend to fail because they rely on the assumption that human knowing is the initiator of human doing, and neglect the fact that human behavior has to be supported or can be produced by physical conditions. Such inattention to the knowing-doing gap reflects, and results from, the asymmetric treatment to anthropocentrism. While knowing-anthropocentrism has been overcome by environmental ethics, doing-anthropocentrism is nearly untouched. Given that in the Anthropocene humans unavoidably take the position of master, non-doing-anthropocentrism cannot be reached by leaving nature out there free from any interference. A more realistic strategy to distance ourselves
Design for Green

from doing-anthropocentrism is to employ technology to improve nature’s, as well as humans’, environmental behavior. Therefore, to tackle the environmental crisis, a focal shift from human knowing to doing is necessary, and the knowing-doing gap has to be bridged or bypassed by behavior-steering technologies.

To take advantage of behavior-steering technologies, we need to have a basic knowledge about them. So in Chapter 3, I introduced three classifications developed in the field of philosophy of technology to clarify the relations between humans and technologies, and strengthen the point that human behavior has been, and still is, highly shaped by technologies-in-use. However, they did not provide a clear picture about the connection between how effective behavior-steering technology is and how it works. I then suggested a simpler but useful classification, which distinguishes two types of behavior-steering technology by their differences in mechanism. While IBSTs (informational behavior-steering technologies) take intrasomatic routes, mobilizing users’ consciousness or unconsciousness to generate behavior, MBSTs (material behavior-steering technologies) take extrasomatic routes, targeting users’ bodies as a locus to apply material conditions or constraints. The pros and cons of each type are clear: Qualitatively and quantitatively, IBSTs are ineffective and therefore regarded as soft and acceptable, but MBSTs are, in contrast, very effective and tend to be regarded as hard and ethically problematic. Its high effectiveness supports MBSTs’ potential to improve people’s environmental behavior in a collective way; it has been the main reason to set limits to the design, implementation, and development of MBSTs.

The main worry concerning behavior-steering technologies is the loss of human freedom, as pointed out in Chapter 4. To address the concerns, ethical guidelines and defensive arguments have been made for IBSTs. For both persuasive technology and nudge, the central principle is to avoid intervening with users’ ends, either by targeting their means only or by minimizing the cost of opting out. Unfortunately, such a principle cannot be applied to the case of the environmental crisis; as it requires collective changes in behavior, we have difficulties in aligning user’s ends with ours because, due to the temporal and spatial dispersion, the causal relationship between one’s behavior and its environmental consequences is hard to identify. As a result, intervening users’ ends is not something avoidable if we would like to significantly improve people’s environmental behavior. Such an ends-intervention is where MBSTs are competent. To be sure, such intervention can be done with personalized IBSTs; however, such IBSTs are likely to make things worse by creating privacy and transparency is-
Conclusion: Activating Technologies

sues. The concreteness and tangibility of MBSTs make the approach of MBST the best option. However, these features also increase the cost of opting out and therefore provoke worry. In other words, while green MBSTs are unavoidably strong and ends-intervening, ethical concerns about users’ freedom tend to rise against the design and its implementation. If the approach of MBST is indispensable in tackling the environmental crisis, then, as pointed out, such a dilemma urges us to further understand the nature of human behavior.

The work of B. F. Skinner helps us to take the next step. In Chapter 5, in referring to Skinner’s *Walden Two*, his theory of behavior and also the idea of behavioral engineering are introduced. As human behavior is shaped through the processes of contingencies of reinforcement, it is the product of its environment. This implies the possibility to change a person’s behavior by changing the technologies s/he interacts with. The technique recommended by Skinner is positive reinforcement, not negative reinforcement, nor punishment. Skinner’s *Beyond Freedom and Dignity* offers various arguments against the objections to behavioral modification. It shows us that the concepts of freedom and dignity are misguided not only because scientifically they are unlikely to exist, but also because they seriously constrain the use and development of the technology of behavior and therefore leave humans to traditional punitive control. With the help of Skinner, I argue on the one hand that behavior-steering technologies are a must for improving people’s environmental behavior because the consequences of environmental problems are too remote to have shaping force, and on the other hand that the worry concerning human freedom can be largely reduced by employing the technique of positive reinforcement. However, Skinner still left two questions for us to solve. First, given that sustainability cannot be realized by creating pro-environmental behavior alone, which means the measures of aversive control are unavoidable to apply, what can we do about the issue of human freedom? Second, in taking an evolutionary perspective on values, Skinner helps very little to justify that being pro-environmental is good and therefore should be the direction in which people are going to be steered. That is to say, what goal should we design if being pro-environmental cannot be well defended?

Fortunately, the answers lie implicitly in Skinner’s experimental approach to behavioral engineering. By looking into two real-world Skinnerian communities—Twin Oaks and Las Horcones—in Chapter 6, I first strengthened the argument that behavioral modification rather than other strategies is the key to social change and that attaching a scientific view to human behavior is crucial to make such a change come true. Moreover, these two communities, together with
Walden Two, reveal the most important dimension of Skinner’s behavioral engineering: small-scale experimentation. Scaling down to a manageable size not only makes the effects of behavior-steering technologies easier to transfer from human-technology relations to human-human ones, but also reduces the difficulty in testing, adjusting, or even revoking the technology-in-design, by which unexpected side effects can be minimized. At this point, I then argued that Skinner’s behavioral engineering resonates with Karl Popper’s idea of piecemeal engineering. By allying with Popper, the two questions left in the previous chapter can be well answered. As knowing what is good or right is epistemologically impossible, there is no need to justify being pro-environmental and design for it anymore; rather, our main task is to identify anti-environmental behavior and design to remove or decrease it. Moreover, without forcing everyone in one direction, we can make all other options enjoyable, pleasant, or satisfying. By doing so, the net consequences of the steered behavior can still remain positive; therefore, the concern about human freedom can still be largely reduced, even though positive reinforcement is not the only technique being employed. In short, piecemeal-behavioral engineering, a combination of Skinner and Popper, can underpin the approach of behavior-steering technology and make the approach not only adoptable but also acceptable for tackling the environmental crisis.

As a problem-oriented research, in Chapter 7, I turned the concepts, perspectives, and arguments developed through previous chapters into a list of seven design recommendations. If there is any possibility to concentrate these recommendations into one sentence, the result must be “Design and experiment with green MBST at the level of community to reduce anti-environmental behavior”. With the example of Village Homes, I illustrated the usefulness, practicality, and soundness of the approach I have been constructing and defending for green behavior-steering technologies. The design of narrow-curvy-cul-de-sacs beautifully demonstrates the advantages and strengths of MBSTs. Without such an MBST, the achievement of the community would never be reached, and the ethos of being green could never be realized. MBSTs, again, are proved to be a must in the advocacy of sustainability.

