Though humans are ‘the tool-using animals’, philosophical action theory has rarely investigated the role artefacts play in the action process. This book investigates this role, driven by its research question: What are actions with artefacts, and how do artefacts influence agents and agency?

*Acting with Artefacts* argues against Davidson’s claim that all actions are bodily movements, and uses the psychology of action to show that there can be actions with artefacts. The resulting picture of actions is used to clarify the concept of ‘affordance’ in design. Furthermore, this book argues that artefacts do not just enable or facilitate actions; they can also alter our reasons for action, and more radically, alter us in our capacity as bounded practical reasoners. Finally, this book shows how artefacts can support our commitment to collective actions, and how engineers can transfer responsibility for artefact use to users.
Acting with Artefacts

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. C.J. van Duijn, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op dinsdag 20 september 2011 om 16.00 uur

door

Auke Jan Koop Pols

geboren te Assen
Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. A.W.M. Meijers

Copromotoren:
dr. W.N. Houkes
en
dr. P.J. Nickel
Samenstelling promotiecommissie

Rector Magnificus, voorzitter
prof.dr.ir. A.W.M. Meijers, Technische Universiteit Eindhoven, promotor
dr. W.N. Houkes, Technische Universiteit Eindhoven, co-promotor
dr. P.J. Nickel, Technische Universiteit Eindhoven, co-promotor
dr. J. Anderson, Universiteit Utrecht
prof. J. Hornsby, Birkbeck, University of London
prof.dr. C.J.H. Midden, Technische Universiteit Eindhoven
dr. M.M.S.K. Sie, Erasmus Universiteit Rotterdam
I asked, “Which was your last war?” He said, “Cutting down that tree was my last war!” I asked him who won, which I thought was a nice question, because it would let him say that he won, and feel proud. He said, “The ax won! It’s always that way!”

- Jonathan Safran Foer, Extremely Loud and Incredibly Close
# Contents

1 Introduction

1.1. General introduction  
1.2. Why investigate actions with artefacts?  
1.3. Goals and research questions of the dissertation  
1.4. Concepts and methodology  
1.5. Structure of the dissertation  

2 Basic actions and the time of a killing

2.1. Chapter abstract  
2.2. Introduction  
2.3. The problem of the time of a killing  
2.4. The problem of the acting dead  
2.5. Killing and causing to die  
2.6. Conclusion  

3 Can there be basic actions with artefacts?

3.1. Chapter abstract  
3.2. Introduction  
3.3. The ambiguous account of basic actions  
3.4. Causally basic actions  
3.5. Teleologically basic actions  
3.6. Basic actions with artefacts  
3.7. Conclusion  

4 Describing affordances: the nested affordances model

4.1. Chapter abstract  
4.2. Introduction  
4.3. Affordances under a description  
4.4. The nested affordances model  
4.5. Applying the nested affordances model  
4.6. Acting with artefacts on affordances  
4.7. Conclusion
5 Artefacts and reasons for action

5.1. Chapter abstract 77
5.2. Introduction 78
5.3. Practical reasons 81
5.4. Artefacts as enabling actions 84
5.5. Artefacts as facilitating actions 85
5.6. Artefacts as prescribing actions 86
5.7. Artefacts as inviting actions 91
5.8. Two counterarguments 96
5.9. Could artefacts have agency? 98
5.10. Conclusion 102

6 How enhancement technologies affect us as bounded practical reasoners

6.1. Chapter abstract 103
6.2. Introduction 104
6.3. The nature of the debate 104
6.4. Ends and instruments 106
6.5. Enabling and facilitating actions 108
6.6. Ending ends 109
6.7. Transcending our bounds 110
6.8. Conclusion 113

7 Robust joint commitment and the use of strategies

7.1. Chapter abstract 115
7.2. Introduction 116
7.3. Joint commitment 118
7.4. Self-control and strategies 124
7.5. Making a robust joint commitment 130
7.6. Robust joint commitment and human-induced climate change 133
7.7. Conclusion 138

8 Transferring responsibility for artefacts through use plans

8.1. Chapter abstract 141
8.2. Introduction 142
8.3. Responsibility and control 145
8.4. Use plans 149
8.5. Combining approaches 151
8.6. The Abcoude dosing lock: a test case 157
8.7. Conclusion 159

9 Conclusion 161
Acknowledgements

The first day I came to work in Eindhoven, I brought a mug with me. It is a big ceramic mug, with a picture of a folder on it. The folder has hands and feet, and a smiling face. The mug was my first Christmas package, from a job where I had to enter names and social security numbers from folders into a computer program, four hours per day, five days per week. Talking, laughing or leaving the building during a break had not been permitted. I had brought the mug with me as a reminder. If I ever despair of my project, I had thought, if I ever consider giving up, I will look at this mug. I will remind myself that the alternative is typing a hundred thousand social security numbers again, and then I will throw myself at my dissertation.

I never had to use my mug for this purpose, though I did muse over it several times, such as the day that Jonathan ‘the Scalpel’ Dancy first dissected one of these chapters with excruciating precision. This was not because writing this dissertation was easy, because it wasn’t. But it has been an interesting and good project, and now I wish to thank the various people who have contributed to it. It is only because of them that I never really had to use my mug for its intended purpose.

First of all, I would like to thank my project supervisor, Anthonie Meijers, for always being involved in the project and available for questions and good advice, despite seeming to be in a perpetual hurry. Then there was my daily supervisor Wybo Houkes, with whom I discussed all the details of my project and the finer art of getting papers published in good journals, which goes above and beyond writing good papers. Philip Nickel joined in later and was indispensable in pointing out vague formulations and loose claims and helping me to tighten them. Joel Anderson has been a thorough reader throughout the project and kept pressing me to explicate why what I was doing was important and a useful contribution to the field. My thanks also go out to the two members of my Eindhoven supervising committee who left earlier in the project, Marc de Vries for being a quick and diligent reader, and Christian Illies for his uncanny ability to point out the sore spots in every presentation, as well as arranging a stay for
me at a summer school at the University of Notre Dame in the U.S.A. I also
would like to thank the other members of my dissertation committee for reading
and commenting on the final draft: Jennifer Hornsby, Cees Midden and
Maureen Sie.

Dissertation writing can be a lonely business, but I had the good fortune to have
a lot of great colleagues to work with. While I would like to thank all of them for
their company, special thanks go out to the following persons: Andreas Spahn,
for being a good friend and teaching me everything I did and didn’t want to
know about Aristotle and Kant. Felicitas Kraemer for the discussions on medical
ethics, and the cashew nuts. Joel Katzav for teaching me not to be too harsh on
students. My roommates Bart van Bezooijen, Christine van Burken and Anya
Topolski for a chat in between writing. And of course, my fellow PhD students:
Krist and Melissa who preceded me and taught by example, and those who came
after me, Bas, Cleitus, Jilles, Litska, Marieke, Maryam, Stefan and Sven. And of
course our secretary Rianne Schaaf for being practical where I was often less so.

Outside the office, I am grateful to Jonathan Dancy and Bart Streumer for
helping me organising my stay in Reading and advising me, especially on
chapter five. For commenting on earlier drafts of these chapters I am grateful to
almost all of the people mentioned above, as well as to Neil Levy, Jessica Nihlén-
Fahlquist and the members of the NSRPP methodology and meta-ethics group,
as well as to a host of anonymous referees, most of whom provided me with
excellent comments.

One does not live by work alone: I am thankful to Frank, Luc, Lukas, Morten,
Patrick and Rens for our relaxing irregular gaming sessions, and to Donijs,
Marieke, Mariska and Mascha for our relaxing regular gaming sessions at the
beginning of my project. I thank my family, especially Jan and Jeannette, and my
family-in-law for some nice discussions on my dissertation topics. Finally, I am
grateful to Leendert and Thijmen for being two of the greatest sons a father
could wish for, and to Marijke, whom this dissertation owes more than mere
words can express.
1 Introduction

1.1. General introduction

In 1956, Oxford University proposed to grant an honourary degree to former U.S. president Harry Truman. Most staff members were in favour of the decision. However, the proposal was fiercely challenged by philosopher and Oxford research fellow Elizabeth Anscombe (1981). Anscombe argued that Truman was a mass murderer since he had ordered the bombing of Hiroshima and Nagasaki, and should therefore not be given the honourary degree. In defense of the proposal, the Censor of St. Catherine’s College spoke to the effect that: “...Mr. Truman did not make the bombs by himself, and decide to drop them without consulting anybody; no, he was only responsible for the decision. Hang it all, you can’t make a man responsible just because ‘his is the signature at the foot of the order.’” (ibid., 66).

Truman was awarded the honourary degree, and the question remained open whether Truman’s signing of the order was actually his committing mass murder. A year later, however, Anscombe published her monograph Intention. Intention laid the foundations for modern analytic action theory, and was called ‘the most important work on human action since Aristotle’ by Donald Davidson, who would go on to define the field further in a series of ground-breaking essays (1980). While Anscombe did not address the mass murder question explicitly, it is implicitly resolved by the book: since Truman did not have to do anything else than sign the order to get Hiroshima and Nagasaki bombed, his action of signing the order was his action of murdering thousands of innocent people, and therefore he could reasonably be considered a mass murderer. As Anscombe had already dryly noted about the speech of the Censor of St. Catherine’s: “The defence, I think, would not have been well received at Nuremberg.” (1981, 65).
1.2. Why investigate actions with artefacts?

*Intention* has answered many questions, but raised many more. How do intentions lead to actions? Are known side effects of actions brought about intentionally? Would it have made a difference if Truman *had* made the bombs by himself? Many of those questions have been addressed since; some have sparked great controversy and led to new debates and new theories of human action (see Mele: 1997 for a good overview).

The question about the bombs is not one of those questions. Indeed, the role of technical artefacts in the action process is rarely, if ever, explicitly examined. This is so despite the fact that a technical artefact figured prominently in *Intention*’s paradigm example, a man using a water pump to pump up poisoned water, thereby replenishing the water supply of a house and poisoning its inhabitants. Since then, many more important claims in action theory have been built on thought experiments involving artefact use without investigating the role the artefact has played in the action, like Davidson’s (1980, essay 3) man who startles a burglar by illuminating a room by flipping the light switch, or Bratman’s (1984) video game player who intends to hit his target with either one of two missiles, but not with both at the same time.

One of the main reasons why action theorists have not examined the role of technical artefacts in actions in any detail might be Davidson’s widely accepted claim that all actions are ‘mere movements of the body’ (1980, essay 3, 59). This claim effectively places artefacts outside actions and makes them part of the external world, along with natural objects, symbols, etc. Whether this claim is valid or not, it has drawn attention away from an interesting and important topic. Technical artefacts are ubiquitous and greatly enlarge the scope of what we can bring about. Analysing how Truman’s intentions led to his moving his hand to sign the form, or what his reasons were to do so can give us important insights, but of crucial importance is that this signing initiated a complex socio-technical process that made his action both a mass murder and the swift ending of a terrible war. Action theorists have tended not to look at this kind of socio-technical processes and related questions about collective actions, distributed

---

1 I will argue that it is not valid in Chapter 3.
responsibility and how artefacts influence agency. Again, this may be because of Davidson’s claim about bodily movements. But whatever the reason, as research in Science and Technology Studies (e.g. Akrich: 1992; Latour: 1987, 1992), (post)phenomenology (Ihde: 1990; Verbeek: 2000/2005), but also the psychology of tool use (Maravita and Iriki: 2004; Goldenberg and Iriki: 2007) has shown, the interaction of acting humans and technical artefacts is characterized by interesting phenomena such as the influence artefacts exert on both our individual and our collective agency, the sensation of artefacts as extending our bodies in action, the sensation of feeling at the tips of our artefacts and the ease with which we can learn to manipulate them, which seems to be unique to our species. All these phenomena could benefit from analytic exploration.

Moreover, action theory distorts its own views by ignoring actions with artefacts, leading it to misidentify what it considers to be paradigm cases of human actions. In this dissertation, I argue that the thesis that all actions are bodily movements is a good example of this: when a carpenter is hammering a nail into a board, or a skilled tennis player is performing a serve, it seems strange to say that the real action performed here is the downward swing of a grasping hand. Another example where theory could go astray when it does not examine the role artefacts play in actions is the lack of attention to the distribution of responsibility for actions done with artefacts. Analytic theories of responsibility build on action theory and similarly often use artefacts in examples and thought experiments: Fischer and Ravizza (1998) use the example of someone driving a car under the supervision of a driving instructor to illustrate different forms of control over and responsibility for a particular situation. However, they do not mention the very real possibility that the engineer of an artefact might also be responsible for some of the effects of user actions with that artefact, e.g. when neither the student nor the instructor can prevent an accident due to an improperly connected braking system.
Acting with Artefacts

Perhaps for this reason, philosophical action theory is not applied much in non-philosophical fields like design sciences and the psychology of action.\(^2\) This is unfortunate: theories of affordances (‘opportunities for action’) in design sciences and ecological psychology are unnecessarily vague because they lack a good definition of what actions are and how they should be described (Michaels: 2003). Similarly, theories on (basic) actions in cognitive (Humphreys, Forde and Riddoch: 2001) and sports (Schack: 2004) psychology could be further developed using the analysis of basic actions that has been performed in action theory. Yet so far there has been no attempt to do so, and it has even been suggested that psychological research offers evidence against the usefulness of current philosophical notions of basic actions (Eilan and Roessler: 2003).

1.3. Goals and research questions of the dissertation

According to Sneddon (2006), any good theory of action should be able to answer two questions. The first is the status question: what is an action? The importance of this question is obvious: one has to know what one is studying. The second is the production question: how are actions brought about? This question is particularly important for explaining actions, and whether that should be done in terms of causal mechanisms (Davidson: 1980; Mele: 2000) or in terms of the agent’s goals, intentions, etc. (Anscombe: 1957/2000; Sehon: 1997).

As this dissertation seeks to give a proper account of actions with artefacts, it should be able to answer the following, specific versions of those questions: what is an action with an artefact? And how are actions with artefacts brought about? With regard to this last question, I will not assume that artefacts act with us or cause our actions in a strong sense, though their characteristics can causally affect the particular bodily movements we make in initiating manipulation (see Milner and Goodale: 1995). While I return to this last point in chapter 4, my main focus in this dissertation will be on how artefacts influence us as agents

\(^2\) Though action theory has been relevant for other non-philosophical fields, e.g. computer sciences. Some examples where action theory and the psychology of action have fruitfully been combined are Vallacher and Wegner (1987) and Pacherie (2008).
either by providing specific reasons for or against certain actions (chapter 5), or more radically, affecting us in our capacity as bounded practical reasoners (chapter 6). Rephrased, the main question of my dissertation thus becomes: *What is an action with an artefact, and how do artefacts influence agents and agency?*

This dissertation has two main goals. First, most of the dissertation addresses the main question. This will help explore the largely unexplored space in analytic action theory of actions done with artefacts, as well as test traditional claims of action theory and see whether they remain standing once the importance of artefacts for what we do is acknowledged. Second, in addition the dissertation establishes how the topic of actions with artefacts relates to fields both within philosophy (practical reasoning; moral responsibility) and outside philosophy (design sciences; cognitive psychology; neurosciences). This will allow us to create interdisciplinary connections and exchange insights, as well as show how the theoretical research might affect practical issues, e.g. in design. This is reflected in the interdisciplinary nature of the dissertation: rather than constructing a single philosophical theory, it aims to build bridges and improve communication between disciplines, strengthening them all.

In order to achieve these goals, I have written a dissertation in three parts. The first part will concern itself with the first part of the question: what is an action with an artefact? The second part will concern itself with the second part of the question: how do artefacts influence agents and agency? The last part of the dissertation is a practical application of research into actions with artefacts: it shows how an action-theoretical account of artefact use, the use plan analysis, can help us define the conditions under which responsibility for (actions done with) artefacts can be transferred from engineer to user.

Each of the first two sections of the dissertation is divided into a part that focuses on developing a theory and a part that focuses on applying it to illustrate the theory’s practical relevance. For the first part of the dissertation, sections 2 and 3 develop the theory while section 4 applies it. For the second part of the dissertation, sections 5 and 6 develop the theory, and application follows later in section 6 and in section 7. This brings the total number of chapters of this dissertation (introduction and conclusion excluded) to seven. Note that the
Acting with Artefacts

chapters were originally written as papers rather than as book chapters, which means that the connection between them is looser than in a typical book. In particular, the three parts function largely independently. Thus, acceptance of my claims about the ways artefacts influence agents and agency does not hinge on acceptance of my claims about actions with artefacts, for example. The structure of the dissertation is shown in Schema 1.

![Schema 1: global structure of the dissertation](image)

1.4. Concepts and methodology

Before I elaborate on the structure of the dissertation, I will briefly clarify what I do and do not mean by the two most important concepts used in this dissertation, action and artefact, and go into the methodology I have used. This will provide a starting point for my investigations into what actions with artefacts are.

First, the concept of action. As actions, I will regard doings of an agent that are intentional under some description (after Davidson: 1980, essay 3). If I fasten a board to a tree by hammering a nail through the board into the tree, I can be said
to perform an action, as I am doing something that is intentional under at least two descriptions (and probably more): fastening a board to a tree and hammering a nail into the board. That actions can be described in different ways is possible partly because actions (hammering the nail into the board) have effects (that the board is fastened to the tree) that can themselves be intended. If this is the case, it makes sense to describe the action in terms of those effects. In these cases, when an agent does something by doing something else, Anscombe and Davidson hold that only one action is done. This thesis is known under various names: in this dissertation, I will call it the action identification thesis (after Wilson: 2009), and elaborate on it in chapter 2.

A logical next question is: if there is one action under different descriptions, what description then really describes the action? Here, Anscombe, Davidson and I part ways. Anscombe (1979) argues that the question just does not make sense: the only possible answer would be to give another description of the action, but she sees no reason to prefer one description over the other. Davidson (1980) argues that the action is the thing we do directly, not by doing something else, which gives us some indication of what should be described. Particularly, Davidson argues that this means that all actions are (and thus can properly be described in terms of) bodily movements. In chapter 3, I will examine Davidson’s claim more closely. Drawing on Hornsby’s (1980) work on descriptions of actions, I will argue that we can keep the action identification thesis and Davidson’s idea of the action as something we do not do by doing something else, but that this does not commit us to claiming that all actions are (and thus can properly be described in terms of) bodily movements.