For readers who would like to take action immediately and apply the approach straightforwardly, it is not necessary to engage with the political implication of the design and implementation of green behavior-steering technologies. The whole approach I have developed is entirely compatible with our liberal democracy. However, it might be worth taking on the political challenge posed
by the environmental crisis. In Chapter 8, the discussion moves to the level of politics, in arguing for the need of an alternative approach to democracy. In prioritizing the individual instead of the collective, liberal democracy appears to be hopeless in solving global-scale environmental problems. However, turning to communitarian democracy does not really help because it is also based on a problematic image of humans. As the non-Kantian image of humans is more “real”, it should be the basis for a new model of democracy. A potential candidate for such a model is, as I have argued, agonistic democracy. Grounded in “the political”, agonistic democracy is not about consensus, but in favor of confrontations for democracy to flourish and vitalize. Moreover, it does not require humans to be rational, and therefore allows for external influences. In an agonistic democracy, both the design and the implementation of behavior-steering technologies are counter-hegemonic practice, by which they justify themselves. Moreover, moral questions like whether being green is good or wrong are suspended and the answers to them are not a precondition for technological interventions. In such a way, agonistic democracy probably is the most proper political framework to accommodate the inevitable interweaving of humans and technologies.

9.2. Toward Educational Environment

Through governmental or non-governmental promotions and campaigns in past decades, the idea of sustainability has become widely spread and recognized. Companies try to manufacture nature-friendly products via nature-friendly processes, and educators attempt to teach the general public to value and appreciate “Mother Nature”. While the former is equipped with technologies to reach the goal, the latter seems not to be. To be sure, educators have various tools to improve their teaching, such as brochures, whiteboards, or PowerPoint, but these tools are made and used mainly to help with the delivery of arguments, theories, and information about substantiality. The technologies that can directly change people’s environmental behavior are still absent, which weakens the effectiveness of environmental education.

The approach of behavior-steering technology can help here. As Skinner reminds us: What needs to be changed is people’s behavior, not their minds. We should pay more attention to human behavior itself and make efforts to understand how people’s behavior has been, and can be, shaped by daily-life artifacts. To be clear, we are not denying the existence of the human mind, and we do not
reject the contributions that education can make. We just do not see the human mind as a necessary detour to the change of human behavior. Behavioral engineering works well, as shown by Skinner, even though there is no such thing as mind assumed in its theory.

Given that most people currently live in a built environment and interact with various artifacts for most of their lives, Skinner’s thoughts on human behavior become salient and significant. If what people do is the product of their environment, it will be an unavoidable responsibility for us to carefully shape and reshape our surroundings. Accordingly, I propose the concept of “educational environment” as a complement to environmental education. Composed of various artifacts designed for improving environmental behavior, a built environment can be more effective than single behavior-steering technology in moving people to a sustainable lifestyle. If the purpose of education is basically to direct students to do the right things and not to make mistakes, then such a built environment is also “educational” because it does exactly what education has been tasked to do. Recall the example of the architectural design of the stairs-elevator that was briefly discussed in Chapter 2. In observing employees and visitors going up and down by the stairs rather than the elevator, we will find difficulty in telling which of them were “educated” by environmental education and which by the built environment.

One may worry that such reliance upon artifacts for moral behavior could have negative effects. For example, Shannon Vallor (2015) argues that just like the invention of technologies doing the work that humans have been doing causes the worry of deskilling, depending on technologies may create the issue of “moral deskilling”. From an Aristotelian perspective on morality, being moral needs frequent practice, but “profound technological shifts in human practices, if they disrupt or reduce the availability of these opportunities, can interrupt the path by which these moral skills are developed, habituated, and expressed” (Vallor 2015, 109). However, the approach developed in this dissertation is unlikely to be the case. If moral skills are about how to do the right things, an educational environment helps to cultivate rather than deprive such skills; it provides the environment in which people practice moral behavior repeatedly and constantly. As my study on the case of the stairs-elevator design has shown, the employees who become habituated to take the stairs tend to look for stairs when entering other buildings (Hung 2014). In other words, moral skills learned inside an educational environment can appear and be applied outside.
On the other hand, if moral skills are about learning how to make decisions, the proposal of educational environment however is not in opposition to such learning. Proposing the idea of educational environment does not mean that we should discard or ignore environmental education. People still need to know why they have responsibilities to the Earth and how these reasons can be justified. Environmental education, working with environmental ethics, is still valuable in offering answers to those questions and thereby helps people to move away from anthropocentric arrogance. Moreover, as I have discussed in Chapter 4, compared to other things, most environmental behaviors are minor and tiny; it can be exhausting for people to deliberate and make decisions on such daily-life behaviors. After people have learned through environmental education, an educational environment can help to release them from the burden of making environmental decisions; it makes moral skills easier, not harder, to acquire. At this point, again, the approach of educational environment is a complement to, rather than a replacement of, environmental education. They are not exclusive approaches for the same goal, and the result can be substantial if they work together.

One may also feel discomfort with the concept of an educational environment, worrying about the risk of manipulation or technocracy. This unease is understandable, particularly in our liberal-democratic society. However, as argued in Chapter 6, the approach of piecemeal engineering can reduce the risk and soften the unease; teaching children about the theories and principles of behavioral modification helps as well, just like what Los Horcones has been and still is doing. In fact, this unease reveals the fact that there is no overwhelmingly convincing argument for the value of nature. No one will get blamed for controlling or governing people if its purpose is to protect human lives or rights from being threatened or harmed. Unfortunately, until now we have not successfully defended the rights of nature. This disappointing fact leads us to another implication that this dissertation may have for the promotion of sustainability.

A lesson we can learn from Mouffe is, as mentioned in Chapter 8, that any settled social order is a hegemony, within which norms, customs, or ethos are the products of hegemonic practices. This implies that we should put the promotion of sustainability in the realm of politics rather than morality. In other words, promoting sustainability should be practiced as a political action, not moral advocacy. To be sure, green movements and green parties are political in themselves, and their development should be an ongoing one. What I mean by political action is more than that. As argued by Mouffe, the political is what un-
derlies politics, meaning that any action concerning the political logically belongs to the realm of politics. Accordingly, every time we design and implement a green behavior-steering technology to steer users toward a more sustainable direction, we are taking a political action. As we are countering the current hegemony that has been produced and maintained by unsustainable practices for centuries, the action of design for green is political in its own right. In viewing sustainability promotion as political action, we escape from the mire of countless ethical debates and move toward something more effective and efficient. To be sure, I am not denying the possibility that someday we might successfully construct a perfect argument for environmental morality, but before that, we need a framework that allows us to do more than what we have been allowed to do thus far.

9.3. Design to Discourage and as Experimentation

Technology, as shown by the theory of technological mediation, has profound impacts on human experience and praxis. Since human beings started making and using tools, technologies have shaped our existence; it already becomes impossible to live without interacting with various technologies. Such a fact of “no-way-out” charges us with the responsibility of paying attention to the mediating effects that technologies may have on human perception and action. However, in the framework of technological mediation, technology only partly shapes human existence. This may bring about confusion because it can be hard to tell, for any behavior, which part is made by the technology-in-use and which part by the person using the technology. Therefore, in designing behavior-steering technology, it might be helpful to take a Skinnerian view and see technology as a determinant.

As Skinner has argued, human behavior is the product of its environment only. From this perspective, all technologies have a determining impact on what humans think and do; they are all behavior-steering technologies, even though they are not designed with the intention to steer user behavior. Taking a Skinnerian perspective helps us to take care of all technologies-in-design and thereby to find their behavior-steering effects. Moreover, such a perspective reminds us to think carefully about the environment we are going to make because it will determine not only how people behave within it, but also, albeit indirectly, the shape of a society. Accordingly, the idea of educational environment is not lim-
ited to advocating sustainability; other values, such as politeness, kindness, and justice, can be the goals of educational environment as well.

The design recommendations for green behavior-steering technology in Chapter 7 can be applied to all behavior-steering technology as well. For example, Recommendation 7, “Aiming at anti-environmental behavior but complementing with positive reinforcers”, can help with the recent promotion of the idea of a “walkable city”. The term has been popularized by the architect Jeff Speck (2013), who believes that problems such as social alienation and public health can be solved by improving walkability in cities. He gives a to-do list for designers to make cities more walkable, including narrowing roads to discourage driving. However, he does not prioritize narrowing roads as a must-do; as I have argued with the case of Village Homes, this may lead the project to failure. If driving is still being reinforced more than walking, then the behavior of walking may not significantly increase, no matter how walkable a city is. Although walking behavior is the target to promote, it cannot be achieved solely by improving walkability. The designs that encourage walking should be complementary to the designs that discourage driving, not the other way around. In short, improving walkability cannot be done without reducing “drivability”.

This is not an infeasible or unreasonable suggestion. In fact, there is already an architect designing in this way for the city of Copenhagen, Denmark, although he does not theorize his strategy:

For many years, Copenhagen had an excellent city engineer who said, “if you can’t park, you don’t drive”. Then he (the architect) said, “I’ll take two percent of the parking out of downtown every year, and I won’t tell anybody so nobody will notice”. So every year, he reduced the amount of parking space and put in more tables, more chairs. ... The place became nicer and nicer, and people changed their habits, taking their bikes or the bus or metro. Over a period of time, the pattern of behavior has changed. You could call it stealth urbanism. (O’Sullivan 2016)

Removing parking spaces from a city apparently increases inconvenience for drivers, and such a strategy shows how crucial “Aiming at anti-environmental behavior first” is for changing the “ethos” of a city.

To be sure, expanding this design recommendation to behavior-steering technology in general may increase the power of the practitioners of technological design and induce worries concerning manipulation and technocracy. However, the other design recommendations I have made can be applied here to ease such worries. It should be noted that such an application does not mean that
these principles and suggestions play the role of moral guides; rather, they are applicable because of the experimental nature of any design. The idea of piece-meal engineering, as discussed in Chapter 6, originates mainly from the practical and epistemic limitations of scientific research. Its effects in reducing ethical and political worries are in fact the “bonus” that we can take advantage of.

The suggestion that the design and implementation of behavior-steering technologies should be regarded as experimentation can find its equivalent in the field of philosophy of technology. For example, Ibo van de Poel (2011; 2013; 2016) has recently proposed the idea of conceiving new technologies as social experiments. He points out clearly that the characteristics of technology-in-development include uncertainty and ignorance concerning its consequences\(^3\), and this makes moral and regulatory evaluations on new technology highly difficult because we lack credible knowledge about its social benefits and harm before it becomes widely used. This is an epistemic limitation concerning the consequences of technology’s social embodiment, which shows again the experimental nature of technology design and implementation. Both van de Poel’s approach and mine emphasize the importance of trial-and-error regarding artifacts, be it behavior-steering technology, technology in general, or emerging technology, as experiments. By taking such a perspective, we can find a balance between blind acceptance of, and hostile resistance against, newly designed artifacts.

Moreover, the approach developed in this dissertation may help with some difficulties that tend to rise in van de Poel’s approach. When suggesting that we view technologies as social experiments rather than merely scientific one, van de Poel notes several unique features of social experiments:

Firstly, they take place outside the lab and involve more and other human subjects than standard experiments, in particular users and bystanders. Secondly, they are not always explicitly carried out or recognized as experiments, so that data gathering or monitoring is sometimes absent. Thirdly, they are less controllable, which makes it more difficult to control experimental conditions and to contain hazards. Social experiments may be difficult to terminate or they may have irreversible consequences. (van de Poel 2011, 287)

\(^3\) According to van de Poel, uncertainty and ignorance are slightly different: “Uncertainty refers to a situation in which we know what might go wrong, but lack the knowledge to express a hazard within a risk. Ignorance refers to the situation in which we do not even know what could go wrong, resulting in unknown hazards.” (van de Poel 2011, 285)
This quotation shows several difficulties that social experiments often confront, such as controllability and irreversibility. However, in developing the approach of behavior-steering technology, we already see technologies-in-design not only as a social experiment but also a scientific one. This means that the approach developed in this dissertation can help to fix these difficulties. For example, by making the size of the user population small, data collecting can be easier, parameters and relations simpler, and revoking or adjusting experiments easier. In short, the idea of piecemeal engineering can enlarge and strengthen van de Poel’s theory.

Experimenting in piecemeal fashion also echoes and helps the approach of moral mediation that has been suggested by Olya Kudina & Peter-Paul Verbeek (2019) to strengthen van de Poel’s conclusions. While also believing that values often change along with the development of technology, they argue that the anticipation of value change cannot be discarded even though the consequences of technological-social experiments are uncertain. They suggest “micro” investigations into how users interact and talk about the technology-in-use, and by doing so the change of values can be detected and the trajectory of such a change depicted. It is not hard to tell that their suggestion, by its nature, is piecemeal in focusing on slight and partial things, and scaling down technological-social experiments can lighten the load of investigation and increase the accurateness of anticipation. Needless to say, the method they have developed to observe and examine the co-constitution of values and technologies can and should be added into the toolkit of behavioral-piecemeal engineering for the design and implementation of behavior-steering technologies.