This dissertation is not concerned with the question whether ‘intelligent’ artefacts can or will ever be able to properly act by themselves. I see no reason why this would never be possible, though, and look at this issue briefly in chapter 5. What this dissertation will also not look at is whether other non-human beings like animals can act, though I say a few words on this in chapter 4. Neither will I concern myself with ‘mental actions’ like deciding or inferring: my focus is on physical actions. It is debated in the literature whether ‘intentionally doing nothing’ or ‘intentionally not intervening’ can count as an action (see Mossel: 2009). This is an interesting topic, and especially relevant for
actions with artefacts, since many artefacts allow us to do things while doing nothing. For example, while sitting in my comfortable chair I can be said to be ‘doing the laundry’ or ‘baking bread in the oven’, if I have turned on the washing machine or the oven earlier. (Falvey: 2000). As I hold that all actions require activity on the part of the agent (see chapter 2), I do not regard intentionally doing nothing as an action. Of course, it can still be part of a plan (e.g. for artefact use), and people can still be responsible for doing nothing and the effects thereof. However, a proper investigation of this topic is beyond the scope of this dissertation.

As technical artefacts I roughly understand material objects that have intentionally been made or (re)designed in order to (better) fulfil a particular function. This can be as simple as an improvised alteration to a natural object in order to facilitate its use for a particular purpose, e.g. breaking a branch at a certain length to obtain a suitable walking stick, or as complex as constructing a car that consists of many parts that are themselves technical artefacts, and that requires a complex physical and social arrangement of roads, traffic laws, etc. in order to function properly. I mostly deal with ‘ordinary’ artefacts like cars and guns (for some reason favoured examples among analytic philosophers), canes, traffic lights, life buoys, garlic presses and the like, though I make a few excursions specifically to artefacts that serve as extensions of the body (in chapter 3) and human enhancement technologies (in chapter 6).

The methodology used in this dissertation is mostly conceptual analysis, though I draw heavily on studies in the (neuro)psychology of action and tool use and studies in artefact design in order to generate results that are practically applicable. This choice of methodology follows from the way the main question is posed. The first part, ‘what is an action with an artefact?’, is a straightforward conceptual question. The second part, ‘how do artefacts influence agents and

---

3 This definition reminisces Dipert’s (1993) definition of ‘tool’: for Dipert, an ‘artefact’ also has to be intentionally adapted to convey the intention that the object should be used as a tool. See for other definitions of technical artefacts Hilpinen (1993, 2008).
agency?', could be addressed in a number of ways. However, there is a sense in which artefacts like the speed bump or the brick wall in front of us influence us as agents - or mediate our actions (Verbeek: 2000/2005), or prescribe certain actions (Latour: 1992), yet it has never been clearly established what ‘influence’ (or ‘mediation’ or ‘prescription’, for that matter) means here. This dissertation will focus on analysing the particular meaning of ‘influence’, which means that conceptual analysis will be needed here too.

Concerning the combination of philosophical and empirical research, I aim to give an account of actions with artefacts that is empirically informed: it should fit psychological research on what people actually do and why. This does not mean that I accept psychological theories on action uncritically. Indeed, where psychologists can inform philosophers of interesting empirical phenomena, philosophers can help psychologists find the right way to conceptualise these phenomena. Through the use of psychological case studies and philosophical reflection on them, I aim to make this dissertation interesting for both parties.

Besides using psychological studies, I also make regular use of Houkes and Vermaas’s (2004) use plan analysis of artefact use and design (especially in chapter 8). Illies and Meijers’s (2009) action scheme has been especially inspiring for chapters 5 and 6. While I do not engage in structural phenomenological analysis, the (post)phenomenology of artefact use and the Hypothesis of the Extended Mind have been a great inspiration especially for chapters 3 and 5. Besides numerous examples and thought experiments on actions with artefacts, this dissertation also contains a case study in chapter 8, where I will investigate (a failure in) the transfer of responsibility for drivers’ use of a traffic calming measure in the Dutch municipality of Abcoude.

1.5. **Structure of the dissertation**

I will now briefly treat each chapter’s main question and show how it contributes to addressing the overall research question.

In order to create a good starting position, chapter 2, ‘Basic actions and the time of a killing’, develops a ‘standard view’ of what an action with an artefact would be. This standard view is developed according to work by Anscombe and Davidson, particularly the action identification thesis and the corresponding
Acting with Artefacts

notion of basic actions that both philosophers endorse. While this thesis is quite popular in the literature, it has recently been attacked by a number of authors (Mackie: 1997; Mossel: 2001; Weintraub: 2003; Sandis: 2006) because it supposedly is not able to deal adequately with the problem of the time of a killing (Thompson: 1971). Basically, the criticism is that the thesis makes a counterintuitive claim about the time when a killing takes place if time elapses between the fatal wounding and the death of the victim. In order to establish a clear ‘standard view’ on actions with artefacts and give it a stable footing, in chapter 2 I defend the thesis against this criticism and argue that it does not commit us to the counterintuitive results that the aforementioned authors claim it does. More specifically, I ask: Can the action identification thesis deal with the problem of the time of a killing?

Having formulated a standard view of actions with artefacts, in chapter 3, ‘Can there be basic actions with artefacts?’ I proceed to criticise this standard view. Particularly, here I address the relation between artefacts and the human body with regard to actions, and argue against Davidson’s claim that all actions are bodily movements (1980, essay 3). The main question of this chapter is: Can there be basic actions with artefacts? By showing that the relation between artefacts and the human body with regard to action can be much closer than usually presupposed, I argue for the possibility of (basic) actions with artefacts.

Chapter 4, ‘Describing affordances: the nested affordances model’, shows how the action identification thesis can have a practical application in the field of design sciences. It does so by showing that a clear notion of action can unify the concepts of affordance used in ecological psychology (Gibson: 1979) and design (Norman: 1988/2002). Actions can be described in different ways and on different levels, ranging from low-level, context-independent descriptions like ‘moving your hand’ to high-level, context-dependent descriptions like ‘hammering a nail into a board’ or ‘signaling a friend to take her turn at a game.’ Most

While Davidson uses the term ‘primitive action’ and some authors prefer ‘simple action’, the term ‘basic action’ is used more often. As the terms are interchangeable, I will use the term ‘basic action’ throughout this dissertation. Strictly speaking, ‘basic action’ is a misleading term since in Davidson’s theory there are no non-basic actions, only non-basic descriptions of actions, but I will keep to the term as it is used in the literature.
animals can at least perform actions under low-level descriptions, like moving their bodies, while signaling a friend to come over requires abstract knowledge (e.g. of the social meaning of a particular gesture) and particular psychological mechanisms (e.g. for planning and face recognition). I argue that affordances can likewise be described in different ways and on different levels. The main question of this chapter is: On which levels can we describe the affordances of artefacts, and what mechanism and knowledge do we need to perceive what an artefact affords on each level of description? The chapter addresses this question by building a Nested Affordances Model that systematises and clarifies existing design recommendations, as well as offering new ones. Thus, it shows that asking the main question, what is an action with an artefact, is not only interesting for action theory, but also relevant for the more practical field of affordance-based design.

The second part of the dissertation addresses the second part of the main question: How do artefacts influence agents and agency? Technical artefacts guide us, persuade us, and mediate our behaviour, as anyone can attest who has ever forgotten (or did not want to bother) to fasten the seatbelt in a car, but was prompted to do so by the loud warning signal, has slowed down before a speed bump, or has been convinced by an anti-RSI-application to stop typing and take a five-minute break. This influence of artefacts on agents and agency has been the subject of much attention in sociology and Science and Technology Studies (Akrich: 1992; Latour: 1992), and (post)phenomenology (Verbeek: 2000/2005). Simply adopting the frameworks of Latour or Verbeek to address this question would not fit well with the analytic literature, however: much of their work diminishes the difference between agents and artefacts, endowes artefacts with (delegated) intentions, or ascribes intentions and agency to human-artefact systems rather than to human beings. The aforementioned authors hold that technical artefacts influence our agency and perception to such a degree that applying core action-theoretical concepts like ‘intentionality’ and ‘agency’ to human-artefact systems rather than to human beings alone is needed to get an adequate account of this influence (Verbeek: 2009).

While taking a systems rather than a human perspective can indeed help to account for the influence of technical artefacts on agents and agency, there are
several problems with such a move, as I will argue in chapter 3. For example, the boundaries of a system are as hard to establish, if not harder, than those of a human being, and many human-artefact systems are transient, existing only for the short period of time in which the user holds and manipulates the artefact. Moreover, even if taking a systems perspective could give us an adequate account of the influence of artefacts on agents and agency, there have been no attempts to account for this influence without taking a systems perspective. Given the problems of the systems perspective, the challenge is whether we can give a satisfying analytic account that does not suffer (so much) from these problems.\footnote{A recent attempt is Illies and Meijers (2009). Yet while Illies and Meijers explain what the influence of artefacts on agents amounts to in analytic terms, they do not explain how artefacts play this role.} Chapter 5, ‘Artefacts and reasons for action’ will take up this challenge. In particular, it will ask the following question: How can we account for the influence of artefacts on agents in analytic terms, without taking a systems perspective?

I assume that for artefacts to influence agents, they have to somehow influence the mechanism usually thought responsible for bringing about actions in agents: for Anscombe and Davidson, this is reasons \slash practical reasoning.\footnote{I use a light conception of practical reasoning that does not require conscious deliberation, but rather a weighing of options or sometimes just ‘acting for a reason’. (That is, the action can be understood as being done for a reason.) Given this light conception, I assume that practical reasoning precedes every action.} Chapter 5 addresses the main question by showing how artefacts affect our practical reasoning by altering our reasons for action. Unfortunately, Davidson’s conception of reasons as belief-desire pairs is less suited for this task. Desire-based \slash internalist accounts of reasons in general seem to be well-suited for explaining actions with artefacts where those artefacts are used as instruments by agents wishing to fulfil a particular desire or reach a particular goal (see Franssen: 2006). However, artefacts like speed bumps are not used for their instrumental value, yet the fact that there is a speed bump ahead of us seems to be a good reason to slow down, which is much easier to account for in an externalist account of reasons, one that is based on facts about what is good to do (including the fulfilment of desires). To put it otherwise, internalists tend to start with investigating desires and practical reasoning (‘what should I do, given what
my goals are?'), while externalists tend to start with investigating facts of the environment (‘what should I do, given what there is to welcome or avoid out there?’) Instruments are typically things sought by agents with particular goals (‘I want to nail this board into a tree. Where is my hammer?’), where artefacts like speed bumps are encountered and have to be dealt with, rather than actively sought out. As I focus on the latter kind of artefacts in chapter 5, I will use an externalist account to investigate how their influence works. In particular, I will use Dancy’s (2000a) account of practical reasons for this task. Additionally, Dancy’s account offers the advantage of distinguishing two other roles besides reason that facts can play, that can affect our practical reasoning. To show how Dancy’s account could have practical relevance for design, I will mention how design recommendations could be derived from it. This is not to say that an internalist account of the influence of artefacts on agents and agency would be impossible to give, but rather that it would likely be more complicated and best suited to explain the influence of those artefacts that are instruments.

While many artefacts can influence agents and agency by altering our reasons for action, human enhancement technologies can influence our agency in a more radical way: by affecting us in our capacity of bounded practical reasoners. What I mean by this is that we are beings with the capacity to set goals and reason about these goals, but that our reasoning is constrained by various factors, e.g. finite time and finite memory. Enhancement technology can affect both this capacity (e.g. by improving our ability to concentrate) and those factors (e.g. by improving our memory). Classically, the discussion on human enhancement has focused on issues like invasiveness and irreversibility (e.g. of the application of medication or brain implants). Chapter 6, ‘How enhancement technologies affect us as bounded practical reasoners’, challenges the idea that these are the most important notions to be discussed in the human enhancement debate. Specifically, it asks: Which kinds of technologies are especially morally problematic with regard to human enhancement? Its relevance for the main question consists in its argument that it is not so much the relation to the body that makes an artefact morally problematic (e.g. invasive versus non-invasive technologies), as well as how its use changes us as bounded practical reasoners. I argue that in the case of enhancement technologies the important question is
whether or not artefacts affect our practical reasoning, and thus influence our agency, regardless of whether they do so by invasive methods or otherwise.

Chapter 7, ‘Robust joint commitment and the use of strategies’, is again the application of existing work on the influence of artefacts on agents and agency to another field: that of collective, rather than individual, action. Artefacts can help us maintain our commitment to our individual actions, as anyone who has ever set an alarm clock can testify (Clark: 2003; Heath and Anderson: 2010). This insight has not yet been incorporated in theories of collective commitment and collective actions, that still regard commitment primarily as an act of the will. This chapter asks: How can the use of strategies contribute to a robust joint commitment? In particular, in this chapter I argue for a revised notion of joint commitment to actions that is strengthened by requiring the application of appropriate intentional, social and technical strategies to maintain this commitment. I will pay special attention here to technical strategies, which can include the use or implementation of technical artefacts. In this way, artefacts can influence not only individual, but also group agency.

One reason why action theory is so important for philosophy, as was mentioned earlier in this introduction, is the relevance of a good concept of action for allocating moral responsibility. I return to this issue in the third and final part of the dissertation, chapter 8: ‘Transferring responsibility for artefacts through use plans’. This chapter zooms out from action theory and the artefacts themselves to address the issue of designer and user responsibility for actions done with artefacts. It asks: What are the conditions under which the transfer of responsibility for an artefact from engineer to user can take place? In particular, it shows how responsibility can be transferred from designer to user by connecting the use plan theory of artefact use and design (Houkes and Vermaas: 2004; Houkes et al.: 2002; Houkes: 2006; Houkes and Vermaas: 2010, esp. chapter 2) to Fischer and Ravizza’s (1998) theory of responsibility.
2 Basic actions and the time of a killing

2.1. Chapter abstract

The problem of the time of a killing is often cited as showing an absurd consequence of the action identification thesis as proposed by Anscombe and taken up by Davidson, and thus, as providing grounds for rejecting it.

In this chapter I will make three claims. First, I claim that the problem of the time of a killing is a threat to the action identification thesis because of two assumptions the thesis makes: since the thesis takes actions to be a kind of doings, it has to assume that agents’ doings last as long as their actions and vice versa. Second, I claim that not making both of these assumptions necessarily leads to another problem, the problem of the acting dead. I will examine several proposals that do not make both assumptions and argue that so far, they have failed to come up with a satisfactory solution to the problem of the acting dead. This means that any theory of action has to choose its poison and face either one of both unresolved problems. Third, I claim that the solution to the problem of the time of a killing can be found in a close examination of what ‘kill’ means. Linguists have argued that its commonly assumed meaning, ‘cause to die’, can in fact not be its actual meaning due to differences in both linguistic characteristics and usage of ‘kill’ and ‘cause to die’. I will examine an alternative, more complex proposal and show how this allows us to keep the action identification thesis, fits colloquial usage of ‘kill’ and deals with the problem of the time of a killing. Additionally, this proposal shows that the problem of the time of a killing (and its solution) and the problem of the acting dead apply to a class of verbs rather than just ‘to kill’, namely, to those verbs that express something becoming something else.

While Davidson argues for both the action identification thesis and the thesis that all actions are bodily movements, in this chapter I will only defend the action identification thesis. Chapter 3 will be devoted to rejecting the thesis that

---

7 An earlier draft of this chapter has been submitted as a paper titled ‘Choosing your poison and the time of a killing’ to Philosophical Studies.
Acting with Artefacts

all actions are bodily movements and claiming that there can be basic actions with artefacts. Chapter 4, finally, will argue that the action identification thesis can be used to clarify the concept of affordance, or opportunity for action, in affordance-based design.

2.2. Introduction

There is a notorious argument against the action identification thesis that is called the problem of the time of a killing (Thomson: 1971). The thesis says, roughly, that when an agent does something by doing something else, the agent only performs one action, but this action can be described in different ways. The problem confronts the thesis with the following scenario: Suppose that Smith shoots Jones on Saturday. Jones collapses, bleeding, and finally dies on Sunday. The question we can now ask, is: ‘When did Smith kill Jones?’ According to Thomson and various authors after her writing on the problem (Mackie: 1997; Mossel: 2001; Weintraub: 2003; Sandis: 2006), those who accept the action identification thesis are committed to saying that since Smith has killed Jones by shooting him, the shooting is the killing. If that were the case, the killing would have happened on Saturday, a day before Jones died, which would be absurd.

In this chapter I will make three claims. The first claim widens the scope of the problem: it is that establishing the time of a killing is a problem for the action identification thesis because the thesis sees actions as a kind of doings (Davidson: 1980, essay 3), and thus has to make two assumptions: that agents’ actions last as long as their doings, and vice versa. The second claim is a challenge to those theories that do not make both assumptions: it is that those theories have to deal with another problem that I will call the problem of the acting dead. This means that any theory of action has to choose its poison and

---

8 The thesis goes by various names, including the ‘Action Sequence Identity Thesis’ (Mackie: 1997), the ‘Anscombe thesis’ (Bennett: 1994; Schnieder: 2008), and the Davidson/Anscombe thesis about act identification (Wilson: 2009). Here I will call it the action identification thesis, or AIT.

9 The same question can be asked for the place of a killing if Jones manages to leave the place of the shooting before collapsing. Though I will only concern myself with the time of a killing, establishing where the killing occurred can be done in a similar way.
confront either the one problem or the other. The third claim, finally, is that the problem of the time of a killing can be overcome by drawing on linguistic research into what ‘kill’ means, and why its usual interpretation as ‘cause to die’ cannot be correct (Fodor: 1970; Wierzbicka: 1975; Morreal: 1976). Though I will thus accept a proposed revision of what ‘kill’ means, my aim in this chapter is to deal with the problem of the time of a killing without accepting a solution that goes against colloquial usage of the word ‘kill’, for example, a solution that accepts that there can be killings before dyings (see Lombard: 1978, 1989; Heinaman: 1983).