9.4. Politicalizing Technology

Both philosophy of technology and STS in recent years focus and work on innovative, advanced, or emerging technologies, such as the Internet of Things, autonomous vehicles, CRISPR (clustered regularly interspaced short palindromic repeats), to name just a few. Investigations and assessments on these technologies help human society to avoid, or prepare for, their unexpected and

---

Social experiments by their nature can be regarded as “field experiments”, which differ from laboratory ones mainly in their “uncontrolled environments”, “individuality”, “uniqueness”, “contingency”, and “spatial openness”; such experiments have become more and more important in recent years along with the growth of social sciences (Schwarz 2014).
undesirable harmful consequences. While such a research trend is no doubt valuable and beneficial, there seems something important left behind: mundane artifacts in people’s daily lives. These artifacts are usually low-tech, inexpensive, and therefore widespread; however, they are what people interact with mostly, closely, but unconsciously. They constitute a major part of our living environment. On roads we move from one place to another; with doors and walls we create not only private spaces but also access to public spaces; in buildings we live and work without being disturbed by annoying animals and bad weather. These mundane artifacts have no less influence than advanced and complex technologies on humans.

Most mundane artifacts have “background relations” with humans, meaning that they work in the background and shape the context of our experience without being experienced by us (Ihde 1990). Such invisibility makes them barely attractive to technology studies. However, as Skinner reminds us, a built environment plays a crucial role in shaping human behavior through the mechanism of reinforcement; it shapes not only us but also our offspring. Mundane artifacts constitute the contingencies of reinforcement, and by so doing, they eventually give shape to human society and culture. Overlooking them may make us miss out on the opportunities to understand and reshape the current civilization in a more fundamental way. Technology studies, therefore, need to go beyond the “high-tech centered mode” and shift the focus to those boring, stubborn, unintelligent, and seemingly uncontroversial things.

Moreover, the approach developed through this dissertation has a theoretical implication for technology studies. Thanks to Langdon Winner’s famous question, “Do artifacts have politics?”, and the answer to it, technology’s politics has received a lot of attention and become an important theme in technology studies. According to Winner (1980), technologies comprise politics because they have so powerful an ability to exclude, discriminate, or oppress some groups of people. In the example of Robert Moses’ bridge, it is the group of the poor and racial minorities; in the example of mechanical tomato harvesters, it is the group of farmworkers and agricultural stallholders; in the example of nuclear power plants, it is the group of non-professionals and the general public. These examples give readers an impression that technological power is something problematic and that it is better to avoid technology’s politics. By suggesting democracy as a solution, Winner’s text implicitly puts technology’s politics and human democracy in opposition to one another. This is what underlies the idea and practice of so-called democratization of technology. Generally, the main purpose of de-
mocratization of technology is to include alternative interests and values into the design of technologies, by which technological exclusion can be prevented and avoided (Sclove 1995; Veak 2006). In other words, the better that human democracy functions, the less politics artifacts would have!

Such a “negative” interpretation of technology’s politics may need a revisit. Empirically speaking, technology is not necessarily opposite to democracy. A prominent example has been mentioned by the philosopher of technology Albert Borgmann (2007). After the chamber of the British Parliament had been damaged during World War II, Prime Minister Winston Churchill told his fellow legislators in a talk about the reconstruction of the Parliament, “First we shape our buildings, and afterwards our buildings shape us.” He then suggested not only that a small new chamber is better because this size may induce the feeling of urgency and therefore can urge the legislators to be proactive in dealing with bills, but also that the arrangement of seats should be divided to two face-to-face areas because seeing the countenances and gestures of the opposite party’s legislators can provoke debates between them. His two suggestions are clearly embodied in the chamber that we see today. If actively taking part in discussions with those who hold different opinions is a virtue to improve the quality of democracy, then what the new chamber, an artifact of course, does is to help with democracy rather than undermine it. Similarly, a case study on the structural arrangement of ships shows that the traditional hierarchies of shipboard life can be largely reversed by creating an environment that can increase the interactions between officers and ratings (Lezaun 2011). The creation of shared spaces makes communication with one another occur on a regular basis, and the reduction in the number of hallways can multiply chance encounters. An “offshore” democracy is made possible not only by humans, but also by technologies.

At the theoretical level, technology is however political, and its politics is impossible to eliminate. As we have learned from Mouffe in Chapter 8, there is always discrimination between we and they, and their relation is inevitably antagonistic. This is “the political” more fundamental than politics and constitutive of human societies. In other words, exclusion is not something that can be eradicated, and power cannot be absent in any form of politics, democracy included. From this perspective, the ideal of democratization of technology is however unreachable because there is always at least a group of people who are unwilling to be affected by a technology. Fortunately, in the framework of agnostic democracy, we can find another way to help them. As suggested by Mouffe, agnostic democracy is not about power justification (by real or fictional consen-
Design for Green

sus); rather, it is concerned with the adversarial relationship between hegemonic practices and counter-hegemonic ones. This means that what we can and should do is to create or strengthen confrontations by politicalizing technologies rather than depoliticalizing them.

To make this point vivid, let me revisit the case of Moses’ bridge and imagine a follow-up scenario as follows. To make the beach accessible to everyone, a meeting was held to achieve consensus on how to rebuild the bridge. During the meeting, farmers worried that a high bridge would shade their crops from sunlight, drivers who cross the bridge every day thought that a high bridge would be too dangerous for them to feel safe, and seismologists warned that the geological condition of this area were not good for building a high bridge. Every group had its interests, and there were always objections. No consensus, no new bridge! However, the problem was eventually solved in a way that no advocates of democratization of technology had expected. A new type of car came onto the market; it was smaller, lighter, but cheaper, i.e., affordable for those who could only afford bus tickets before, creating a happy ending for the story. In fact, this imagined scenario is not ungrounded. In his paper questioning Winner, the sociologist Bernward Joerges (1999) states that he has been surprised by the reaction of his students who live in New York: none of them agrees with Winner that the bridge is political because in their experience most people in the downtown area own a car and have no problem to pass through under the bridge to the beach.

It should be noted here that the politics of the bridge is not gone; rather, it is balanced out. The bridge was never democratized, merely countered by another technology, a new type of car. This new technology does not work without power; rather, it has already given shape to our current car-centered culture. This example is entirely Moufffeian. In the framework of an agnostic democracy, designing to counter hegemony is exactly the way to practice, exercise, and vitalize democracy. Technological power, therefore, is indispensable for democracy. As we acknowledge, technologies have morality and such “moral power” is impossible to eliminate; we then turn to the question of how to make use of it. In the same vein, when we find artifacts have politics and that such “political power” is however inevitable, we should turn to the question of how to take advantage of it. While moralizing technology helps artifacts to be integrated into our moral life, we should also make artifacts helpful for our political life by,
apparently, *politicalizing technology*. This is where behavior-steering technology finds its place in the world of democracy.\(^{35}\)

### 9.5. Conclusion

The environmental crisis is made not merely by technologies themselves, but also by the lifestyle created and maintained by them. As people have been shaped by technologies to live in an unsustainable way over the past centuries, it is impossible to change how they live without changing the technologies they use and enjoy. All technologies are behavior-steering. If people’s unsustainable behavior is the product of the unintended behavior-steering effect of past and current technologies, we then need intended behavior-steering technologies to alter it. Perhaps behavior-steering technologies cannot completely solve the environmental crisis, but not taking advantage of them will make us miss out on a great chance to tackle the problem. Designing and implementing behavior-steering technologies often causes worries and objections, but the approach developed in this dissertation is meant to help answer and ease them. Piece-meal-behavioral engineering is not only a principle underpinning such practices, but hopefully also the best strategy to realize them.