The chapter is structured as follows. In section 2.3, I will explicate both the action identification thesis and the problem of the time of a killing, examine some earlier attempts at dealing with it, and argue that the problem hinges on the assumptions that agents’ doings last as long as their actions and vice versa. In section 2.4, I will show that any theory that does not make both assumptions necessarily falls prey to the problem of the acting dead. I will investigate several proposed solutions to the problem by authors who do not make both assumptions (Thomson: 1971; Mackie: 1997; Mossel: 2001), and argue that those proposed solutions do not solve the problem. In section 2.5, I return to the problem of the time of a killing and show how drawing on linguistic research into what ‘kill’ means can help us deal with that problem. Particularly, I will follow Fodor (1970), Wierzbicka (1975) and Morreal (1976) in arguing that the meaning of ‘kill’ cannot be, as is commonly assumed, ‘cause to die’, due to differences in linguistic characteristics and usage of both phrases. Instead, the meaning of ‘kill’ has to be rather more complex. It has already been observed that the problem of the time of a killing vanishes if we just do not try to locate killings in time very precisely (Thomson: 1971; White: 1980; Mossel: 2001; Sandis: 2006). Explicating the meaning of ‘kill’ can help us see why this is looking for the solution to the problem in the right direction, rather than evading it, and why this solution applies to a whole class of verbs, those that express becomings, rather than just to ‘to kill’.
Acting with Artefacts

2.3. The problem of the time of a killing

In this section I will first present the problem of the time of a killing, then go into each of its conditions in turn. The scenario, first, is as follows:

(A) Smith shoots Jones on Saturday. Jones collapses, bleeding.
(B) Jones dies from blood loss on Sunday.

The problem is basically a *reductio ad absurdum*, though the result is a colloquial, rather than a logical, absurdity:

(1) Smith has shot Jones on Saturday.
(2) According to the action identification thesis, if an agent does something by doing something else, the agent has only performed one action. Since Smith kills Jones by shooting him, the shooting is the killing.
(3) Therefore, Smith has killed Jones on Saturday.
(4) Therefore, Jones was killed a day before he died, which is absurd.

Except for step (1), which is simply taken from the scenario, all of these steps have been contested. I will now examine each of these steps in turn, starting with explicating the action identification thesis as presented in (2).

To get a grip on the action identification thesis, let me set the stage with an extended quote from Anscombe (1957/2000) about a man pumping water, thereby poisoning the inhabitants of a house:

"Are we to say that the man who (intentionally) moves his arm, operates the pump, replenishes the water supply, poisons the inhabitants, is performing *four* actions? Or only one? (...) more circumstances are required for [the moving of his arm] to be [the pumping of water] than for [the moving of his arm] just to be [the moving of his arm]... In short, the only distinct action of his that is in question is this one, [the moving of his arm]. For moving his arm up and down with his fingers round the pump handle is, in these circumstances, operating the pump; and, in these circumstances, it is replenishing the house water-supply; and in these circumstances, it is poisoning the household. So there is one action with four descriptions, each dependent on wider circumstances, and each related to the next as description of means to end...” (§26, 45-46).

So saying that ‘one action is, in these circumstances, another action’ means for Anscombe the same as saying that ‘one action *can be described*, in these circum-
stances, in other terms.’ Davidson (1980, essay 3) has elaborated on this. According to him, we regularly do things by doing other things (like the pumping man); if there is some description under which what we do is intentional, we can be said to perform an action. Neither Davidson nor Anscombe gives a formal definition of the action identification thesis implied in the passage above, but I will adopt one from Bennett (1994) here:

(AIT) “If someone’s by p-ing, and F is the act which makes it the case that she fs, and P is the act which makes it the case that she ps, then F is P.”

So if Smith kills Jones by shooting him, Smith performs an act F which makes it the case that he kills, and Smith performs an act P which makes it the case that he shoots, and F is P. What is (or properly describes) this act is, according to Davidson, what Smith does directly, not by doing anything else.

A worry expressed by Mackie (1997) is that AIT would not work because the is-relation between actions is symmetric and reflexive (F is P implies P is F), while the by-relation is asymmetric and irreflexive (Smith kills by shooting; but he does not shoot by killing). The worry is that, since we can swap F and P at will, the same could happen to f and p. However, close examination of AIT can put this worry to rest: f-ing and p-ing are ways to describe an action, while the identity relation claims that there is only one action to which both descriptions apply. The important point here, already made in the quote by Anscombe, is that whether a certain description applies to an action depends on the particular circumstances in which the action is performed. Killings can be shootings-in-

10 The thesis holds for acts, which can be either intentional actions or unintentional doings of an agent. So if I turn on the light by backing into the switch, my backing into the switch is my turning on of the light, even though both acts are unintentional. In this dissertation I will concern myself only with intentional actions.

11 There has been disagreement about whether these basic actions, or in Davidson’s terms, primitive actions, are bodily movements (Davidson: 1980, essay 3), or rather willings (Ginet: 1990), tryings (Hornsby: 1980) or could include artefact use (Annas: 1978). Though I will use bodily movements as examples of basic actions throughout this chapter, my defense of the action identification thesis does not hinge on any specific conception of basic actions. I will return to this issue in chapter 3 and argue that there can be basic actions with artefacts.
particular-circumstances, namely, if the shooting results in a death. Shootings, on the other hand, cannot be killings-in-particular-circumstances. Or in other words: a killing can be a kind of shooting (namely, a shooting that results in a death), but a shooting cannot be a kind of killing. This ensures that the asymmetry of the by-relation is not threatened, even though the action described is in both cases the same (see White: 1980).

As AIT assumes that actions are a particular kind of doings, it has to make two assumptions. The first is that agents have to be doing something for the duration of their actions, or that agents’ doings last as long as their actions. The second is that agents cannot be doing something after they have ceased to be acting, or that agents’ actions last as long as their doings (after Hornsby: 1980, 29). It is this combination of assumptions that makes step (2) of the problem valid and opens up AIT to the problem of the time of a killing. For it implies that if Smith has done something, in this case, moved his finger, his action is over: Smith needs to do nothing else in order to kill Jones. Yet if Smith has performed the action that can be described as killing Jones, it seems reasonable to infer that he has killed Jones. And if Smith has killed Jones, Jones has been killed, and therefore, Jones is dead – but he isn’t yet (Bennett: 1973, 321).

Bennett (1973) has challenged the step from (2) to (3) by arguing that, just like objects can acquire new properties, events (including actions) can too. He argues that Smith’s shooting was indeed a killing, though it did not become a killing until Jones died on Sunday. Anscombe (1979) similarly argues that some descriptions of acts can only become true later, and that careful formulation will solve our problems here. For example, since Smith’s action caused the gun to go off, we could say: “That act which (as things turned out) was the killing of [Jones] by [Smith] caused the gun to go off.” (227).

Yet there are at least three problems with such an approach. First, Smith’s shooting of Jones did not acquire a new property when Jones died. Rather, the shooting had the property of being a cause of death all along. On Saturday, the shooting is the event that causes Jones to die on Sunday, or that makes it the case that Jones dies on Sunday. We might not know this to be true until Jones does indeed die on Sunday, but those facts hold regardless of whether Jones’s
death has already occurred or not. Assuming that ‘kill’ is ‘cause to die’, the action
must have had the property of being a killing all along.

Second, Lombard (1989) has argued that indeed there cannot be a killing
unless there is a death, but that this does not imply that there cannot be a killing
until there is a death, which Bennett assumes. The core of his argument is that,
logically, ‘there cannot be a killing unless there is a death’ means: ‘If there is no
death, then there cannot be a killing’. This, again, means: ‘If there is a time \( T \) at
which there is no death, then there is a time \( T^* \) at which there cannot be a
killing’. It does not mean: ‘If there is no death now, there cannot be a killing
now,’ which would be required to state that there cannot be a killing until there is
a death. So from the proposition that there cannot be a killing unless there is a
death, we cannot derive the proposition that there cannot be a killing until there
is a death.

Third, Mossel has argued that Bennett’s solution only postpones the prob-
lem, not solves it. He notes that as soon as Jones has died on Sunday, it becomes
true that Smith’s shooting is a killing, and thus that Smith has killed Jones on
Saturday, a day before Jones died. In short, Bennett’s argument does not seem to
be able to prevent the move from (2) to (3).

I do not know of anyone who has challenged the move from (3) to (4), or
from ‘Smith has killed Jones’ to ‘Jones was / has been killed’. However, it has
been argued that, since each step in the problem so far seems plausible, maybe
the end result is not so absurd after all (Davidson: 1980, essay 8; Lombard:
1989). The only argument against this position seems to be the observation that
this is simply not how we use the phrases ‘kill’, ‘has been killed,’ etc. If we hear
that Jones has been killed, we would assume that he is dead, and we would at
least be surprised to find out otherwise. This argument from usage is strength-
ened by the observation that we do have a separate way to describe those
woundings that have not yet resulted in death, but will eventually do so: as fatal
woundings (see White: 1980).

As my aim in this chapter is to deal with the problem of the time of a killing
without accepting a solution that goes against colloquial usage of the word ‘kill’,
I will not settle for accepting killings before dyings, but rather argue that the
problem can be solved by establishing the meaning of ‘kill’. Before I do so,
Acting with Artefacts

however, I will deal with those who would address the problem of the time of a killing by arguing that it implies that we should abandon AIT, or more precisely, either of the two assumptions that I have claimed both support AIT and open it up to the problem. Specifically, I will argue that those who abandon either of those two assumptions open themselves up to the problem of the acting dead, and that so far no satisfying solution to this problem has been put forward.

2.4. The problem of the acting dead

The two assumptions that lie behind AIT are that agents' actions last as long as their doings, and vice versa. Either one or both of those assumptions can be renounced. The first assumption is renounced by Thomson (1971), who holds that the duration of an action of killing extends from the shooting up to and until the death of the victim, regardless of what the murderer does after the shooting. The second assumption is renounced by Mackie (1997) and Mossel (2001), who hold that agents can be doing something even though they are not active. For example, an agent could legitimately say: “I’m baking bread, doing the laundry and executing a database query,” while sitting in a comfortable chair, doing nothing but waiting for the dough to rise, and for the washing machine and the computer to finish.

No matter which assumption is renounced, those who renounce it run into trouble if they claim to be able to handle this scenario adapted from the time of a killing:

(A) Smith shoots Jones on Saturday. Jones collapses, bleeding.
(B) Smith dies of a heart attack, moments after the shot.
(C) Jones dies from blood loss on Sunday.

For those who assume that actions can extend beyond agents' doings, the problem goes as follows:

(1a) Smith has done something, namely, shooting Jones.
(2a) Smith’s action of killing Jones lasted from the shooting until Jones’s death (Thomson's assumption).
(3a) Therefore, Smith’s action of killing Jones took place (for the most part) when Smith was already dead. But that would be absurd.\textsuperscript{12}

For those who assume that doings can extend beyond agents’ actions, the problem is very similar:

(1b) Smith has performed an action, namely, shooting Jones.
(2b) Smith was doing something from the shooting until Jones’s death (Mackie’s / Mossel’s assumption)
(3b) Therefore, Smith was doing something while (for the most part) he was already dead. But that would be absurd.

There seems to be no good reason to object to assumption (1a/b), regardless of one’s position. Smith does something, namely, he is shooting; and he performs the action of shooting Jones. When Jones has been shot, everyone will agree that both the doing and the action are now over - and that this shooting is in some way or other related to Smith’s killing of Jones.

(2a/b) would probably be the easiest step to deny, for the simple reason that not accepting one of AIT’s two assumptions does not imply accepting that \textit{all} doings extend beyond an agent’s actions or vice versa, only that \textit{some} do. It could then be argued that in this particular case the doing does not extend beyond the agent’s action or vice versa. Of course, this would place the burden of proof on the arguer, who would have to come up with a criterion for when these doings or actions do extend beyond the other, and why they do not do so in this case.

Mackie (1997) suggests such a criterion. He presents us with the following case: Brown prints a paper by pressing a combination of keys on his computer, then waits for the printer to finish printing the document. Mackie argues that here we can say that Brown is printing the paper while he waits for the printer, and this is because Brown is doing at least \textit{something}: “Suppose that I die immediately after pressing the combination of keys that will start the printing process."

\textsuperscript{12} Thomson defends her position further in (1977); Heinaman attacks it in (1983) using a variant of the problem of the acting dead. As Heinaman has already shown the problems in extending actions beyond doings, I will concentrate here on showing the problem in extending doings beyond actions.
It does seem that in such a case we will not say that I am doing anything at all, not even printing out my article.” (51). So an agent can be doing something without being active, as long as that agent is still around.

It does not seem strange for Brown to say that he is printing the paper while he is waiting for the printer to finish, but this example is problematic for two reasons. First, we cannot obviously adapt Mackie’s example to the original time of a killing-scenario. If Smith shoots Jones and then waits for him to die, we would not say that he is killing Smith during that time. This does not necessarily render Mackie’s point invalid, but it does show that at least additional criteria are needed. Mackie suggests that the notion of control might be relevant here: Brown is in control of the printing process in a sense that Smith is not in control of Jones’s dying. Mossel (2001) adds that Brown has to actively supervise the process of printing, which may itself not require any physical actions. However, it is not clear that these criteria would do the required work: Smith might have saved Jones’s life by calling an ambulance, or providing him with a miracle cure, but that would not sway our intuition that Smith was not actively killing Jones, even if he stood by, supervising Jones’s demise. At best, Smith was making sure that Jones would die (as a result of his shooting); but that does not imply that Jones was killing Smith during that time. So positing control of and supervision over the process is not sufficient either.

Second, one might wonder whether Brown’s printing of the paper while he is waiting is the same kind of doing as his printing of the paper while he is pressing the combination of keys. Consider the following change in Mackie’s example: Suppose that Brown dies right after pressing the combination of keys. The printer dutifully does its job. Later, a colleague comes in and remarks, shocked: “The last thing he did was print out his paper.” This seems a correct observation: the intended result, the paper having been printed, has been achieved. In both cases, Brown does something in order to print out a paper, resulting in him having printed out a paper. Since we can say that Brown is printing the paper in the case where he remains alive, but not in the case where he is dead, yet the result is in both cases the same, this suggests that the ‘is printing’ refers to a plan or project that Brown has, rather than to the kind of doing that is directly involved in actions.
Basic actions and the time of a killing

It is true that Brown has a plan or project to have a paper printed in Mackie’s example, but not (anymore) in my example. What this suggests, however, is not so much that one of the assumptions behind AIT is false, as well as that the phrase ‘is doing’ can be used in different ways: to denote actions, or what the agent is doing now (Brown is sitting; waiting), and to denote plans or projects, or what the agent is doing today, these days, etc. (Brown is preparing a journal submission; setting himself up for a promotion) (see also Heinaman: 1983; Falvey: 2000). Or more specific: some things we do either by doing something else, or directly. According to AIT, those doings are actions, and those are also the doings that are referred to in the problem of the acting dead. Some things, on the other hand, we do by doing something else and something else again and..., like making falafels or preparing a journal submission, and those doings are plans that are being executed. AIT, as defined, is only about those doings that are actions, otherwise it would simply be false (e.g. Smith could have escaped from the police by masquerading as a construction worker, but there was no action of escaping, let alone one that was also an action of masquerading). This may seem to beg the question about what actions are; but it is rather stating that there are two conceptually different entities, actions and plans, and ‘is doing’ may refer to executing either of them. Showing that people can be engaged in doings in the plan sense is neither a rejection of AIT, nor a proper rejection of step (2) in the problem of the acting dead, which is about doings in the action sense.13

Two factors further complicate this picture. The first is that a plan can consist of only one action (and possibly waiting for its results, as with Brown’s plan for printing a paper), so that it may not always be possible to distinguish a doing in the action sense from a doing in the plan sense. The second is the status of activities: those are doings that stretch out over time, yet resemble actions more than plans, as the agent has to keep doing something for the duration of the

13 This observation also takes care of a point of Mackie that agents do not always have to be doing something for the duration of their action. He gives the example of slowly killing your flatmate by repeatedly administering small doses of poison over a longer period. As I have argued, this is rather a killing in a plan sense, consisting of multiple actions (the poisonings), none of which itself is the killing.
activity (Anscombe’s man pumping water; Jones’s wife killing Smith by strangling him). It might be that because ‘kill’ can denote both an action and an ongoing activity, ‘Smith is killing Jones’ sounds strange in the sense that ‘Brown is printing a paper’ does not. As a criterion for whether an inactive agent is doing something, however, it would clearly not be of the same kind as Mackie’s control and Mossel’s supervision. Moreover, it would be too vague: even ‘printing’ can denote an activity, namely, the operation of a manually operated printing press. Thus, both ‘is printing’ and ‘is killing’ can have at least two different meanings, depending on their context of use, and those meanings are easy to confuse.

The final possibility left to deal with the problem of the acting dead is to challenge step (3) and argue that this inference is either invalid or not absurd. The latter possibility seems most promising, as there is a sense in which we can do things after our deaths, for example: ‘Jones surprised his wife by leaving her his secret falafel recipe.’ This suggestion is not compatible with Mackie’s and Mossel’s proposed criteria, however: Jones was not in control of his wife’s being surprised, nor did he supervise the whole process that led to her being surprised. Furthermore, it seems that these kind of cases can in principle be reduced to cases of an agent doing something by doing something else during his life: by signing his will on Friday, Jones ensured that his wife would inherit his secret falafel recipe after his death (by which, in turn, he would surprise her.) All the things that agents can be said to do after their deaths have to be arranged by them during their lives.

In this section I have argued that the problem of the acting dead arises for anyone who does not accept one or both of the assumptions made by AIT, that agents’ actions last as long as their doings and vice versa. I have also argued that distinguishing between doings as actions and doings as plans is important both to understand AIT and to understand why Mackie’s and Mossel’s solutions to the problem are unsatisfactory as they stand. This does not imply that the problem of the acting dead could not be overcome, or would be harder to overcome than the problem of the time of a killing. It does imply that those who are quick to reject AIT (or rather, one of the two assumptions underlying it) based
on the problem of the time of a killing have a problem to solve as well if they wish to offer a plausible alternative.

In the final section I will revisit the problem of the time of a killing and suggest a possible solution based on what ‘kill’ means. The common assumption, which often appears in dictionaries, is that ‘kill’ means ‘cause to die’ (e.g. Lombard: 1989; Weintraub: 2003; Sandis: 2006). Thomson (1971), however, has already expressed doubts against this interpretation, and a number of arguments have been brought forward against it. I will now examine these arguments and the consequences they have for the problem of the time of a killing.

2.5. Killing and causing to die

Not only does it seem quite reasonable to say that ‘kill’ means ‘cause to die’, it also promises us a way out of the problem of the time of a killing. Weintraub (2003) has pointed out that causes and effects are events, while causal interactions, or causings, are not. This implies that causal interactions are not temporally localisable. This means that the question: ‘When did Smith kill Jones?’ asks for the temporal location of something that is not temporally localisable, so the question is meaningless (see Fodor: 1970, 437n9). Sandis (2006), however, rightly observes that causings are temporally localisable, though not very precisely, by simply taking the time frame including cause and effect. In our case, we could then say that Smith killed Jones in the weekend, which seems unproblematic.