In a world where human beings are inevitably influenced, shaped, mediated, nudged, or reinforced by technical artifacts, it becomes no longer possible to draw a clear boundary between humanity and technology. The interweaving between humans and technologies requires us to come up with a new framework, configuration, or arrangement to accommodate every intertwined human-technology complex. This is where the environmental crisis becomes the entry point to start an exploration. After all, the crisis is not only about the natural environment, but also to the current civilization, political system, concept of freedom and dignity, and image of humans. The assumptions we have

\(^{35}\) To be sure, I am not the only one who notices the potential of Mouffe’s work for technology studies. Delvenne & Parotte (2019) also argue, by referring to Mouffe, that Technology Assessment is in fact political rather than neutral and therefore should be regarded and practiced as political actions. What they suggest probably can be called “politicalizing technology assessment.” It is not hard to imagine that when a technology-in-design is politically assessed, the artifact favored and approved by the assessment will entail politics. In other words, the strategy they have suggested can be a part of the approach of politicalizing technology.
been relying upon are becoming more and more unreliable. We are in need of something new.

At this point, let me quote Skinner to end the discussion:

It is not enough to “use technology with a deeper understanding of human issues”, or to “dedicate technology to man’s spiritual needs”, or to “encourage technologists to look at human problems”. Such expressions imply that where human behavior begins, technology stops, and that we must carry on, as we have in the past, with what we have learned from personal experience or from those collections of personal experiences called history, or with the distillations of experience to be found in folk wisdom and practical rules of thumb. These have been available for centuries, and all we have to show for them is the state of the world today. (Skinner 1971, 10)

Embarking on a journey of transformation is by no means easy. The present dissertation can be a steppingstone for it.


Evans, Nicholas, Stephanie Eickers, Leonie Geene, Marijana Todorovic, and Annika Villmow. 2017. “Green Nudging: A Discussion and Preliminary Evaluation of Nudging as an Environmental Policy Instrument.” Berlin, Germany:
Forschungszentrum für Umweltpolitik, Freie Universität Berlin.


Design for Green


Summary

While technology can be designed to improve people’s behavior, it always comes with the worry of technological intervention to human freedom and autonomy. This dissertation, taking the environmental crisis as an entry point, argues for behavior-steering technology as an ethically acceptable and politically feasible approach to improve people’s behavior. Such an approach develops through four research questions: 1) Why is behavior-steering technology necessary for mitigating environmental problems? 2) What kind of behavior-steering technology is most helpful and why is that? 3) How can we respond to ethical concerns about the design and implementation of behavior-steering technology? 4) What would a political framework that can accommodate the practice of behavior-steering technology look like? The following eight chapters are devoted to answer these questions.

In Chapter 2, I argue that traditional strategies, represented by environmental education, to induce behavioral change tend to fail because they rely on the assumption that human knowing is the initiator of human doing, and neglect the fact that human behavior has to be supported or can be produced by the built environment. Such inattention to the knowing-doing gap reflects, and results from, the asymmetric treatment to anthropocentrism. While knowing-anthropocentrism has been overcome by environmental ethics, doing-anthropocentrism is nearly untouched. Given that in the Anthropocene humans unavoidably take the position of master, non-doing-anthropocentrism however cannot be reached by leaving nature out there free from any interference. A more realistic strategy to distance ourselves from doing-anthropocentrism is to employ technology to improve nature’s, as well as humans’, environmental behavior. Therefore, to tackle the environmental crisis, a focal shift from human knowing to doing is necessary, and the knowing-doing gap has to be bridged or bypassed by behavior-steering technologies.

To take advantage of behavior-steering technologies, Chapter 3 introduces three existing classifications concerning the relations between humans and technologies, and suggests a simpler but useful classification, which distinguishes two types of behavior-steering technology by their differences in mechanism. While IBSTs (informational behavior-steering technologies) take intrasomatic routes, mobilizing users’ consciousness or unconsciousness to
generate behavior, MBSTs (material behavior-steering technologies) take extrasomatic routes, targeting users’ bodies as a locus to apply material conditions or constraints. The pros and cons of each type are clear: Qualitatively and quantitatively, IBSTs are less forceful and therefore regarded as soft and acceptable, but MBSTs are, in contrast, very forceful and tend to be regarded not just as effective but also as hard and ethically problematic. Its high effectiveness supports MBSTs’ potential to improve people’s environmental behavior in a collective way; it has been the main reason to set limits to the design, implementation, and development of MBSTs.

In the beginning of Chapter 4, I point out that the main worry concerning behavior-steering technologies is the loss of human freedom. To address the concerns, ethical guidelines and defensive arguments have been made for IBSTs. For both persuasive technology and nudge, the central principle is to avoid intervening with users’ ends, either by targeting their means only or by minimizing the cost of opting out. Unfortunately, such a principle cannot be applied to the case of the environmental crisis; as it requires collective changes in behavior, we have difficulties in aligning user’s ends with collective goals, because, due to the temporal and spatial dispersion, the causal relationship between one’s behavior and its environmental consequences is hard to identify. As a result, to significantly improve people’s environmental behavior, intervening users’ ends is however unavoidable. Such an ends-intervention is where MBSTs are competent. To be sure, ends-interventions can be done with personalized IBSTs, but such IBSTs are likely to make things worse due to privacy and transparency issues. The concreteness and tangibility of MBSTs make the approach of MBST the best option. However, these features also increase the cost of opting out and therefore provoke worry. In other words, while green MBSTs are indispensable in tackling the environmental crisis, ethical concerns about users’ freedom tend to rise against the design and its implementation. This is a dilemma that urges us to further understand the nature of human behavior.

The work of the psychologist B. F. Skinner helps us to take the next step. Chapter 5 firstly introduces Skinner’s theory of behavior and also the idea of behavioral engineering by referring to his novel *Walden Two*. As human behavior is shaped through the processes of contingencies of reinforcement, it is the product of its environment. This implies the possibility to modify a person’s behavior by changing the technologies s/he interacts with. The technique recommended by Skinner is positive reinforcement, not negative reinforcement, nor punishment. Skinner’s *Beyond Freedom and Dignity*, then, offers various arguments
against the objections to behavioral modification. It shows us that the concepts of freedom and dignity are misguided not only because they too ambiguous in a scientific sense to be reliable terms to understand human behavior, but also because they seriously constrain the use and development of the technology of behavior and therefore leave humans to traditional punitive control.