However, this solution depends on interpreting ‘kill’ as ‘cause to die,’ and it is not immediately clear what ‘Smith caused Jones to die’ would mean, or what kind of causation would be involved. For that reason, Davidson (1980, essay 3, 48-49) has suggested that ‘kill’ rather means something like ‘do something that causes someone to die’, where the causation involved is event causation. Yet if we accept this, the problem of the time of a killing stands again: Smith did something on Saturday that caused Jones to die (on Sunday), namely, shooting

---

14 Sandis also claims, contra Weintraub, that causings are events.
him. This implies that the killing took place on Saturday, and we have the absurdity again.

Apart from this unclarity about what kind of causation is involved, there are more problems with viewing ‘kill’ as ‘cause to die.’ Fodor (1970) has shown that some transitive verbs may be analysed in this way (e.g. ‘melt’ = ‘cause to melt’), but not ‘kill.’ He has also mentioned that a sentence of the form ‘Smith caused Jones to die on Sunday by shooting him on Saturday’ would be well-formed, while ‘Smith killed Jones on Sunday by shooting him on Saturday’ would not be. Morreal (1976) adds that ‘cause’ is not an action verb and cannot be modified, where ‘kill’ is and can be. For example, we can say: ‘Smith killed Jones slowly and painfully’, but not: ‘Smith caused slowly and painfully Jones to die’. We can say: ‘Smith caused Jones to die slowly and painfully,’ but here it is the dying that is modified, not the killing.

Wierzbicka (1975) claims that ‘kill’ cannot mean ‘cause to die’ exactly because of its implications for the time and place of the event. She argues that a sentence like: ‘Peter killed the cat in the attic on Saturday’ implies that both Peter and the cat were in the attic when Peter acted, and that as a result of that the cat died in the attic on that same day. On the other hand, the sentence: ‘Peter caused the cat to die in the attic on Saturday’ does not carry such an implication: Peter might simply have set a lethal trap in the attic on Friday and then left the house. Taking this and other considerations into account, Wierzbicka defines the meaning of a sentence like ‘Smith killed Jones’ as follows: ‘At one time Smith and Jones were in some one place; Smith was doing something and something was becoming in contact with Jones’s body because of that; Jones was dying because of that; after that, Jones was dead because of that.’ 15 This is arguably rather more complex than ‘cause to die’, but given that ‘kill’ is used in a rather complex way, Wierzbicka states that that was only to be expected.

Suppose we adopt Wierzbicka’s definition. The question ‘When did Smith kill Jones?’ then means: ‘When was the one time in which Smith and Jones were in some one place...?’ The answer to this is then: ‘in the weekend.’ Not because a

15 Wierzbicka proceeds to work out and defend the meaning of the components of this formulation in the remainder of her paper.
killing is a causing and thus not precisely localisable in time, as Sandis (2006) argues, but because the definition of ‘kill’ requires a unity of time. Or: the time given in the answer should include the time of the shooting, the dying and the death.

At first glance, this may seem to be an abandonment of AIT. For if two actions (or events in general) are the same, they should occupy the same stretch of time (Davidson: 1980, essay 8). Yet Wierzbicka tells us that the shooting occurred on Saturday and the killing occurred in the weekend. The important point here, as noted by Anscombe in the quote at the beginning, is that one action is (or can be described as) another only in certain circumstances. Wierzbicka has spelled out those circumstances for ‘kill’, arguing that an action cannot be described as a killing if the unity of time is not honoured in the description. This is not incompatible with stating that there only is one action. One might legitimately ask how Smith killed Jones, or when and where the shooting occurred, or which description would best capture the action (see chapter three). Yet straightforwardly describing the shooting as a killing and then claiming the killing occurred before the dying is simply using ‘kill’ in a way that does not reflect its meaning. In contrast, we can unproblematically say that Smith fatally wounded Jones by shooting him on Saturday, or that he widowed Jones’s wife by killing Jones in the weekend. Given the circumstances at the time of the shooting, the action could then also be described as a fatal wounding, and given the circumstances at the time of the killing, the action could then also be described as a widowing.

Returning to our original problem scenario, we can say that this elaboration of the meaning of ‘kill’ renders the move from step (2) (accepting AIT) to step (3) (saying that Smith killed Jones on Saturday) invalid. This solution differs from that of Bennett (1973) and Anscombe (1979) in that it does not depend on old events retroactively gaining new properties, or a description of the action becoming true later, after Jones’s death, but on stating conditions for use of the word ‘kill’ that have to be fulfilled whenever the word is used. This overcomes

\footnotetext{16} This is different from Thomson’s (1971) claim that the action has to include those circumstances (see Davidson: 1980, essay 3; White: 1980).
Mossel’s worry that as soon as Jones has died, it becomes true that Smith’s shooting of Jones on Saturday was Smith’s killing of Jones on Saturday. After Jones’s death we can certainly say that Smith has killed Jones by shooting him, but not that Smith killed Jones on Saturday, or before Jones was dead; the definition of ‘kill’ does not allow this.

One last worry remains, which is Lombard’s claim that there cannot be a killing unless there is a death, but that this does not imply that there cannot be a killing until there is a death. Lombard’s claim is reasonable if we, with Lombard, assume that ‘kill’ means ‘cause to die’: one cannot cause a death unless there is a death, but to say that one cannot cause a death until there is a death would be wrong, as causes precede effects. Assuming Wierzbicka’s definition, however, there is only a killing unless and until there is a death. The important point here is that Wierzbicka does not suggest this definition because she makes the unless-until inference, which Lombard argues is fallacious: her reasons are rather observations on the common usage of ‘kill’, or what ‘kill’ implies, and the problems that arise from equating ‘kill’ with ‘cause to die’. We can say, though, after White (1980), that there cannot be a fatal wounding unless there is a death. Again, the fact that there is a difference between ‘kill’ and ‘fatally wound’ suggests that there might be more to ‘kill’ than just ‘cause to die’. Accepting this can help us deal with the problem of the time of a killing while preserving the action identification thesis.

I will end by noting that, while this definition of ‘kill’ helps us to deal with the problem of the time of a killing, it does not help us to deal with the problem of the acting dead. In particular, remember that Mackie and Mossel wanted to be able to say that, in some circumstances, agents can be doing something while they are no longer active. Suppose that they wanted to say this about Smith killing Jones, they would have to say that Smith was doing something (killing) during the weekend. Wierzbicka’s definition, however, only allows us to say that Smith was doing something (killing) in the weekend.

To this, Mackie and Mossel might reply that there are differences between Smith killing Jones and Brown printing a paper, such that they would want to
Basic actions and the time of a killing

say that Brown was printing a paper during, say, the whole afternoon, but not that Smith was killing Jones during the weekend. However, Wierzbicka argues that her analysis of ‘kill’ holds for all verbs that express that (someone does something because of which) something becomes something else, including printing. So if Brown prints a paper by pressing the right combination of keys, and his printer takes the whole afternoon to print the document, according to Wierzbicka we can only say that Brown printed the paper in the afternoon, not during the afternoon, whether or not he died after pressing the right combination of keys. This means that following Wierzbicka would not even give Mackie and Mossel their starting example, let alone solve the problem of the acting dead for them.

Finally, Wierzbicka’s claim makes it possible to generalise the problem of the time of a killing (and its solution) and the problem of the acting dead from ‘kill’ to all verbs of a similar kind: verbs that express a becoming, or describe actions (and other events) in terms of their consequences. Questions like: ‘When did Brown print the paper?’, or: ‘When did Jones microwave the falafels?’ might be less likely to require a precise answer than questions about killings, but the problem structure, and that of the solution, is in all cases the same.

2.6. Conclusion

The problem of the time of a killing is often presented as a reason to abandon the action identification thesis, as it would have absurd implications. In this chapter I have argued for three claims. First, that the problem threatens the action identification thesis because the thesis equates the duration of actions with that of doings. Second, that abandoning the assumptions that agents’ actions last as long as their doings and vice versa confronts one with another problem, the problem of the acting dead, to which so far no satisfactory solution has been given. Third, that by examining the meaning of ‘kill’ and acknowledging that it is not ‘cause to die’, we can come to a definition of ‘kill’ that does not lead to any absurd consequences, that does adequately deal with the problem of the time of a killing and that does allow us to keep the action identification thesis.
3 Can there be basic actions with artefacts?\textsuperscript{17}

3.1. Chapter abstract

Actions with artefacts are often things we do by doing something else, like turning on the light by flipping the switch. If we accept the action identification thesis that I have defended in chapter 2, and if we follow Davidson in assuming that (basic) actions are those things we do not do by doing something else, mere movements of the body, it follows that there are no (basic) actions with artefacts. Rather, we should conceive of actions with artefacts in the following way. All actions we can perform are bodily movements, as we can move some body parts directly, not by doing something else. As actions are events, those bodily movements can affect the world through event causation. Of course, we may intentionally bring about certain consequences and use technical artefacts by intentionally acting, and we may describe our actions in terms of the consequences we intentionally (try to) bring about, such as ‘lighting the room’, or our use of artefacts, such as ‘turning on the light’. Technical artefacts have made it possible for us to bring about events that are remote in time and space, yet have enormous consequences, such as president Truman who brought about the bombing of Hiroshima and Nagasaki by signing an order. But the basis, the intentional action, is always a bodily movement according to Davidson.

I will call this the ‘standard view’ of actions with artefacts, and argue in this chapter that it is false. I will argue that this is so because humans are natural tool users: when we become skilled enough in using a tool, we can accurately and automatically act with it, so much so that it seems to become an extension of the

\textsuperscript{17} An earlier draft of this chapter has been submitted as a paper to Philosophical Explorations. The editor has requested a revision and resubmission of that paper. Earlier drafts of this chapter have been presented at the Graduate Research Seminar of the University of Reading Philosophy Department and the Sixth European Conference of Analytic Philosophy.
Acting with Artefacts

body, and transparent in use. This means that it is not clear why we would be able to do things directly with our body, but not with at least some artefacts.

In particular, this chapter argues for two claims. The first, critical claim is that while Davidson’s notion of basic actions is ambiguous, neither a causal nor a teleological interpretation of basic actions leads to Davidson’s conclusion that all basic actions are bodily movements. The second, constructive claim is that a teleological notion of basic actions will encompass actions performed with artefacts. This claim is supported by the phenomenology and the psychology of tool use, and has consequences for the metaphysical question of what an action is, as well as for the importance of the psychology of tool use for the philosophy of action.

3.2. Introduction

All actions are bodily movements. At least, that is what Danto (1963) has suggested and Davidson has explicitly claimed and defended in his essay Agency (1980). This claim has subsequently been endorsed by a number of action theorists, including Frankfurt (1978), Wilson (1989), Vollmer (2001), Smith (2004) and Haddock (2005). Others do not adhere strictly to the claim, but remain quite close to it, saying that actions are constituted by bodily movements (Baker: 1998; Schlosser: 2006), that agents’ contributions to overt actions end with the termination of their guided bodily motions (Mele: 2000), or that bodily movements are the overt components of actions (Searle: 1983; Mossel: 2005; Pacherie: 2008). Clearly, the relation between actions and bodily movements is considered to be a tight one.

Phenomenologically speaking, the temptation to claim that all actions are bodily movements is understandable. After all, we can do things directly with our body, not by doing anything else: we flick a light switch by moving our hand, but our hand we move directly (Davidson: 1980). Our body is transparent to us in action: we do not have to (consciously) attend to our bodily movements while performing routine actions, but can focus on the task at hand. This can be the

Whenever I refer to Davidson, I refer to Davidson (1980, essay 3), unless stated otherwise.
flipping of the light switch rather than the moving of our finger, or the flying of the model plane rather than our fingers bending to grip and twist the controls (Sartre: 1943/1977; Merleau-Ponty: 1945/2002). And we seem to have direct control over our bodies, whereas we control features of the external world, like model planes, only indirectly (Vollmer: 2001; Mossel: 2005).

The phenomenology of embodied action seems to provide us with good reasons to award bodily movements a special status in action. However, when we look at the phenomenology of tool use, or actions performed with technical artefacts or instruments, it turns out that the phenomena mentioned earlier with respect to the body may occur with respect to tools as well. Examples that spring to mind are permanently used tools like prostheses and contact lenses, but it has been claimed that artefacts like a blind man’s cane phenomenologically extend our body in acting and become transparent to us in use (Merleau-Ponty: 1945/2002), and that artefacts in action can withdraw from our attention much like our body parts do, like the hammer wielded by the expert carpenter. This allows us to focus on the task at hand, hammering a nail into a board, rather than on manipulating the artefact: the hammer becomes 

\[\text{zuhanden}\text{} \] rather than \[\text{vorhanden}\text{} \] (Heidegger: 1927/1962), or available rather than occurrent (Dreyfus: 1995). It seems that, insofar as we wish to use phenomenology in determining what actions are, there might be no good reason to restrict ourselves to bodily movements and exclude all cases of tool use. Doing so could not only lead us astray with regard to the metaphysical question of what actions are, but also with regard to integrating the philosophy and psychology of action. For example, it could lead us to focus exclusively on the psychology of bodily movements for explaining actions at the cost of ignoring valuable insights from the psychology of tool use.20

---

19 For the purposes of this dissertation, I do not consider a tool to be significantly different from an artefact. For a possible distinction between instruments, tools and artefacts, see Dipert (1993).

20 While I focus on basic actions with artefacts, at a presentation in Reading I was asked whether it would also be possible, under certain circumstances, to perform basic actions with animals. Also, Olle Blomberg (2011) defends a claim very similar to one that there may be basic actions with other agents, namely, that the content of an intention-in-action may include the actions of other agents. Given the close bond, trust and intimate knowledge of each other’s capacities and
In this chapter I will argue that the phenomenology and psychology of tool use present a dilemma to those who hold that all actions are bodily movements. I claim that if ‘bodily movements’ here is interpreted in a narrow sense (as Davidson seems to do), where the body stops at our skin and bodily movements can only be movements of our flesh-and-blood limbs, the claim that all actions are bodily movements has to be abandoned. On the other hand, those who wish to keep this claim will have to acknowledge that ‘bodily movements’ has to be interpreted in a broad sense, where the use of tools, from prostheses to canes and hammers, could constitute bodily movements, and those tools should under certain conditions be considered to be body parts.

In this chapter I will focus on the first option and argue that, given a narrow interpretation of what counts as bodily movements, the claim that all actions are bodily movements is untenable. Particularly, I will examine the arguments given for the claim that all actions are bodily movements and argue that they do not hold. I will then suggest how our concept of actions could be extended beyond bodily movements. I will use the concept of basic actions here and the act identification thesis, as developed by Anscombe (1957/2000) and Davidson (1980). I will argue that keeping the basic actions and act identification thesis, but rejecting the claim that all actions are bodily movements, offers us the possibility of developing a concept of actions that is supported by, rather than ignores, the phenomenology of tool use. For this I will explicate the relations between notions like doing directly, transparency and skill in action, whose relevance extends beyond that of determining what basic actions are. For those who would rather interpret bodily movements in a broad sense, this analysis will show what kinds of tool use could reasonably be considered to be bodily movements.

The structure of my argument is as follows: In section 3.3, I argue that Davidson’s definition of basic actions can be interpreted in two ways: causally and teleologically. In section 3.4, I argue that if we interpret basic actions as causally

---

behaviour present in e.g. some riders and their horses, or skilled dancers, these claims seem to me plausible and well worth investigating, though doing so goes beyond the scope of this dissertation.
Can there be basic actions with artefacts?

basic, actions can be antecedents of bodily movements rather than bodily movements themselves. Thus, not all causally basic actions are bodily movements. In section 3.5, I argue that if we interpret basic actions as teleologically basic, actions can encompass more than bodily movements. Thus, not all teleologically basic actions are bodily movements. From this, I will conclude that Davidson’s claim that all basic actions are bodily movements (in a narrow sense) is on either interpretation untenable. In section 3.6 I show that interpreting basic actions as teleologically basic allows us to accommodate both the phenomenology of embodied action and that of tool use, and keep basic actions and the act identification thesis. I also suggest that skill is strongly involved in determining an agent’s basic actions, or what the agent can do directly.

Given recent developments in the philosophy of mind (Dennett: 1984; Clark: 2003) and the philosophy of technology (Ihde: 1990; Verbeek: 2000/2005), one might wonder whether adjusting the notion of basic actions is sufficient to properly accommodate the phenomenology of tool use. Indeed it has been argued that we should focus on systems rather than on embodied individuals, and that it is the system that thinks or remembers (Clark), intends or acts (Verbeek), in short, that has agency. By assigning agency to the system rather than the individual, the distinction between embodied individual and tool becomes irrelevant.

Adopting this systems perspective certainly has its advantages. Human beings are essentially embodied. Given that artefacts can phenomenologically extend our body and withdraw from our attention, extending the unit of analysis from the embodied agent to the embodied artefact-using agent would be a logical step. Apart from that, it would simplify explanations of action in some cases, as the same individuals may behave differently depending on which system they are a part of. And insofar as we can control which systems individuals can and do become a part of, e.g. by letting some technology become available on the free market, we can predict the behaviour of these systems or even guide it.21

---

21 For ethical reflection on the desirability of this, see Verbeek (2006).
Acting with Artefacts

However, adopting a systems perspective only relocates our problem rather than solves it, and creates several new problems besides. An agent’s skin may not be the clear boundary it seemed to be at first glance, but we need to draw a boundary somewhere between a system and its context in order for a systems perspective to work, and finding the right place to draw this boundary is as much a challenge to those adopting a systems perspective as it is to classical action theorists (see Kroes et al.: 2006). Furthermore, while some artefacts may be attached to an agent, like prostheses, or readily available, like watches, many artefacts are used only episodically and put aside once the goal is reached or dropped. Having systems as agents rather than embodied individuals means having unstable agents that can constantly change themselves and their capacities drastically. This would make it hard to maintain a meaningful general concept of agency. One of the downsides of this is that responsibility ascription would become problematic: it has been argued that taking a systems perspective can sometimes help in responsibility ascription (Johnson and Powers: 2005), but it might hinder it just as well and does not seem to fit with everyday practice (Illies and Meijers: 2009). Besides, it does not seem to be necessary for a proper account of responsibility for actions performed with artefacts (ibid.; Pols: 2010, or chapter 8). This is not to say that a systems perspective could never yield us a good theory of action; but I would rather attempt to amend our existing theory here than to simply abandon it in favour of another that probably requires much more effort to amend.

3.3. The ambiguous account of basic actions

In this section I show how Davidson’s account of basic actions allows for two possible readings and thereby, two kinds of basic actions that could be bodily movements.

To recapitulate the theory on actions, Davidson holds that there is a fairly definite subclass of events that are actions. He delimits this subclass in two important ways. First, he notes that agents often do something by doing something else, for example, alerting a burglar by illuminating a room by turning on the light by flipping the switch by moving their hand. Like Anscombe
Can there be basic actions with artefacts?

(1957/2000), Davidson claims that there is only one action here, but described in different ways (essay 1). This is the action identification thesis.

Of course, happenings can similarly be described in different ways, so Davidson needs a second restriction to distinguish actions from mere happenings and involuntary doings: all actions may be bodily movements, but not all bodily movements are actions. His solution is that ‘...a man is the agent of an act if what he does can be described under an aspect that makes it intentional.’ (essay 3, 46). So, actions are doings that are intentional under some description. To distinguish intended actions from intended consequences, Davidson claims that (basic) actions are ‘the ones we do not do by doing something else, mere movements of the body’ (ibid., 59). So turning on the light is not a basic action, because we do it by flipping the switch. Neither is flipping the switch basic, but moving your hand is: you do not (normally) move your hand by doing anything else, according to Davidson. So any proper description of the action itself will be in terms of bodily movements.

A quick note to avoid confusion: there are actually two questions here, what is the basic action, and what is the basic description of an action. The entanglement of both questions is hard to get rid of, as we cannot talk about basic actions without describing them, yet the first question is about what actions are, while the second is about how they should be described. Here I will proceed from the assumption that the basic description of an action is always identical to the description of the basic action.

Davidson claims that basic actions are those we do not do by doing something else. Yet it is not quite clear what function the word ‘by’ fulfils in this sentence. A good assumption to start with would be to assume that it defines a causal relation between events. There are two reasons to assume this. First, the notion of basic actions was introduced by Danto (1963), in whose theory basic actions are causally connected to non-basic actions, where both kinds of actions are events. The second reason is that while Davidson denies that actions can cause other actions, event-causality is still the backbone of his theory; it is what allows agents to bring about events and their results in the world.
Hornsby (1980) has followed this idea to formulate a notion of a causally (most) basic description, or the basic\(_C\) description of an action.\(^{22, 23}\) Her idea behind causal basicness is that a cause is ‘more basic’ than its effect. Hornsby argues that some descriptions of actions really describe actions in terms of their (event-causal) effects, and some describe the action itself, and those last descriptions are causally more basic. Formally:

“A description \(d\) of a particular action \(a\) is a more basic description than another description \(d'\) iff the effect that is introduced by \(<a, d>\) causes the effect that is introduced by \(<a, d'>\).” (71)

Where \(<a, d>\) is an action/description pair that can introduce an effect, and

“A description [of an action] is (...) one of the basic descriptions – if there is no other description that is more basic than it is.” (72).

This means that the most basic\(_C\) description of an action cannot mention anything that is itself an effect. While it may not always be easy to say whether a description is really effect-free, the strength of Hornsby’s definition is that it starts with finding out whether a description is more basic\(_C\) than another. Establishing this can be done more easily, by looking at the places of the events described in the causal chain. Hornsby’s definition of basic\(_C\) descriptions also supports her claim that all actions are tryings that occur inside the body: tryings are among the most basic\(_C\) descriptions of an action, since a trying is not an effect itself (if I do not try to raise my arm, I cannot be said to have done anything towards raising my arm.)\(^{24}\) So the basic\(_C\) description of an action describes the causally basic action.

\(^{22}\) Whenever I refer to Hornsby, I refer to Hornsby (1980) unless stated otherwise.

\(^{23}\) Baier (1971) has argued that many more possible forms of basicness can be defined. Causal and teleological basicness have the most obvious potential explanatory value, however. In addition, none of the other forms of basicness Baier lists would make it the case that all basic actions are bodily movements.

\(^{24}\) Hornsby still holds that all actions are tryings, though she has abandoned the constraint that they have to occur inside the body (Hornsby, personal communication).
Can there be basic actions with artefacts?

However, as Hornsby notes, causally basic descriptions have only limited explanatory value. Often, when we want to explain why an agent did something, we want to know what the agent thought he was doing, and how, and why, and causally basic descriptions cannot tell us that. An alternative account of basicness that could cover this, Hornsby argues, would have to draw upon what the agent did intentionally. This would be the teleologically (most) basic description of an action, or the BASIC description:

“A description $d$ of an action $a$ of agent $x$ is more BASIC than another description $d'$ of $a$ iff in virtue of $a$’s occurrence ‘$x$ intentionally [$d'$] by [$d$]’ is true.” (78)

(Where $[]$ denotes the correct inflection of the verb used.) And:

“A description of an action is the BASIC description of it iff there is no other description MORE BASIC than it.” (79)

Note that people do not normally use teleologically basic descriptions when describing what they do: they tend to favour high-level descriptions, or descriptions in terms of intended consequences (Vallacher and Wegner: 1987). Rather, it is the most basic description an agent can give of her action, for example, when we ask the agent repeatedly: ‘How do you do that?’ This does not mean that the agent cannot ever, by slowly executing the action over and over again and closely studying her movements, discover how to break up the action, meaning that the descriptions of its components are turned into BASIC descriptions; just that this should be a discovery rather than the usual way in which an agent intentionally proceeds to perform a certain action.

An advantage of this definition is that we do not have to worry about the possibility that behind the basic description there might lie an even more basic description, which is always a possibility with causally basic descriptions. As the agent is the ultimate arbiter about what is the BASIC description of his or her actions, discovering it will usually be no more difficult than getting an honest
answer from the agent – though there are cases where agents might be mistaken about what exactly they do intentionally.\(^{25}\)

As I will show in a moment, both classes of basic descriptions do not completely overlap. A natural next step would therefore be to argue, as Hornsby does, that one kind of basic descriptions describes the actions, and the other kind does not, though it might still be useful for other purposes. To make my case against Davidson’s claim that all basic actions are bodily movements as strong as possible, however, I will argue that no matter which kind of basic descriptions we take to describe the actions, we cannot claim that all basic actions are bodily movements.

Before I do so, however, I would like to put one potential worry to rest: that ‘by’ is an inherently causal notion that on its own is not strong enough to support a teleological reading, and thus, a class of teleologically basic descriptions. I will illustrate this point with an example from Hornsby. Take the following two sentences:

\[(1) \text{ “A contracted his muscles by clenching his fist.”} \]
\[(2) \text{ “A clenched his fist by contracting his muscles.”} \]

\(^{25}\) A challenge to this definition of teleologically basic actions arises from research done by Marcel (2003). Marcel had subjects hide one of their arms from view under a table, then tricked them into misjudging the position of that arm (say, the subjects were led to believe it was to the right of a target where it really was to the left). Those subjects would then intend to move their arm towards the target, intend to do so by moving it to the left – but in fact did move their arm to the right, towards the intended target. This shows the need for further refinement of the notion of a basic action.

On the other hand, this finding fits well with the notion of teleologically basic actions: subjects learning of this effect may well learn to move their arms in a certain way by moving them towards the target rather than by moving them directly to the left or right, trusting their visuomotor systems to fill in the details. In addition, if it would turn out that some or most of the bodily movements supposed to be intentional actions do not really classify as such, the natural place that is suggested by Marcel’s experiment to look for those actions would be among our doings under higher-level descriptions (like pointing to the target). This would thus offer a further challenge to the claim that all actions are bodily movements.
Here, (1) is a teleological explanation (A’s goal was to contract his muscles, the means used to attain the goal was clenching his fist) and (2) is a causal explanation (the fist clenching was caused by the muscle contracting).

As Hornsby thinks that ‘by’ is an inherently causal notion, denoting a cause–effect relationship, she is disposed to deny that statements like (1) could ever be true (94). Hornsby contrasts this with Von Wright’s (1971) claim that statements like (2) can never be true, and his denial that a scientific explanation of the causation of human movement can tell us anything about human actions. Moreover, since Von Wright does believe ‘by’ to be an inherently causal notion, he is forced to accept the uneasy conclusion that (1) is a case of backward causation.

I propose that both Hornsby’s and Von Wright’s conclusions are unwarranted. Instead, I will argue for a third option: that ‘by’ is an ambiguous notion. This means that both statement (1) and (2) can be true and sensible without accepting phenomena like backwards causation. Both forms of explanation can then have their value for action theory.

Wreen has pointed out that ‘by’ can have a non-causal meaning that is not necessarily fundamental to causation, namely: ‘indirectly achieved or performed by’ (1987, 124). It does not seem strange to say: “A contracted his muscles and this was indirectly achieved (or performed by) clenching his fist.” It may not be entirely clear why this is teleological and, more specifically, non-causal. Kosrovani (1991) helps here in noting that (1) is only compelling assuming that A acted intentionally. Roughly, we could say that causal basicness deals with cause-effect relationships (in processes) and teleological basicness with means-end relationships (in plans). The confusion then arises because causes and effects almost always overlap with means and ends, for example, when I turn on the light by flipping the switch. They do not overlap, however, in those specific cases when an agent intentionally wants to bring about something which is (the result of) (part of) the unconscious and automatic process involved in performing a

\[\text{For the discussion on the validity and significance of Wreen’s point, see Hornsby (1987), White (1987), Pfeifer (1988), Wreen (1988) and Kosrovani (1991), in that order.}\]
teleologically basic action. Here, ‘attaining ends Y by (means of) X’ does not imply that X causes Y. This does not mean that there is no direct or indirect causal connection between X and Y whatsoever, just that a non-causal reading of ‘by’ can offer added explanatory value for actions by referring to the plans and capabilities of the agent.

In the remainder of this chapter, I will proceed to argue that no matter which kind of basic descriptions we take to describe the actions, we cannot claim that all basic actions are bodily movements. Specifically, in the next section, section 3.4, I examine how we can identify causally basic actions, and argue that they can be the antecedents of bodily movements. In section 3.5, I then treat teleologically basic actions, and argue that they can encompass more than bodily movements.

3.4. Causally basic actions

The thesis that all actions are bodily movements seems hard to defend when actions are interpreted as causally basic. This is because causally basic descriptions of actions are effect-free, that is, they do not describe events that are effects of other events. Yet it is undeniable that bodily movements have causal antecedents. A possible consequence of our definition of causally basic actions is then that those antecedents may turn out to be the ‘real’ actions: muscle flexing, neural activity, brain activity – and if it turns out that an external event has caused this specific pattern of brain activity, it seems we should call that the action. Even if we add additional criteria to stop the regress at the agent, for example, that actions originate in an agent, or that actions are done by agents in some way, the basicC description of an action is not necessarily in terms of bodily movements. Several authors claim that mental events precede or cause our bodily movements, where these events are usually either acts of the will (volitions) or tryings. This would mean that we make our bodily movements by doing something else after all, namely willing or trying to make them. If this is true, we can say that ‘trying is an essential constituent of intentional action as such’ (O’Shaughnessy: 1973, 365), that ‘every basic action is a causally simple mental action’ (Ginet: 1990, 20), or even stronger, that ‘every action is an event of trying
or attempting to act, and every attempt that is an action precedes and causes a
contraction, of muscles and a movement, of the body.’ (Hornsby: 1980, 33).27

Davidson, having anticipated this line of criticism, states that this would not
mean that bodily movements are not basic actions, because you still do not move
your body by doing something else: ‘doing something that causes my finger to
move does not cause me to move my finger; it is moving my finger.’ (50). But
this counterargument seems to beg the question: after all, the claim Davidson
has to refute is precisely that you do move your body by doing something else,
namely willing or trying to move your body. Davidson could amend his claim by
arguing that you do not move your body by intentionally doing something else,
but that would mean committing himself to a teleologically basic account of
actions. As we shall see in the next section, this would not save his claim that all
basic actions are bodily movements.

This chapter will not go deeper into the arguments for and against actions as
willings / tryings. For my purposes, it suffices to note that it is not only possible
to construct an account of causally basic actions that are not bodily movements,
but that it has indeed been done and argued for extensively.

3.5. Teleologically basic actions

An action can be explained not only by looking at the causes of what we do, but
also by looking at what the agent is trying to accomplish. When people are asked
to describe what they are doing, they generally favour high-level, goal-directed
descriptions (“turning on the light”) over low-level, mechanical descriptions
(“flipping the switch”) (Vallacher and Wegner: 1987). This is quite natural:
elementary motor skills and behaviours like reaching, grasping and walking are
learned at a very young age and after that only draw our attention in special
circumstances. This ‘automation’ or ‘chunking’ (Graybiel: 1998) of basic move-
ments is important for two reasons. First, it frees up cognitive capacity that we
can then use for other tasks, like keeping an eye on our progress towards our

27 The ‘I’ here stands for intransitive, passive, while Hornsby uses a ‘T’ to denote transitive, active
meanings of ambiguous verbs. An example: ‘If I move my body, then my body moves.’
Acting with Artefacts

goals. Second, it makes the automated movements more fluid and resistant to
disruption from those other tasks, even if they tax the working memory (Masters
and Maxwell: 2004). In other words: for a child, most teleologically basic actions
might be bodily movements, but adults can be expected to have sufficient motor
skills to perform their actions without explicit regard to specific bodily move-
ments: they do not have to move their bodies intentionally in order to use
artefacts or perform certain tasks intentionally.

Some philosophers reason similarly. Anscombe remarks: ‘In general, (...) one
does not deliberate about an acquired skill; the description of what one is doing,
which one completely understands, is at a distance from the details of one’s
movements, which one does not consider at all.’ (1957/2000, 54). On a more
modest note, it has been claimed that simple motor actions requiring only or
mostly procedural knowledge, like speaking words, typing or tying your shoe-
laces, are basic actions (Baier: 1971; Annas: 1978).

Davidson argues against this that there is always some description of such an
action in terms of bodily movements, if only ‘I move my body in just the way
required to tie my shoelaces’ (51). Just like Davidson’s answer to the volitionists,
however, this answer begs the question. Davidson’s description ‘in terms of
bodily movements’ brings in a high-level, teleological description (‘to tie my
shoelaces’) through the back door. If we were to reformulate this action in terms
of bodily movements without referring to the goal of getting your shoelaces tied,
we would probably end up with a definition that is either too general to properly
describe the complicated bodily movements involved in shoelace-tying, or too
specific for agents to recognize as a proper description of their action.

Even if we allow Davidson to use this action description, however, we can still
argue that it does not describe the teleologically basic action in this case. The
reason for this is the following: Davidson has to assume that we can intention-
ally move our bodies in just the way required to tie our shoelaces directly, not by
doing something else. Yet empirical studies have shown clear differences in the
bodily movements of normal subjects pantomiming tool use and actually using
tools (Goodale, Jakobson, and Keillor: 1994). Similarly, patients with the neuro-
psychological disorder apraxia may hardly be able to pantomime tool use while
they can still use actual tools to some degree (Goldenberg, Hentze, and Hermstedt: 2004). These findings imply that ‘moving your body in just the way required to handle a tool’ is often only possible if you handle a tool. But if that is the case, it follows that we can only intentionally move our bodies in just the way required to handle a tool by intentionally doing something else, namely, handling that tool. So Davidson’s description in terms of bodily movements does not describe the teleologically basic action here. When we interpret ‘by’ teleologically, doing something not by doing something else, or doing something directly, implies that actions may encompass more than bodily movements.

Several authors have argued that actions do not encompass more than bodily movements. Vollmer (2001) holds that ‘most actions are bodily movements’ (638), and defends this position by arguing that we have a special kind of control over our body. ‘Actions are movements that agents have control over (...) [and] control of this kind, in turn, depends on inside (proprioceptive) awareness of the body and its movements, as well as on a number of physiological conditions.’ (ibid.) Mossel (2005) reasons similarly: he argues that the body is under the agent’s ‘direct control’, and that the external world is at best under our ‘indirect control’. Thus, by positing direct control over your means of action as a prerequisite for ‘doing something not by doing something else’, Vollmer and Mossel claim to have an argument for restricting actions to bodily movements.

However, it is instructive to look at the example Mossel gives to illustrate his point. He writes: ‘I control the flight of my model plane by moving the stick on its remote control unit, and therefore indirectly. However, in raising my arm I control its rising directly.’ (143). Controlling a model plane is a good example of indirect control. However, a model plane is a complex technical artefact, and the mapping of the movements of the stick to those of the plane depends on a pragmatic designer choice rather than on mechanical necessity. When we look at artefacts that mechanically transform hand movements (like tin-openers or screwdrivers), or simply extend them (such as a blind man’s cane), the distinction blurs.28 Whether we control our tin-openers directly can be debated, but it

---

28 See Goldenberg and Iriki (2007) for the relevance of this ordering of tools for neuropsychological research.
seems strange to say that we have only indirect control over our canes or walking sticks.

Both Vollmer and Mossel emphasize the importance of proprioceptive awareness or experience involved in direct control. However, the role of proprioception in making you aware of (controlled) goal-directed movements seems very limited (Frith: 2005). Indeed, Marcel (2003) has argued that proprioceptive awareness is neither necessary nor sufficient for action awareness, which suggests that it can only be of limited use in controlling actions. Even if proprioceptive awareness would be important for direct control, however, it would not be immediately clear why this awareness should be limited to the body. One possible reason could be that our proprioceptive organs extend throughout the body but not into tools. Yet when we use a hammer to pound a nail into a board, it seems that we do (proprioceptively) experience the hammer, its weight and its length. Without looking, we feel whether we have just hit the nail or not, and if we did, whether it went smoothly into the wood. This simple example might not settle the issue of whether we can sense stimuli at the tips of our tools, but at least it speaks in favour of it.²⁹

Neither ‘doing directly’ nor ‘direct control’ point us unambiguously towards the conclusion that basic actions of any sort must be bodily movements. It seems that causally basic actions can include antecedents of bodily movements, and that teleologically basic actions can encompass more than bodily movements. Hence, neither conception of actions leads to Davidson’s claim that all basic actions are bodily movements.