With the help of Skinner, I argue finally on the one hand that behavior-steering technologies are a must for modifying people’s environmental behavior because the consequences of environmental problems are too remote to have shaping force, and on the other hand that the worry concerning human freedom can be largely reduced by employing the technique of positive reinforcement. However, Skinner still left two questions for us to solve. First, given that sustainability cannot be realized by creating pro-environmental behavior alone, which means the measures of aversive control are unavoidable to apply, what can we do about the issue of human freedom? Second, in taking an evolutionary perspective on values, Skinner helps very little to justify that being pro-environmental is good and therefore should be the direction in which people are going to be steered. That is to say, what goal should we design if being pro-environmental cannot be well defended?

The answers lie implicitly in Skinner’s experimental approach to behavioral engineering. By looking into two real-world Skinnerian communities—Twin Oaks and Las Horcones—in Chapter 6, I first strengthen the argument that behavioral modification rather than other strategies is the key to social change and that attaching a scientific view to human behavior is crucial to make such a change come true. Moreover, these two communities, together with Walden Two, reveal the most important dimension of Skinner’s behavioral engineering: small-scale experimentation. Scaling down to a manageable size not only makes the effects of behavior-steering technologies easier to transfer from human-technology relations to human-human ones, but also reduces the difficulty in testing, adjusting, or revoking the technology-in-design, by which unexpected side effects can be minimized. At this point, I then argue that Skinner’s behavioral engineering resonates with Karl Popper’s idea of piecemeal engineering. By allying with Popper, the two questions left in the previous chapter can be answered. As knowing what is good or right is epistemologically impossible, there is no need to justify being pro-environmental and design for it; rather, our task is to identify anti-environmental behavior and design to remove or decrease it. Moreover, without forcing everyone in one direction, we can make all other options enjoyable, pleasant, or satisfying. By doing so, the net consequences of the
steered behavior can still remain positive; therefore, the concern about human freedom can still be largely reduced, even though positive reinforcement is not the only technique being employed. In short, piecemeal-behavioral engineering, a combination of Skinner and Popper, can underpin the approach of behavior-steering technology and make the approach not only adoptable but also acceptable for tackling the environmental crisis.

As a problem-oriented research, in Chapter 7, I turn the concepts, perspectives, and arguments developed through previous chapters into a list of seven design recommendations. To concentrate these recommendations into one sentence: Design and experiment with green MBST at the level of community to reduce anti-environmental behavior. With the example of Village Homes, I illustrate the usefulness, practicality, and soundness of the approach I have been constructing and defending for green behavior-steering technologies. The design of narrow-curvy-cul-de-sacs beautifully demonstrates the advantages and strengths of MBSTs. It not only makes driving less preferred for the residents as well as visitors, but also creates a beneficial condition for the remaining nature-friendly designs to take effect. Without such an MBST, the achievement of the community would never be reached, and the ethos of being green could never be realized. MBSTs, again, are proved to be a must in the advocacy of sustainability.

For readers who would like to take action immediately and apply the approach straightforwardly, it is not necessary to engage with the political implication of the design and implementation of green behavior-steering technologies. The approach developed through Chapter 2 to 7 is compatible with our liberal democracy. However, it might be worth taking on the political challenge posed by the environmental crisis. In Chapter 8, the discussion moves to the level of politics, in arguing for the need of an alternative approach to democracy. In prioritizing the individual instead of the collective, liberal democracy appears to be hopeless in solving global-scale environmental problems. However, turning to a communitarian democracy does not help, because this is also based on a Kantian, rationalistic image of the human being. From a Skinnerian perspective, this image is problematic, and hence it cannot serve as a basis for democracy. I then argue that a potential candidate for replacing the current models of democracy is agonistic democracy by the political theorist Chantal Mouffe. Grounded in “the political”, agonistic democracy is not about consensus, but in favor of confrontations for democracy to flourish and vitalize. Moreover, it does not require humans to be rational, and therefore allows for external influences. In an agonistic
democracy, both the design and the implementation of behavior-steering technologies are counter-hegemonic practice, by which they justify themselves. Moral questions like whether being green is good or wrong are suspended and the answers to them are not a precondition for technological interventions. In such a way, agonistic democracy probably is the most proper political framework to accommodate the inevitable interweaving of humans and technologies.

To conclude, Chapter 9 puts this research into a broader context, discussing its implications for three closely related fields and practices. For sustainability advocacy, the concept of “educational environment” is proposed to be a complement to environmental education, by which people can be “taught” by their surroundings and practice environment-related skills. For design practice in general, the strategy “design to discourage” is suggested to emphasize the importance of removing problematic behavior rather than creating beneficial one; discouragement is more fundamental than encouragement in shaping people’s behavior to address societal problems. For STS, I make a proposal called “politicalizing technology” as opposed to the idea of democratizing technology; while technology’s politics is however inevitable (in contrast to Langdon Winner), the question of how to make good use of such politics, not that of how to de-politicize technology, is more relevant and beneficial to democracy.
Samenvatting


In Hoofdstuk 2 beargumenteert ik dat traditionele strategieën om gedragsverandering te induceren via milieueducatie vaak falen, omdat die ervan uitgaan dat gedragsverandering ontstaat via bewustwording en veronachtzamen dat deze verandering ondersteund moet worden en ook voortgebracht kan worden door de fysieke omgeving. Dit miskennen van de kloof tussen ‘weten’ en ‘doen’ hangt nauw samen met de asymmetrische manier waarop het antropocentrisme behandeld wordt. Hoewel de milieuethiek het anthropocentrisme beoogt te overwinnen op theoretisch niveau (het ‘weten’), is het in de praktijk vrijwel onaangetast gebleven (‘het doen’). Gegeven het feit dat de mens in het anthropoceen onvermijdelijk de positie van ‘heerser’ inneemt, kan het anthropocentrisme in de praktijk alleen maar overwonnen worden door ook de natuur en de omgeving van de mens in de discussie te betrekken. Een realistische strategie daartoe is het gebruik van technologie om het ecologisch gedrag van mens en natuur in een wenselijke richting te beïnvloeden. Het aanpakken van de milieucrisis vergt een verschuiving in focus van ‘weten’ naar ‘doen’, en de kloof tussen beide kan overbrugd worden door gedragsbeïnvloedende technologie.