In section 3.6, I consider more empirical evidence to support my claim that teleologically basic actions can involve more than bodily movements, as well as insights from phenomenology. While I do not give a fully developed account of

²⁹ The claim that we can sense stimuli in the tips of our tools has been made in passing by various authors (Merleau-Ponty: 1945/2002, 165-166; Wittgenstein: 1953/1958, § 626; Paillard: 1993, 40), and there exists neuropsychological evidence that is at least consistent with this claim (Yamamoto and Kitazawa: 2001). However, to my knowledge it has not been philosophically investigated what such a claim would amount to, and what its implications would be for the philosophy of perception.
Can there be basic actions with artefacts?

the mechanisms underlying teleologically basic actions here, I do offer suggestions on how to (re)interpret and connect some of the underlying notions in order to find out what such an account would look like.

3.6. Basic actions with artefacts

In the previous section I have argued that teleologically basic actions can encompass the use or manipulation of artefacts. However, this claim hinges on other claims I have made that are in need of clarification, namely the claims that artefacts can phenomenologically extend our bodies, that we can do things directly with artefacts, that they can become transparent to us in use, and that we can exercise direct control over them. In this section, I will go into what these claims mean and how they interrelate, and test their validity. In particular, I suggest that the notion of skill can help support these claims, and help us better understand what teleologically basic actions we can perform, why our repertoire of teleologically basic actions can change over time, and why they can be performed with as well as without artefacts.

Artefacts and bodies can have a very intimate relationship. Subjective experience of artefacts may vary depending on circumstances, but there will certainly be cases where an artefact (e.g. glasses or a prosthesis) is experienced as part of the body or part of the agent. Merleau-Ponty (1945/2002) has investigated the phenomenology of artefact-body relations, claiming that artefacts can phenomenologically extend our bodies, that is, be experienced as if they were body parts.

“To get used to a hat, a car or a stick is to be transplanted into them, or conversely, to incorporate them into the bulk of our own body.” (166). Most famous is his example of the blind man, whose body is phenomenologically extended to his cane: after a while, he no longer feels the cane, only the touch of the ground or an obstacle. The sense of his body is extended beyond his bodily boundaries: “The blind man’s stick has ceased to be an object for him, and is no longer perceived for itself; its point has become an area of sensitivity, extending the scope and active radius of touch, and providing a parallel to sight.” (165). In this sense, we do gain direct control over the artefact because it phenomenologically extends a part of the body we have direct control over.
Interestingly enough, Merleau-Ponty’s observation has been backed up by neuroscientific research claiming that even simple tools can easily be incorporated into our body schema: an internal system keeping track of our body shape and posture (Maravita and Iriki: 2004; Cardinali et al.: 2009). The body schema is action-oriented, used by subconscious processes to fine-tune and coordinate the specific movements involved in goal-directed actions. Though we can be consciously aware of our body posture, the body schema guides our movements on a subconscious level, allowing us to direct our conscious attention to the task at hand without falling over or losing track of limbs outside our visual field.

The body schema also seems to be crucial for our ability to do things directly with our body, as is illustrated by deafferented subjects whose body schema is impaired: they have to consciously and visually direct their limb movements. (Cole and Paillard: 1995; Gallagher and Cole: 1995). In other words, we act intentionally with our bodies when performing simple tasks like grasping a cup or flipping a light switch. We do not need to form additional intentions in order to get from intending to doing; we do those things directly. Deafferented subjects, on the other hand, have to constantly act on their bodies, forming separate intentions for all the movements that together constitute their grasping a cup or flipping a switch in order to actually do so.

In humans, the body schema is quite plastic: it can easily incorporate artefacts as if they were elongations of the hand. This incorporation takes place very quickly, but only when the artefact is actively used as opposed to being passively held, and only for the functional length of the artefact (Farnè, Iriki, and Làdavas: 2004).

---

50 The body schema is commonly contrasted with the body image, a set of intentional states with the body as object, or how subjects feel and think about their body (Gallagher and Cole: 1995; Dijkerman and De Haan: 2007). That there exist subconscious processes subserving actions and keeping track of our posture and limb positions seems to be commonly accepted. It has been argued, however, that grouping these under the heading of ‘body schema’ and contrasting it just with the ‘body image’ yields concepts that are too coarse to be very useful, and that we should either clarify them or develop more, fine-grained distinctions (De Vignemont: 2007, 2010; Carruthers: 2008; Berlucchi and Aglioti: 2010).

51 We would need what Pacherie (2008) calls motor intentions, but these are not the kind of intentions that figure in Hornsby’s definition of teleologically basic actions.

52 I have borrowed the distinction between ‘acting on’ and ‘acting with’ from Johnson and Grafton (2003).
Can there be basic actions with artefacts?

2005). This might (partly) explain why it seems that we can sense stimuli in the tips of our artefacts, as I have mentioned in the last section. Thus, insofar as a functioning body schema is necessary for us to do things directly with our body, it also allows us to do things directly with artefacts.

Merleau-Ponty’s observations about artefacts phenomenologically extending the body can also be analyzed differently. Let me turn for that to a second phenomenological claim about artefacts: their transparency.

An agent’s body is a unique object in the sense that it is transparent to the agent when acting with it. It is *zuhanden* or available in the Heideggerian sense. In Merleau-Ponty’s words: “...My body itself I move directly, I do not find it at one point in objective space and transfer it to another, I have no need to look for it, it is already with me – I do not need to lead it towards the movement’s completion, it is in contact with it from the start and propels itself towards that end.” (1945/2002, 108). In Davidson’s terms: the agent moves external objects by means of moving her body, but her body she moves directly. It seems that transparency in this sense requires a functioning body schema: deafferented patients would likely not describe their body as it is for them in the way Merleau-Ponty does. This transparency of our body, moreover, seems to be what makes it possible that we can do things directly with it.

However, transparency is not limited to the body. In Merleau-Ponty’s example, the blind man’s cane ceases to be an external object, becoming instead part of his phenomenal body. Thereby, it becomes transparent to the blind man, allowing him to feel and act directly with it. The process of an artefact becoming transparent to an agent will usually be gradual: the blind man needs time to get used to his cane and become skilled in its use. The cane may never attain complete transparency in the way that our body parts do: it may extend our body and be incorporated in our body schema, but not change our sense of body-ownership, that is, give us the sensation that we ‘own’ it like we ‘own’ our body parts (De Preester and Tsakiris: 2009). Yet for my purposes that is not necessary. What is necessary is that artefacts can become transparent enough to do things directly with them. Based on both the phenomenological insights and psychological research quoted, it seems they can.
Verbeek (2000/2005) argues that an artefact can only become transparent to its user if three conditions are fulfilled: the artefact has the capacity to function as it should, otherwise it cannot become transparent to any user, the user needs to be skilled in the use of that artefact, otherwise no artefact of that type can become transparent to that user, and it should be possible to perceive the results in a suitable way, otherwise the user cannot know whether the action has actually succeeded. This last condition might not even be necessary for an artefact to become transparent, as Anscombe (1957/2000; see also Moran: 2004) has remarked that observation is an aid but not a determinant of action. For example, I may be happily writing notes while looking the other way, reading a text on my computer screen, and still experience my pen as transparent. Not unsurprisingly, those conditions seem to hold for body parts as well: if my arm is broken, or has been paralyzed all my life until recently, it will not be transparent to me. Consequently, I cannot do things directly with that arm.

The requirement of functionality draws attention to another aspect of transparency. As the functionality of an artefact always refers to a specific task that can be performed with that artefact, so is transparency always relative to a specific task: the body or an artefact is never transparent in itself. For example, my body may be transparent to me in grasping a cup or flipping a light switch, but not when I learn to dance the tango or do a challenging yoga pose. Similarly, my claw hammer may become transparent to me in hammering, but not in using it to wriggle and pull a broken nail out of a board. In short, we can say that our body is not always transparent to us in acting, and that it is not only the body that can become transparent to us in acting.

As Verbeek has argued, the notion of skill is important for transparency. Our body is usually transparent to us in action because we are highly skilled in moving our bodies: so skilled that we can automatically and swiftly execute an

---

33 Verbeek speaks about ‘availableness’, as opposed to ‘occurrentness’, terminology introduced by Heidegger (1927/1962) to describe the (potential) transparency of tools.

34 Verbeek uses a telescope and a microscope as examples. In these cases, the third condition is necessary as the standard action done with those artefacts is perceiving something. Still, it seems that this condition can be reduced to the first two: that the telescope / microscope has the capacity to function as it should, and that the user is skilled in using the telescope / microscope.
extensive repertoire of bodily movements without paying attention to or observing these movements. If we do something directly, not by doing anything else, by moving our bodies, it is because we are skilled in moving our body, and those skills have been accumulated gradually and over years of continual practice. On the other hand, a baby who grasps a toy but who does not yet have these skills can hardly be said to grasp it directly, but does so by reaching, groping and (over)compensating for unwanted arm movements.

One of the advantages of talking in terms of skill rather than transparency or doing directly is that the concept of skill is applicable to domains as diverse as the cognitive, the physical and the social, and has been well-investigated in the psychology of action. Looking explicitly at physical or motor skills, several features are commonly regarded as central to the concept. The most important is automaticity: performing a task you are skilled in implies automatic execution with little or no need for conscious attention to the mechanics (Summers: 2004). In its most extreme, it may become impossible to consciously attend to, let alone control, those mechanics. This constraint on our conscious control may be disadvantageous at times – unlearning of a bad posture or movement is a slow and difficult process. On the other hand, it has been argued that taking conscious control of the mechanics can severely tax working memory and make the action prone to disruption (Masters and Maxwell: 2004; Maxwell, Masters, and Van der Kamp: 2007; but see Montero: 2010). Highly automated movements are simply quicker, smoother and more precise, core features of motor skills (2007 APA Dictionary of Psychology). In experts this is characterized by low repetition variability: when repeated, the skilled movements of experts are always highly similar (Summers: 2004).

While several models of skill acquisition have been proposed, hierarchical models (e.g. Miller, Galanter, and Pribram: 1970; Schack: 2004) seem most promising to elucidate the philosophical concept of basic actions. Hierarchical models consist of levels, the highest of which contains our action plans, the one below that individual actions, then the movements involving those actions, down to the most basic sensorimotor control systems. The highest level is under our conscious control, the lowest is grounded in our physiology.
Basic actions start out low in the hierarchy, as the coarse movements of (groups of) muscles children first learn to control. Through practice, these movements are both automated and ‘chunked’ (Graybiel: 1998) or incorporated in ‘motor programs’ (e.g. Schmidt: 1991) on a higher level. Those chunks or programs can themselves be automated again, chunked or incorporated in higher motor programs, and so on. Conscious control is used on the higher levels, particularly to plan actions and to evaluate the outcome (Schack: 2004). Thus, any basic action requires a ‘bedrock’ of automated movements, which will be relatively small in young children and extensive and well-developed in experts. In this way, the notion of skill emphasizes that the class of basic actions is flexible and changes over time for any person. Deafferented patients are an exception to this rule: they do not seem to acquire new motor skills, which suggests that a functioning body schema is necessary for motor skill acquisition (Cole and Paillard: 1995). In short, just as we have to learn to ‘use’ our bodies, we have to learn to use artefacts. And just as we can become skilled in moving our bodies, we can become skilled in using artefacts - skilled enough in using them so that they become transparent to us in use, or phenomenologically extend our bodies, and thus, allow us to do things directly with them.

What does this mean for the metaphysical question of what actions are? The easy answer would be that, rather than being bodily movements, it turns out that actions are either bodily movements or skilled tool use, but this would beg the question of what is common about them. For the moment, we have simply lopped the ‘mere movements of the body’ off Davidson’s definition, which means that we are back at: ‘what we do directly, not by doing something else’ and Hornsby’s refined definition of teleologically basic actions as discussed in chapter 3.3. While this definition is admittedly less clear and concise than ‘bodily movements’, it does allow us to use both the phenomenology of embodied action

---

55 This analysis does not seem to hold for all artefacts. For example, though glasses can become transparent in use, we do not speak of ‘skill in using glasses’. We do speak, however, of ‘getting used to glasses’, which in this case roughly means the same. Note that this is not because the primary function of glasses is to aid perception: one can become skilled in using artefacts such as a blind man’s cane or a microscope. The difference rather seems to lie in ‘passive’ versus ‘active’ use.
and that of tool use to investigate what actions are. Also, I have suggested how
the notion of skill might help us to better understand what is involved in doing
things directly, and thus, how we should conceive of actions.

Before ending, let me make a brief remark on the notion of direct control,
and whether it can play a useful role in this discussion. Skills concern tasks (e.g.
one can be skilled in fencing, or playing the piano), while control is related to
objects or entities (e.g. one controls one’s body or one’s sabre). In order to
become skilled in a task, it seems one needs a certain measure of control over
the object with which the task is performed, be it a body part or an artefact. The
advantage of skill, then, is that it does not distinguish between tasks performed
with or without artefacts. It does mean that skill cannot distinguish between
tasks requiring direct control and tasks requiring indirect control: one can
indeed be skilled in flying model planes. Thus, while the concept of skill may
explain why basic actions can be done with artefacts in so far as they are the
actions you do directly, or do not do by doing something else, it applies both to
tasks requiring direct and indirect control. However, as I have argued in the
section 3.5, the notion of direct control can on its own not make a distinction
between the body and some of the artefacts that phenomenologically extend it, so
even without the notion of skill a conception of actions based on direct control
does not give us the doctrine that all actions are bodily movements. Clearly, there
is need for a philosophical investigation into whether there is a meaningful
distinction between direct and indirect control, and if so, what direct control
would amount to, but this is beyond the scope of this chapter.

3.7. Conclusion

This chapter has argued for two claims. First, that basic actions as defined by
Davidson can be causally or teleologically basic, but that no matter which kind of
basicness one endorses, basic actions are never necessarily bodily movements
interpreted in a narrow sense. Rather, they can also be prior to bodily move-
ments in an account of causal basicness, or be events of which the bodily
movements are only a part in an account of teleological basicness.

Second, this chapter has argued that a teleological notion of basic actions will
include some actions that are performed with artefacts. I have analysed the
relation between artefacts and the body and shown that teleologically basic actions can encompass the manipulation of artefacts, provided the artefacts have the capacity to function as they should and the user is skilled in using them.

These claims have two general implications for further research. First, they emphasize the importance artefacts play in our everyday actions and show how action theory can benefit from insights from the phenomenology and psychology of tool use. Second, by questioning the popular claim that basic actions are bodily movements, they might well pose a challenge for the Davidsonian causal theory of action. It is undeniable that brain activity causes bodily movements. Davidson has argued that we can make a reliable connection between mental states (including reasons) and brain states through the doctrine of anomalous monism. Furthermore, he has argued for the doctrine that basic actions are bodily movements. Given both doctrines, we can translate the claim ‘brain activity causes bodily movements’ into ‘reasons cause actions’, which lies at the heart of the Davidsonian causal theory of action. The validity of the doctrine of anomalous monism is already debated (e.g. Honderich: 1982; Hutto: 1998), as is the equation of reasons with mental states (Dancy: 2000a; Alvarez: 2008). This chapter has contested the doctrine that basic actions can be equated with bodily movements, suggesting another possible challenge to the Davidsonian causal theory of action.

With the defense of the action identification thesis and an investigation of basic actions behind us, it is time to answer the first part of the main question: What is an action with an artefact? Based on the previous chapters, my answer is the following. An action is a doing, that is done by doing something else or done directly, that is intentional under some description. An action with an artefact is then a doing, that is done by doing something else or done directly, that is intentional under at least one description that describes the action in terms of artefact use. This description of the action may be in terms of its consequences: this is the case if the artefact is intentionally used by doing something else (say, flipping the switch by moving my hand in a certain way). In chapter 3, however, I have argued that this description of the action may also be the teleologically most basic description: this is the case if the artefact is intentionally used directly, not by doing something else (say, hammering a nail into a board or prodding an
Can there be basic actions with artefacts?

obstacle with a blind man’s cane.) For this, the artefact has to become transparent in use, which seems to require incorporation of the artefact in the body schema.
4 Describing affordances: the nested affordances model

4.1. Chapter abstract

Artefacts offer opportunities for behaviour, ‘affordances’, that can be described on a low, manipulation level (‘flipping a switch’) as well as on a high, social activity level (‘winning a war’), and several levels in between. However, design research into (perceived) affordances has not distinguished those levels of description. This chapter addresses the issue how affordances in design are connected on different levels. Its main claim is that affordances are nested, that is, that many different descriptions can apply to a single affordance. In this, I argue, affordances are similar to actions: just like I can turn on the light by flipping the switch, we can say that the switch affords turning on the light by affording flipping.

On this claim a nested affordances model is built, specifying what mechanism and knowledge the artefact user would need to be able to perceive affordances on each level. The chapter also shows several ways in which the nested affordances model can contribute to affordance-based design. Apart from using the action identification thesis from chapters 2 and 3 to clarify the affordance concept, at the end of this chapter I will also examine the implications of my claim from chapter 3, that there can be basic actions with artefacts, for affordance-based design.

An earlier draft of this chapter has been submitted as a paper to Design Studies. The editor has requested a revision and resubmission of that paper. An earlier draft of this chapter has been presented at the Conference of the Society for the Philosophy of Technology 2009.
4.2. Introduction

It is obvious that we can perceive objects and their properties, but it has also been argued that we can directly perceive opportunities for behaviour (Gibson: 1979). Gibson calls these opportunities for behaviour affordances, and states that they are ubiquitous in our everyday environment: a single chair can afford sitting on, standing on, throwing, etc. Not only has Gibson’s claim been supported by various experiments in ecological psychology (e.g. Milner and Goodale: 1995; J. Norman: 2002), designers have found it very useful to work with the concept of affordances, and have investigated what ensures that affordances of artefacts will be perceived (D. Norman: 1988/2002). This is an important issue, since only the perceived affordances of an artefact can influence the behaviour of the intended user.

However, assuming that we can perceive affordances in artefacts, two questions arise. The behaviour of people can be described in different ways and on different levels (Anscombe: 1957/2000; Vallacher and Wegner: 1989; Michaels: 2003). Low-level descriptions contain mostly context-independent terms like making bodily movements, reaching, grasping, etc. High-level descriptions, in contrast, situate such behaviour in its context: reaching for someone may be playing tag in one context, apprehending a suspect in another. In between are various levels of descriptions that are context-sensitive to some degree. The first question is then: on which levels can we describe the affordances of artefacts? For example, a coffee cup filled with coffee seems to afford grasping, drinking coffee, and getting energized: is each a different affordance of the same cup, or a different description of the same affordance? And the second question is: what mechanism and knowledge do we need to perceive what an artefact affords on each level of description?

---

37 In this dissertation I will concentrate on visual perception and to a lesser degree on tactile perception, as vision and touch are the senses that can with most accuracy locate objects and their affordances relative to the observer.

We seem to subconsciously perceive basic manipulation opportunities, as Gibson (1979) and J. Norman (2002) use the terms, but we also perceive possible technical (D. Norman) and social (Searle: 1996) uses of artefacts. Not only is there currently no attempt to connect such approaches, but the literature also shows that different fields use the affordance concept to denote opportunities for behaviour on different levels of description. Ecological psychology emphasizes that affordances are directly perceivable, which is helpful for research in perception, but constrains the application of the affordance concept to opportunities for behaviour under low-level descriptions. D. Norman, in contrast, allows affordances to be opportunities for behaviour under high-level descriptions by stating that the perception of affordances may require interpretation and background knowledge. Thus, Norman’s interpretation of the affordance concept comes at the cost of severing the ties with ecological psychology. This chapter will help to connect the fields of design sciences and ecological psychology by showing how Norman’s perceived and interpreted affordances relate to directly perceivable Gibsonian affordances, allowing both fields to profit from each other’s findings. For example, it can enable designers to use methods from ecological psychology in order to determine which artefact affordances would make the relevant actions with that artefact as efficient and comfortable as possible for the intended user (see e.g. Warren: 1984; Shaw, Flascher and Kadar: 1995). Likewise, it can give ecological psychologists a better grip on case studies from design for establishing what kind of affordances are perceived under certain circumstances.

The main question this chapter addresses is a combination of the two questions I mentioned earlier: on which levels can we describe the affordances of artefacts, and what mechanism and knowledge do we need to perceive what an artefact affords on each level of description? I will address this question in several steps. In section 4.3, I will argue that affordances are similar to actions in that they are nested: just like any given action can be described in many different ways, some of which might be intentional and others not (Anscombe: 1957/2000; Davidson: 1980: essay 3), any given affordance of an artefact can be described in many different ways. And just as action descriptions are grounded in basic actions, affordance descriptions can be said to be grounded in basic affordances. In
Acting with Artefacts

section 4.4, I will capture this idea in a model, the nested affordances model. I will argue that we can describe affordances of artefacts on four levels: how the artefact can be manipulated, what the reliable effects of those manipulations will be, what can be done with the whole artefact in itself, and what can be done with the whole artefact as component of a socio-technical system. I will also investigate what psychological mechanism and knowledge is needed to perceive what an artefact affords on each level of description. In section 4.5, I will give a general recommendation for design based on perceived affordances, show how the nested affordances model can ground and specify existing recommendations like ‘give feedback’, and how it can extend the scope of affordance-based design to socio-technical systems. Finally, in section 4.6 I will connect this chapter to the last and examine the consequences of my claim that there can be basic actions with artefacts for affordance-based design. I will argue that, as the theory of affordances does not presuppose specific capabilities on the part of the actor, it is well-suited to deal with both humans and humans wielding artefacts, or human-artefact systems.

4.3. Affordances under a description

Suppose I startle a burglar (unintentionally) by turning on the light (intentionally) by flipping a switch by moving my hand in a certain way. How many actions have I done? The philosophers Anscombe (1957/2000) and Davidson (1980: essay 1) hold that there is only one, basic, action here: moving my hand. This action can be described under all kinds of descriptions, some intentional, others not. By moving my hand, I do all of these things, or bring about all of these effects. Moving my hand is the basic action here, since I do not (intentionally) do it by doing something else.39

As affordances are opportunities for action, it seems only natural to mirror them to the structure of actions. In that case, there would be basic affordances that

39 See chapter 3 for criticism of the claim that all (basic) actions are bodily movements.
afford basic actions. Those affordances would then afford actions under all other kinds of descriptions, by affording these basic actions. The light switch affords a basic action, namely, moving your hand against it in a certain way. The light switch affords actions under other descriptions as well: turning on the light, startling the burglar, etc. These, then, would all be the same affordance under different descriptions. Of course, the light switch in itself does not afford turning on the light: it needs to be connected to electric wiring and a light bulb. The same holds for basic actions: my action of moving my hand is not an action of turning on the light if the light switch is disconnected. However, given that the light switch is connected, the one action is the other, since the agent does not need to do anything else to turn on the light beyond moving his hand against the switch in a certain way. Similarly, the one affordance is the other, since the light switch does not have to afford anything else to the agent beyond affording the moving of his hand against the switch in a certain way, in order to afford the turning on of the light. What is important here is that the switch affords these actions under various descriptions by affording the basic action. If the switch does not afford the moving of your hand against it in a certain way, e.g. because it has broken off, then it does not afford flipping, turning on the light, startling the burglar, etc.

A potential worry with the idea of basic affordances might be the discrepancy between actions, that are inherently intentional, and affordances as affording behaviour: we might wish, as Gibson does, to say that the environment also has affordances for simple animals that might not have intentions in the philosophical sense of the word. This worry could be addressed in line with a proposal by Scarantino (2003), by letting affordances afford ‘goal-directed behaviour of the organism’ rather than ‘intentional actions’. However, as I will talk only about affordances in design for humans, I will not explore this option further here.
Another worry might be that the affordance under the lowest description might guide unintentional aspects of the basic action, rather than the basic action itself, for example when the shape of a doorknob affects the aperture of a grasping hand. Here, however, an aspect of the affordance is described rather than the affordance itself. What we could say regarding this example is that certain characteristics of objects (the shape and size of the doorknob) influence components of action (the aperture of the hand) (e.g. Ellis and Tucker: 2000). However, matching the aperture of the hand to the shape and size of the doorknob is not a basic action, as it is not something that we do intentionally (see e.g. Milner and Goodale: 1995). Nor do we grasp the doorknob by just opening our hands to a certain degree. Rather, the opening of our hands is part of our grasping the doorknob.

In this section I have proposed that affordances mirror actions in the sense that if an artefact affords something by affording something else, there is only one affordance, the basic affordance, and this affordance affords a basic action. This basic affordance can be described in various ways, however, as it affords actions under various descriptions. The next step is to identify the most important levels of description on which the affordance can be described, and investigate what mechanism and knowledge the perceiver needs in order to perceive the affordance on that level of description. I will explicate this in the nested affordances model. For clarity I will talk about ‘perceived affordances’ on each level, though what the user actually does is perceive an opportunity for an action, under a certain description, that is realised by the presence of the basic affordance and its specific context.

4.4. The nested affordances model

In this section I will argue that there are four main levels on which we can describe an affordance, and show how these levels connect.

In the previous section, I argued that there are basic affordances that correspond to (intentionally) basic actions. These affordances I will call manipulation opportunities. They are affordances at the first, lowest level of description, as they
afford intentional actions at the lowest level of description: grasping, pushing, etc. As such, they map nicely onto the original Gibsonian concept of affordances that can be directly perceived and fit the neuropsychological evidence for the direct perceivability of affordances (J. Norman: 2002). In practice, we rarely encounter completely unknown artefacts, but if we do, the only affordances of those artefacts we will be able to perceive will be manipulation opportunities: can we stand on them? Push them? Pull at some parts, or twist them? We will then, of course, need to experiment with the artefact in order to find out whether those perceived affordances are real affordances. Just as we can only find out if we can do something new by trying to do it, we can only find out if some unknown artefact affords X-ing by trying to X with or on the artefact. Perceived affordances can tell us where to start experimenting, and what actions we might try out first.

By experimenting, or gaining knowledge about the artefact in some other way, we might discover reliable connections between manipulations and their effects. We might find out that pulling the trigger of a loaded gun will reliably fire it, or that pressing the ‘A’ key on the keyboard of our computer will make an ‘A’ appear in our text editor. In short, we learn what the effects of our manipulations are. I will call perceived affordances on this level effect opportunities. Although abstract knowledge can help you in learning effect opportunities – the ‘A’ printed on the key that will make an ‘A’ appear in your text editor is there for that reason, effect opportunities can be directly perceiveable. All that would strictly be needed for an observer to directly perceive effect opportunities would be to have a simple psychological mechanism that can correlate causes and effects.

According to You and Chen (2007), it is the task of product semantics to link affordances to their intended functions, via design but also via labels, symbols, etc., removing the need for experimentation. The nested affordances model shows how exactly product semantics is related to affordance design: it allows the user to perceive effect opportunities, or higher-level descriptions of basic affordances.
Affordances can also be perceived on a third level, the level of what the user can do with the artefact, rather than how to act on it. I will call these perceived affordances use opportunities. What the user can do with an artefact on this level often depends on a number of basic affordances: just like goal-directed actions can be strung together in goal-directed plans, as when cooking a recipe, a number of basic affordances of an artefact taken together can enable a whole artefact to afford an action under a high-level description. My computer affords writing a paper because its affordances realise effects (typing an ‘a’, deleting a letter, etc.) that contribute to the use opportunity of the whole. In other words: by affording text insertion, deletion, etc., my computer affords writing a paper. Perceiving use opportunities seems to be what Searle (1996, 4) meant when he said that we often perceive objects as functional, for example, that we see bathtubs rather than enamel-covered iron concavities containing water. Sometimes, perceiving the context or socio-technical system of which the artefact is a part is necessary to perceive its use opportunities (e.g. Vaesen: 2008, 56).

It does seem that we need mental representations or abstract knowledge in order to be able to perceive use opportunities; they require too much interpretation or background knowledge to be directly perceivable. Norman argues that the designer needs to provide us with the information needed to construct a correct mental model of the artefact, that is, knowledge of what effects our actions will have on the functioning of the whole artefact. Houkes and Vermaas (2004; see also Houkes, Vermaas, Dorst, and De Vries: 2002) argue that the designer should communicate the use plan of the artefact to the user, that is, knowledge that a certain sequence of actions will lead to the realisation of a goal. The designer also needs to communicate information regarding the proper context of use together with the use plan, as well as the need for auxiliary items, if any, and whether the user would need special skills in operating the artefact. It is possible for users to construct use plans themselves rather than relying on one provided by the designer (if any is provided), but that would involve gaining knowledge about the artefact as well, e.g. by experimenting.41

41 See chapter 8, especially 8.4, for a closer look at the use plan analysis.
Finally, affordances of artefacts can also be perceived on a fourth level. This is because artefacts rarely operate in isolation: they are usually part of larger socio-technical systems that themselves have specific functions (Kroes, Franssen, Van de Poel, and Ottens: 2006; Ottens, Franssen, Kroes, and Van de Poel: 2006). Gibson (1979, 139) seems to acknowledge this when he writes: “...the real postbox (...) affords letter-mailing to a letter-writing human in a community with a postal system. This fact is perceived when the postbox is identified as such...” Just as an artefact can offer a use opportunity by virtue of having a particular set of affordances, it can also offer an opportunity for action under a still higher-level description by virtue of being part of a system consisting of artefacts, humans, institutional arrangements, etc. I will call perceived affordances on this fourth level activity opportunities. These would correspond with social actions in action theory: raising your hand is asking a question if the social/institutional setting is that of a talk, meeting, etc. A debit card reader affords conducting a payment, and passing your debit card through the reader is conducting a payment, provided all kinds of other artefacts and institutional arrangements are in place. We need abstract social and institutional knowledge in order to perceive activity opportunities.

In this section I have explicated the nested affordances model and argued that there are four main levels on which affordances can be perceived. An overview of the model can be found in Table 1.

Let me add that not every artefact will have different perceived affordances on each level. Complex technical artefacts such as computers and cars will have them, but for simple artefacts, some distinctions between levels may collapse. For example, while a spoon can be manipulated in different ways, manipulating the spoon always means manipulating the whole artefact, so there will be no use opportunities that are not also effect opportunities or manipulation opportunities. This redundancy, however, is necessary to accommodate the higher-level affordances that more complex artefacts have.
Acting with Artefacts

<table>
<thead>
<tr>
<th>Affordance</th>
<th>Corresponding concept action</th>
<th>Mechanism and knowledge needed</th>
<th>Examples of actions afforded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunity for manipulation</td>
<td>Basic action</td>
<td>Neuropsychological mechanisms (Milner and Goodale: 1995; J. Norman: 2002)</td>
<td>Pulling a trigger, hitting a glass pane, pressing a button...</td>
</tr>
<tr>
<td>Opportunity for use</td>
<td>Plan</td>
<td>Mental models (D. Norman: 1988/2002); use plans (Houkes and Vermaas: 2004)</td>
<td>Shooting a person, obtaining an emergency hammer, writing a paper...</td>
</tr>
<tr>
<td>Opportunity for social activity</td>
<td>Social action</td>
<td>Abstract, institutional and social knowledge (Searle: 1996; Kroes et al.: 2006; Ottens et al.: 2006)</td>
<td>Murdering an enemy, escaping a crashed vehicle, working out a psychological theory...</td>
</tr>
</tbody>
</table>

Table 1: The nested affordances model

4.5. Applying the nested affordances model

In this final section I will show three ways in which the nested affordances model can contribute to affordance-based design. First, I will give a general recommendation for design based on perceived affordances and show how this recommendation should be realised using the nested affordances model. Second, I will show how the nested affordances model can ground and specify
design recommendations like ‘give feedback’. Third, I will show how the nested affordances model can extend the scope of affordance-based design to socio-technical systems.

On the most general level, an affordance-based theory of design can give either a ‘hard’ or a ‘soft’ recommendation. The ‘hard’ recommendation would be that any artefact should have all the right / desired affordances, and none of the wrong / unwanted / dangerous ones. Maier and Fadel (2009) make this recommendation when they state that design is the specification of a system structure that possesses certain desired affordances in order to support certain desired behaviors, but does not possess certain undesired affordances so as to avoid certain undesired behaviors (23). However, Maier and Fadel describe an ideal situation that is unattainable in practice. The nested affordances model shows why: sometimes an affordance that allows for a desired use opportunity also allows for an undesired one. For example, the sharp edge of a knife affords cutting bread, but also cutting yourself. The ‘soft’ recommendation for design is more practical, as it acknowledges this, and puts the emphasis rather on perceived affordances:

Good design based on perceived affordances is the specification of a system structure that does possess certain desired affordances in order to support certain desired behaviors, and that ensures that those desired affordances will be perceived by the intended users. It does not possess undesired affordances whenever possible, but if undesired affordances are inevitable, it ensures that those undesired affordances will not be perceived by anyone in order to avoid certain undesired behaviors.

An added strength of this recommendation is that it is asymmetrical: only the intended user has to perceive the right affordances, while the undesired affordances should not be perceivable by anyone. For example: a chainsaw does not have to be designed so that a child can see how it works, but it should be de-
Acting with Artefacts

signed so that no child can see a possibility to seriously harm itself or others with it.\[42\]

The nested affordances model shows how to realise this recommendation. If a certain system structure should possess an affordance described on a high level, say, a use opportunity, this can only be realised by ensuring the system a) has the proper affordance(s), and b) that the system structure is such that this affordance (these affordances) can be redescribed as the desired use opportunity. For example, if an artefact should afford ‘lighting the room’, it should afford a manipulation like ‘flipping a switch’, and the artefact should be built so that the light comes on when the switch is flipped. In other words: the user should be able to light the room by flipping the switch. The next step is to make sure c) that the use opportunity can be perceived, so that the user can see what can be done with the artefact, and d) that the manipulation opportunity can be perceived, so that the user can see how to do that. Note that it is possible to have c) without d): I might perceive the light bulb and know that it can be used to light the room, but if I do not manage to find the switch, I will not be able to actually use the light.

In the remainder of this section I will show how the nested affordances model can ground and specify existing recommendations for design like ‘give feedback’. I will also show how the nested affordances model can extend the scope of affordance-based design from technical artefacts to socio-technical systems.

---

\[42\] In Gibson’s (1979) description of affordances as opportunities for an animal in an environment, objects can afford both actions (have goal affordances) and events (have happening affordances; Scarantino: 2003). For example: a fire affords cooking meat (goal affordance) but also burning yourself (happening affordance). As this dissertation is concerned only with goal affordances, the recommendation is to hide them so as not to encourage or suggest inappropriate use. Undesired happening affordances, on the contrary, should be made clearly visible: if the artefact affords harming the user, the user should know. Though happening affordances are currently underexplored in affordance-based design, their treatment is outside the scope of this dissertation.
In order to perceive effect opportunities, the user should be able to register a reliable connection between her actions and the resulting behaviour of the artefact. Therefore, the artefact’s behaviour should be predictable, and the user should be given feedback on the effects of her manipulations. Norman emphasizes the importance of giving immediate and obvious feedback. This can tell the user what result has been accomplished and ensures that the user can determine the relationship between actions and their effects on the system.

The nested affordances model adds that feedback is necessary to go from perceiving manipulation opportunities to perceiving effect opportunities. Also, the model can offer a more specific recommendation. The user should not just receive feedback, but she should receive feedback on different levels. At least, the user should receive feedback that the manipulation was successful. For example, when pressing a button, there is direct feedback: the button is pushed backward (or downward). This direct feedback is absent in for example touchpads, where the user has to rely on indirect feedback to perceive that the action was successful (a light, a tone, etc.). Where direct feedback is important for knowing that a manipulation is successful, indirect feedback is important for knowing that a successful manipulation is also a successful use of the artefact. Good pedestrian traffic lights, for example, offer feedback on all relevant levels. They have buttons that are pushed inwards when pressed. They have lights near the button to show that the button press was successfully registered and constituted a successful use of the artefact. As Norman says, immediate feedback is important here. The light turning green is delayed feedback that the action was succesful, but until the feedback occurs, the user cannot distinguish between delayed feedback on a successful action and no feedback on an unsuccessful one. Finally, other lights turn red to stop the traffic, showing that the socio-technical system works as it should. Since with a traffic light pressing the button is operating the traffic light, that is, there is no distinction between effect opportunities and use opportunities, the light turning green is both feedback on the successful action and on the successful use of the artefact.

Intimately related to the recommendation of giving feedback is the recommendation to give feedforward, or to provide the user with clues as to what effect will result from a specific manipulation. This can be a design solution, for
example, placing the control (say, a light switch) close to the function (the light). It can also be symbolic, like putting a symbol or a label on a switch. Norman argues that if design depends on labels, it may be faulty (78). The nested affordances model shows that this depends on the level that the label announces an opportunity on. It may be difficult to find an intuitive design alternative for a label that announces an effect opportunity, for example, a light switch with a picture of a glowing light bulb on it. However, the label can also announce a manipulation opportunity, say, a light switch with the word ‘press’ on it. Here, Norman’s recommendation holds, for the information denoting a manipulation opportunity should be made clearly available, not hidden behind clumsy or ‘artful’ design. Following Gibson, we can say that the (basic) affordances should be fully specified in the stimulus information (1979, 140).