Om gedragsbeïnvloedende technologieën te kunnen benutten introduceert Hoofdstuk 3 een aantal bestaande classificaties van de relaties tussen mensen en gedragsbeïnvloedende technologieën, om vervolgens een simpelere, meer
bruikbare classificatie voor te stellen. Deze classificatie onderscheidt twee typen van gedragsbeïnvloedende technologie op basis van hun werkingsmechanismen. Terwijl IBST’s (informational behavior-steering technologies, oftewel informationeel-gedragsbeïnvloedende technologieën) een ‘intrasmatische’ route nemen, die het bewustzijn of onderbewustzijn van gebruikers mobiliseert om gedrag te beïnvloeden, nemen MBST’s (material behavior-steering technologies, oftewel materieel-gedragsbeïnvloedende technologieën) een ‘extrasmatische’ route, waarbij de lichamen van gebruikers het aangrijpingspunt zijn voor materiële condities en beperkingen. De voor- en nadelen van elk type zijn duidelijk: kwalitatief en kwantitatief zijn IBST’s minder ingrijpend en worden daarom beschouwd als zacht en acceptabel, hoewel minder effectief, terwijl MBST’s zeer ingrijpend kunnen zijn, en daarom vaak als hard worden gezien en niet alleen effectief maar ook ethisch problematisch zijn. De hoge effectiviteit ervan onderstreep het potentieel van MBST’s om het ecologisch gedrag van mensen collectief te verbeteren; tegelijk is het ingrijpende karakter ervan de voornaamste reden om grenzen te stellen aan het ontwerp, de implementatie en de ontwikkeling van MBST’s.

In het begin van Hoofdstuk 4 wijs ik erop dat de grootste zorg ten aanzien van gedragsturende technologieën het verlies van menselijke vrijheid is. Om deze zorgen weg te nemen zijn er ethische richtlijnen en argumenten ontwikkeld ter ondersteuning van IBST’s. Het centrale principe, voor zowel persuasieve technologie als ‘nudging’ is het vermijden van interventie in de doelen van gebruikers, door ofwel uitsluitend gericht te zijn op de middelen ofwel het eenvoudig te maken om een opt-out te kiezen. Helaas is een dergelijk principe niet van toepassing in het geval van de milieucrisis. Omdat daar een collectieve gedragsverandering noodzakelijk is, zouden de individuele doelen van gebruikers van technologie op een lijn moeten zijn met het collectieve doel, en dat is problematisch vanwege de grote afstand in ruimte en tijd tussen iemands gedrag en de uiteindelijke ecologische gevolgen. Ten gevolge hiervan is het onvermijdelijk, als we het ecologische gedrag van mensen significant willen beïnvloeden, om te interveniëren in de doelen van gebruikers.

Voor dergelijke interventies zijn MBST’s het best toegerust. Interventie in doelen kan in principe ook gerealiseerd worden middels met behulp van gepersonaliseerde IBST’s, maar dit zou de situatie niet verbeteren, omdat er dan problemen ontstaan ten aanzien van privacy en transparantie. De concreetheid en tastbaarheid van MBST’s maakt de benadering van MBST de beste optie. De kenmerken van MBST’s maken een opt-out echter ingewikkelder en dat roept
Samenvatting

zorgen op. Met andere woorden: hoewel groene MBST’s onontbeerlijk zijn in het aanpakken van de milieucrisis, leiden ethische zorgen over de vrijheid van gebruikers tot bezwaren tegen het ontwerpen en implementeren van dergelijke technologieën. Dit dilemma noopt ons tot meer onderzoek naar de aard van menselijk gedrag.


Met behulp van Skinner beargumenteer ik enerzijds dat gedragbeïnvloedende technologieën noodzakelijk zijn voor het veranderen van het ecologisch gedrag van mensen, omdat de gevolgen van milieuproblemen te ver van hen af staan om invloed te hebben op hun gedrag, en anderzijds dat de zorgen ten aanzien van de menselijke vrijheid grotendeels kunnen worden weggenomen door gebruik te maken van positieve bekraftiging. Skinner laat echter nog twee onopgeloste vragen voor ons over. Ten eerste is er de vraag hoe om te gaan met de menselijke vrijheid wanneer duurzaamheid niet gerealiseerd kan worden door het creëren van pro-ecologisch gedrag alleen, zodat het toepassen van onaangename vormen van beïnvloeding onvermijdelijk is. Ten tweede biedt het werk van Skinner, vanuit een evolutionair perspectief op waarden, weinig basis om te beargumenteren dat ‘pro-ecologisch’ gelijk staat aan ‘goed’, en de richting kan aangeven voor het sturen van menselijk gedrag. Voor welk doel zouden we moeten ontwerpen als het niet per definitie verdedigbaar is om ‘pro-ecologisch’ te zijn?

Het antwoord ligt besloten in Skinners experimentele benadering van gedragstechniek. Door twee reëel bestaande Skinnerianse gemeenschappen te
Design for Green

bestuderen—Twin Oaks en Las Horcones—ontwikkeld ik in Hoofdstuk 6 allereerst het argument dat gedragsmodificatie, in tegenstelling tot andere strategieën, de sleutel vormt tot sociale verandering en dat een wetenschappelijke benadering van menselijk gedrag cruciaal is om een dergelijke verandering te realiseren. Bovendien laten deze twee gemeenschappen, samen met Walden Two, zien wat volgens Skinner de belangrijkste dimensie van de gedragstechniek is: kleinschalige experimenten. Schaalverkleining tot een beheersbaar formaat maakt het niet alleen gemakkelijker om de effecten van gedragsbeïnvloedende technologieën te vertalen van mens-technologie relaties naar mens-mens relaties, maar ook om de technologie-in-ontwerp te testen, aan te passen of te herroepen, waardoor onverwachte neveneffecten geminimaliseerd kunnen worden. Ik beargumenteer dat de gedragstechniek van Skinner in dit opzicht in lijn is met Karl Poppers idee van ‘piecemeal engineering’.

Door deze verbinding met Popper kunnen de twee vragen beantwoord worden die in het vorige hoofdstuk overbleven. Net zoals het epistemologisch gezien onmogelijk is om te weten wat goed en juist is, is het niet nodig om te zoeken naar een rechtvaardiging van pro-ecologisch gedrag om daarvoor te kunnen ontwerpen. Het is juist zaak om anti-ecologisch gedrag te identificeren en het ontwerp te richten op het verminderen of wegnemen ervan. Bovendien is het mogelijk om alle andere opties aangenaam, plezierig of bevredigend te maken zonder iedereen in één en dezelfde richting te dwingen. Op deze manier kan het nettoresultaat van het beïnvloede gedrag nog steeds positief blijven, terwijl de ongerustheid over de menselijke vrijheid nog steeds grotendeels weggenomen kan worden, hoewel positieve bekrachtiging niet de enige gebruikte techniek is. Kortom, ‘piecemeal’ gedragstechniek, die Skinner en Popper combineert, kan de benadering van gedragsbeïnvloedende technologie onderbouwen en de benadering niet alleen aannemelijk maken, maar ook aanvaardbaar voor het aanpakken van de milieucrisis.

Vanuit mijn probleemgeoriënteerde benadering vertaal ik in Hoofdstuk 7 de concepten, perspectieven en argumenten die ik in de voorgaande hoofdstukken heb ontwikkeld in een lijst van zeven ontwerpaanbevelingen. Om deze aanbevelingen in één richtlijn samen te vatten: ontwerp en experimenteer met groene MBST op gemeenschapsniveau om anti-ecologisch gedrag te reduceren. Ik illustreer de bruikbaarheid, uitvoerbaarheid en deugdelijkheid van deze benadering van groene gedragsbeïnvloedende technologieën met het voorbeeld van Village Homes. Het ontwerpen van nauwe, bochtige doodlopende straatjes.
Samenvatting
demonstreert op fraaie wijze de voordelen en sterke punten van MBST's. Niet alleen maakt het autorijden minder aantrekkelijk voor zowel bewoners als bezoekers, maar ook creëert het gunstige voorwaarden voor de effectiviteit van overige milieuvriendelijke ontwerpen. Zonder dergelijke MBST's zou deze gemeenschap dit resultaat nooit bereikt hebben en had het groene ethos nooit gerealiseerd kunnen worden. MBST's blijken, opnieuw, noodzakelijk te zijn om het duurzaamheid te bevorderen.

Voor lezers die onmiddellijk actie willen ondernemen en de benadering direct willen toepassen, is het niet per definitie nodig om zich bezig te houden met de politieke implicaties van het ontwerp en de implementatie van groene gedragsbeïnvloedende technologieën. De benadering die in Hoofdstuk 2 tot en met 7 is ontwikkeld, is verenigbaar met onze liberale democratie. Het is echter wel degelijk de moeite waard om de politieke uitdaging aan te gaan waar de milieucrisis ons voor plaatst. In Hoofdstuk 8 verplaatst de discussie zich daarom naar het niveau van de politiek, door te pleiten voor een alternatieve benadering van democratie. Door het individu boven het collectief te plaatsen, lijkt de liberale democratie schijnt niet in staat te zijn om de wereldwijde ecologische problemen op te lossen. Een wending naar een communautaire democratie helpt niet, omdat die eveneens gebaseerd is op een Kantiaans, rationalistisch mensbeeld. Vanuit Skinneriaans perspectief is dit mensbeeld niet adequaat, en daarom kan het niet als basis voor democratie dienen.

Ik beargumenteer dat een potentiële kandidaat voor een ander model de agonistische democratie is, zoals voorgesteld in de politieke theorie van Chantal Mouffe. Agonistische democratie heeft zijn grondslag in “het politieke” en draait daarom niet om consensus, maar pleit voor confrontaties van die democratie doen floreren en vitaliseren. Bovendien vereist het niet dat mensen altijd rationeel zijn en biedt het ruimte voor externe invloeden. In een agonistische democratie zijn zowel het ontwerp als de implementatie van gedragsbeïnvloedende technologieën contra-hegemonische praktijken, waarmee zij zichzelf rechtvaardigen. Morele vragen, zoals of het groen zijn goed of fout is, worden opgeschort en de antwoorden erop zijn geen voorwaarde voor technologische interventie. Op deze manier is agonistische democratie waarschijnlijk het meest passende politieke raamwerk om de onvermijdelijke vervlechting van mensen en technologieën te accommoderen.

Als conclusie plaatst Hoofdstuk 9 dit onderzoek in een bredere context, door in te gaan op de implicaties ervan voor drie nauw gerelateerde velden en praktijken. Voor duurzaamheidsbevordering stel ik het concept van een
“educatieve omgeving” voor als aanvulling op milieueducatie: mensen kunnen “onderwezen” worden door hun omgeving in het (uit)oefenen van milieugerelateerde vaardigheden. Voor de praktijk van ontwerpen in het algemeen stel ik de strategie van “ontmoedigend ontwerp” voor, om het belang te benadrukken van het wegnemen van problematisch gedrag, in plaats van het oproepen van bevorderlijk gedrag. Ontmoediging is belangrijker dan aanmoediging in het beïnvloeden van menselijk gedrag om maatschappelijke problemen aan te pakken. Voor STS stel ik het “politiseren van technologie” voor, tegenover het idee van het “democratiseren van technologie”. Omdat de politiek van technologie onvermijdelijk is, is de belangrijkste en meest bevorderlijke vraag voor de democratie niet (in tegenstelling tot Langdon Winner) hoe technologie gedemocratiseerd kan worden, maar hoe we goed met haar politieke lading om kunnen gaan.
About the Author

Ching Hung (1982) was born in Chiayi, Taiwan. He completed his PhD in Philosophy of Technology at the University of Twente, the Netherlands, between September 2013 and November 2019. Since 2014, he has been a columnist for several digital and print magazines in Taiwan, writing to bring the insights of philosophy of technology and STS (science and technology studies) to the public. Ching’s background is truly interdisciplinary. He holds a Bachelor of Science in Nuclear Science and a Master of Arts in History, both from National Tsing Hua University, Taiwan. In the same university, he also received one-year doctoral training at the Institute of Sociology, where enrolled him with a scholarship. Before his PhD study, he had worked as a fulltime research and teaching assistant at the Institute of Science, Technology and Society, National Yang-Ming University. His current research is about rethinking and relocating the position of human beings in their relations with technology and nature, through which he believes the present political, economic, educational, and social systems can be largely improved.
Simon Stevin Series in Ethics of Technology

Delft University of Technology, Eindhoven University of Technology, University of Twente & Wageningen University

Editors: Philip Brey, Anthonie Meijers and Sabine Roeser

Books and Dissertations
Volume 1: Lotte Asveld, Respect for Autonomy and Technology Risks, 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 6: Daan Schuurbers, Social Responsibility in Research Practice: Engaging Applied Scientists with the Socio-Ethical Context of Their Work, 2010
Volume 8: Ilse Oosterlaken, Taking a Capability Approach to Technology and Its Design: A Philosophical Exploration, 2013


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 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
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.
We are faced with a very serious environmental crisis. Resource depletion, air pollution, biodiversity loss, climate change, etc., accompany each another, resulting largely from our current way of life. Manufacturing of nature-friendly products is far from being able to protect the environment; technologies can and should be made to change people’s environmental behavior. However, designing and implementing such technologies tend to provoke ethical and political concerns about human freedom and autonomy. Determining how to take advantage of behavior-steering technology without causing worries, therefore, is unavoidably a key challenge to practitioners as well as theorists. This dissertation takes on this challenge by synthesizing philosophical insights about human-technology interactions, psychological explanations for the nature of human behavior, and political theories concerning democracy and social reform. It not only offers a conceptual basis in arguing for the necessity and feasibility of behavior-steering technology, but also develops a piecemeal-behavioral approach to realizing this goal. Through the chapters of this dissertation, it becomes clear that, above all, to address the environmental crisis eventually requires us to conceive of a more realistic image of human beings and to rebuild contemporary ethical and political frameworks based upon it.