Knowing what an artefact does in itself is good, but no artefact works in isolation. Artefacts are connected to humans and to each other, their use and behaviour regulated by institutions and social procedures. In other words, many artefacts are part of socio-technical systems, systems that themselves have intended functions and need to contain artefactual, human and institutional parts in order to fulfil those intended functions (Kroes et al.: 2006). An example would be a luggage monitoring service at an airport that requires an artificial scanner, a human operator and regulations concerning luggage handling and safety in order to fulfil the service’s intended function within the larger civil aviation system. Given this additional context of the scanner, the scanner affords actions under new descriptions that take the socio-technical system in account. For example, as an artefact, the scanner affords ‘highlighting certain substances in someone’s luggage’. As part of a socio-technical system, the scanner also affords ‘detecting bombs’, ‘improving the safety of plane travellers’, et cetera. Given that many artefacts (like scanners) are designed to function within the context of specific socio-technical systems, it is important for the designer to show how the affordances of the artefact as it is relate to the affordance(s) the artefact has by virtue of the socio-technical system to which it belongs.

While Norman does not concern himself with socio-technical systems or the place artefacts have in them, building a bridge between artefacts as artefacts and
Describing affordances: the nested affordances model

artefacts as parts of socio-technical systems is a natural step in the nested affordances model. Strictly speaking, it may not be the responsibility of the designer of the artefact to show what activity opportunities the artefact affords by virtue of belonging to a certain socio-technical system, but rather that of the designer of the socio-technical system. Still, it might have distinct advantages to do so, if only for marketing purposes. People do not buy modems because they afford modulating and demodulating analog signals to encode and decode digital information, though that could be said to be the technical function of a modem. Rather, people with computers buy modems because they can be made a part of a socio-technical system and then afford emailing, surfing the internet, etc.

The necessity to show how the use opportunities of artefacts relate to their activity opportunities is even greater with social artefacts. The use opportunities of technical artefacts usually contribute causally to the activity opportunities they offer as part of the socio-technical system to which they belong: the design of cars, trains and roads has a direct influence on the (quality of the) activity opportunities they offer within a transportation system. This is not the case with social artefacts (banknotes, wedding rings, etc.). Here, the activity opportunities artefacts offer within the system (the financial system, the institution of marriage) might be only very loosely related to the affordances of the artefacts themselves. It might be tempting to say that the function of those objects is symbolic, and that the relation between what the artefact affords and the activity opportunities it offers as part of the system is totally arbitrary: a coin is money as well as a banknote. This might be the case in theory. However, we only have to think of the Ningi to see that it is not true in practice.

---

43 A means of payment from Douglas Adams’s Hitchhiker’s Guide to the Galaxy, the Ningi is a triangular rubber coin 6 800 miles along each side. The absurdity of this idea illustrates that we can at least put some constraints on social artefacts. In the case of physical money: that it should afford carrying, that it should not afford melting or breaking quickly, etc.
Acting with Artefacts

In this section I have given a general recommendation for design based on perceived affordances. I have also shown how the nested affordances model can ground and specify existing recommendations for design. Finally, I have shown how the nested affordances model can extend the scope of affordance-based design from technical artefacts to socio-technical systems.

4.6. Acting with artefacts on affordances

Those who have read this chapter immediately after Chapter 3 may be left with a question relating both chapters. In this chapter I have argued that affordances mirror actions in the sense that there are basic affordances, and that objects may afford something by affording something else – a light switch connected to a lamp may afford the lighting of the room by affording the flipping of the switch. We can speak of a basic affordance whenever a basic action is afforded. The possibility of basic actions with artefacts, however, seems to allow for basic affordances that can only be utilised by humans using artefacts, and this would be strange if affordances would be limited to those action possibilities of the human body.

Fortunately, affordance theory is explicitly not restricted to action opportunities for the human body. For example, previous research has looked at how broad doorways should be for wheelchair users to perceive those doorways as passable, or affording passage (Shaw, Flasher and Kadar: 1995). Indeed, artefacts both offer affordances to the user (one can ‘act on’ an artefact), and present the user with new affordances in the environment, or extend the user’s capacities to act (one can ‘act with’ an artefact; see Johnson and Grafton: 2003). For example: a hammer affords grasping for its user due to its handle. A nail affords fastening objects to surfaces – but only for a hammer user. Thus, artefacts can change the affordances present in our environment: the environment can offer new opportunities for action to a human with a hammer. The environment can also lose affordances, however: a heavy box may afford lifting for me if I happen to have two free hands, but not if I happen to have only one hand free because I am holding a hammer.

In this line of thought, Turvey and Shaw (1979; see also Michaels: 2003) proposed the general notion of ‘effectivities’ to denote the combination of
physical and mental capacities or abilities an organism needs to have in order to act on a certain affordance. Effectivities may include both (control over) body parts and tools, and so illustrate the flexibility of the affordance concept. Another proposed extension of the affordance concept has been to propose object-object affordances rather than agent-object affordances.\(^4\) For example, one cogwheel can be said to afford turning to another cogwheel, provided they are sufficiently close to each other (Maier and Fadel: 2009). Whether useful or not, this application of the affordance concept takes it far from the field where it was conceived, the psychology of perception. Such extensions would also sever the link between basic actions and basic affordances that I have proposed, as objects cannot act.

Given that affordance-based design can deal relatively easy with human-artefact systems, it could be useful to investigate for anyone who both works on design and takes a system perspective. I will not go into the possibilities here, as I have already mentioned various problems that arise for those who would take such a perspective in chapter 3. For those who would face those problems, various examples can be found in this chapter that could serve as a starting point for investigation. For example, Gibson’s mailbox that affords mailing a letter to a letter-writing human in a community with a postal system in section 4.4, or my modem that affords emailing to a human with an operational computer in section 4.5.

The most important problem when trying to apply affordance-based design to systems of increasing complexity would probably be the increased difficulty of affordance perception. One has to be aware in some way of both one’s effectivities and the environment in order to perceive actual affordances in that environment, and the more effectivities one has at one’s disposal, and the more complex the environment, the harder this gets. Fortunately, affordance perception itself can be facilitated by using technology (maps, glasses, etc.). Unfortunately, perception aids may reveal some affordances and occlude others (Ihde: 1990). The choice

---

\(^4\) Besides object-agent affordances, which would be similar to the happing affordances described in an earlier footnote, Maier and Fadel also mention agent-agent affordances. These would be the opportunities for action that agents can offer each other, ranging from the physical (using a human as a living shield) to the social (one’s neighbour may afford gossiping) and everything in between.
Acting with Artefacts

which affordances will be revealed and which will be occluded is at best up to the designers rather than the users, and at worst is not a choice at all, but merely an accidental design feature.

4.7. Conclusion

In this chapter I have asked the question: on which levels can we describe the affordances of artefacts, and what mechanism and knowledge do we need to perceive what an artefact affords on each level of description? I have addressed this question in two stages. First, in section 4.3, I have argued that affordances are like actions in the sense that there are basic affordances, and those can be described in different ways. Second, in section 4.4, I have worked out this idea in the nested affordances model, a structured account of the levels on which we can describe affordances. The nested affordances model states that we can describe affordances of artefacts on four levels: that of the manipulation opportunities, that are very basic, Gibsonian affordances; effect opportunities, that describe affordances in terms of the effects of their manipulations; use opportunities, that describe affordances in terms of their effect on the whole artefact; and activity opportunities, that describe affordances in terms of their effects on the particular socio-technical system the artefact belongs to. I have also investigated what mechanism and knowledge we need in order to perceive what an artefact affords on each of these levels of description. In section 4.5 I have shown how the model can give a general recommendation for design based on perceived affordances, as well as specify known design recommendations and extend the scope of affordance-based design to socio-technical systems. Finally, in section 4.6 I have shown how my claim from chapter 3, that there can be basic actions with artefacts, fits into affordance theory.

Gibsonian affordances are very useful within ecological psychology, while Norman’s adjusted concept of perceived affordances is very useful for design. By introducing the concept of basic affordances and creating the nested affordances model, I have shown how both concepts relate to each other, building a bridge between design and ecological psychology. This provides an opportunity for synergy between two fields that have much to offer each other.
5 Artefacts and reasons for action

5.1. Chapter abstract

Chapters 2 to 4 have addressed the first part of the main question: What is an action with an artefact? In particular, in chapter 2 I have defended the action identification thesis against the time of a killing, in chapter 3 I have argued that there can be basic actions with artefacts, and in chapter 4 I have argued that affordances are nested, that is, that multiple descriptions can apply to the same affordance, just like multiple descriptions can apply to the same action.

Chapters 5 to 7 will address the second part of the main question: How do artefacts influence agents and agency? Chapter 5 will show how artefacts can influence agents by altering their reasons for action. Chapter 6 will show how artefacts can influence agency by affecting us in our capacity as bounded practical reasoners. Chapter 7 will show how artefacts can influence collective actions and agency by offering a way to strengthen our commitment to collective actions.

Artefacts can influence agents and agency in several ways. They can be instruments, enabling and facilitating actions, changing what we can do while leaving the choice up to us. However, they can also change what we actually do. Analytic philosophy of technology has generally been concerned with the former, while continental philosophy of technology has been more interested in the latter. Despite the fact that each field has its share of successful and interesting theories, there are few attempts to incorporate insights from one field into the framework of the other.

In this chapter I develop an analytic account of how artefacts can influence what we actually do that is inspired by investigations into this phenomenon in

An earlier draft of this chapter has been submitted as a paper to Inquiry. The editor has requested a revision and resubmission of that paper. An earlier draft of this chapter has been presented at the Philosophy & Ethics seminar of Eindhoven University of Technology.
Acting with Artefacts

continental philosophy of technology. Specifically, it accounts for this influence in terms of the way artefacts alter our reasons for action. In developing this account, I will build on Dancy's (2000a) account of practical reasoning. Not only can a reasons approach tell us how artefacts influence agents, it also helps us in analysing why artefacts sometimes fail to influence agents, contrary to designer expectations or intentions.

5.2. Introduction

When it comes to affecting human agents, it seems artefacts can play two roles. In their first role they can enable or facilitate human actions. Artefacts make it possible for us to do things we would not otherwise be able to do, and thereby adopt new goals, or help us do things we would otherwise be able to do, but in more time, with greater effort, etc. (Houkes and Vermaas: 2004; Illies and Meijers: 2009). In this role, artefacts are instruments, means that can be used to achieve a certain end. In short, artefacts in this role alter what people can do. What people actually do, which artefacts they use for which ends, is up to them.

Artefacts can also play a second role, where they influence, guide or mediate human actions. In this role, artefacts alter what people actually do. Authors like Akrich (1992) and Latour (1992) have argued that artefacts can prescribe actions to us. Following their lead, Verbeek (2000/2005) has argued that artefacts mediate actions in part by inviting or inhibiting us to act in a certain way. For example, speed bumps can be said to prescribe slowing down to the oncoming driver, while SUVs seem to invite reckless driving.

If we compare investigations into these roles by analytic and continental philosophy of technology, there is a marked difference in focus. Analytic philosophy of technology tends to look at the first role, analysing e.g. functionality (Kroes and Meijers: 2006), use know-how (Houkes: 2006) and (instrumental) goodness (Franssen: 2006), phenomena that are all relevant for studying how artefacts enable and facilitate actions. Continental philosophy of technology, on the other hand, is mostly concerned with the second role, analysing how artefacts change the way we see the world (Ihde: 1990), act in it (Verbeek:  

78
Artefacts and reasons for action

2000/2005), or on a larger scale, how artefacts help shape scientific practice (Latour: 1987), politics (Winner: 1986), or societies (Akrich: 1992). At the basis of this difference in focus seems to be a different conception of the relation between subject and object. Very generally, we could say that analytic philosophy of technology works with a conception of the subject as a free, autonomous, rational planning agent with certain desires and goals. Objects in the external world, including artefacts, may evoke particular beliefs and desires in the subject, but they only become relevant when the subject wishes to act on those desires or to achieve particular goals, leading to a focus on artefacts as instruments. In contrast, much continental philosophy of technology immediately dismisses the subject-object distinction. Rather than starting from a given subject, continental philosophy of technology often starts with relations that define, shape, or co-constitute both subject and object. Subjects and objects cannot and should not be considered in isolation: they are always situated, embodied or embedded in particular settings. Relations may emerge, change or disappear, but all of this may be brought about by any entity in the relation, whether human, artefact, or institution. Latour uses the term actants to encompass all these entities that can all ‘act’ in particular ways, rather than making a distinction between subjects/agents and objects/non-agents. Naturally, this leads to a focus on the second role of artefacts, that of influencing human actions.

By and large, both fields have developed their ideas on the relation between artefacts and human actions independently, and there has been little attempt to account for insights of the one within the framework of the other. This lack of cross-pollination is unfortunate, not only because there is no lack of different

46 I use the term ‘continental philosophy of technology’ here in a broad sense, encompassing certain parts of sociology (hence the reference to Latour and Akrich) as well as Science and Technology Studies and (post)phenomenology, as there is considerable overlap in questions posed and methods used by those fields.

47 A recent exception is Illies and Meijers (2009), who explain what the second role of artefacts amounts to in analytic terms, but not how artefacts play this role. Another exception is Schyfter (2009), who offers a social constructivist/collectivist account of some of the phenomena covered in Kroes and Meijers (2006). Strictly speaking, however, Schyfter’s work is rather incorporating insights from one analytic framework in another analytic framework.
and interesting insights in both fields, but also because engineers and designers could benefit from insights from both fields, yet cannot be expected to reconcile the different philosophical and methodological approaches by themselves.

In this chapter I will explain how artefacts can play their second role, how they can influence what we actually do, in terms that are acceptable to the analytic philosopher (or at least, the analytic philosopher who assumes the subject-object distinction). I will do so by constructing an account of how artefacts alter our reasons for action, that will hook up the investigation of the second role of artefacts to theories of practical rationality and (good) reasons for action. This account does not require the assumption that every action is necessarily preceded by rational deliberation. Rather, it just requires that every action can be understood as being done for a reason. To put the account to the test, I will show how it can deal with the phenomena covered by Akrich’s and Latour’s prescription and Verbeek’s invitation.

With respect to giving advice to engineers and designers, the reasons account also allows us to go beyond the actual influence of artefacts on agents and look at cases where artefacts fail to influence them, contrary to designer intentions or expectations. For example, suppose that people regularly crash on a certain speed bump. A reasons approach can then suggest possibilities why this speed bump didn’t influence drivers as it should have done: the relevant facts might not have been perceivable, the offered reason for slowing down might not have been considered a good or relevant one, etc. I will work this out in more detail in the next section.

To show how the influence of artefacts on agents can be explained in terms of reasons, in section 5.3 I explain what I take to be reasons for action and how they work, using Dancy’s (2000a) externalist account of practical reasons. In sections 5.4 and 5.5 I show how artefacts can play their first role, or what it means for them to enable (5.4) and facilitate (5.5) actions, and how this alters our reasons for action. In sections 5.6 and 5.7 I show how artefacts can play their second role, or how they influence what we actually do by prescribing (section 5.6) or inviting (section 5.7) them, in terms of providing or changing reasons for action. In section 5.8 I address two possible counterarguments to my claim that a reasons
account can explain the influence of artefacts on agents. Finally, in section 5.9 I will raise an argument against the systems perspective as an explanatory account of agency and actions, and I will suggest what it would take for us to ascribe at least a form of agency to artefacts, rather than to humans or human-artefact systems.

5.3. Practical reasons

For Dancy, reasons for action are facts (things that are the case) that favour or disfavour a certain course of action for an agent. For example, if I drive in my car and approach a speed bump, the fact that I am approaching a speed bump would be a reason for me to slow down. Besides being reasons, facts can play two more roles in Dancy’s account, which allows us to perform a more fine-grained analysis of the influence of artefacts on agents. The first role is that of the enabler/disabler. An enabler is something that enables a fact to be a reason, for example: the facts that I would ruin my car or that I would hurt my back when driving too fast over a speed bump enable the fact that I am approaching a speed bump to be a reason for me to slow down. Similarly, a disabler disables a reason to count as such. The fact that my car has excellent suspension systems can disable the fact that I am approaching a speed bump to be a reason for me to slow down.

The second role that facts can play is that of the intensifier/attenuator, which can make a reason a stronger or weaker reason for action. The fact that I am taking my frail grandmother for a ride does not in itself give me a reason for slowing down if there is no speed bump, but it may intensify my existing reason for slowing down (the fact that I approach the speed bump), because she will suffer as well if I drive too fast over it (Dancy: 2000b, 2004).48

As a note on terminology, facts can change: they can be brought about or ‘destroyed’, depending on circumstances. The fact that I am approaching a speed bump can be enablers/disablers and intensifiers/attenuators can be found in Dancy’s unpublished paper ‘Practical Reasoning and Inference’. http://experimentalphilosophy.typepad.com/2nd_annual_online_philoso/files/jonathan_dancy.pdf. Accessed 19 May 2010.
bump is brought about (made the case) by me taking a particular road in which a speed bump has been constructed. As soon as I take a side road, avoiding the speed bump, the fact that I approach it has been destroyed: it is no longer the case.

Dancy’s commitment to the claim that reasons favour or disfavour certain courses of action allows him to say that an agent can have a reason to act even if he does not desire to act, even if that agent would be fully informed and had deliberated rationally about what to do. Indeed, an agent can have a reason to act without knowing it, though this agent can still be said to act rationally, if he cannot be blamed for not knowing all the relevant reasons (Parfit: 1997). For example, the fact that someone is drowning nearby can enable the presence of a life buoy to be a reason for me to use that buoy to save the drowning person, even if the person drowning is my hated academic opponent whom I do not desire to save. It may even be a reason for me to act when I am slumbering on the beach, though then sadly I will not be aware of it. This is what makes Dancy and Parfit externalists about reasons, as opposed to reason-internalists like Williams (1981) who hold that an agent only has a reason for action if he, knowing all the relevant facts and after rational deliberation, is motivated to do that action (Parfit: 1997). In short, we can say that the internalist allows only desires to ground reasons for action, while the externalist allows facts to do so, both about desires and otherwise, and these reasons may not always be motivating. In the case of the drowning academic opponent, the reason to save this person does not come from my motivation: it has to come from what is actually good, or

49 Note that the distinction between internalism and externalism thus depends on motivation rather than knowledge: both internalism and externalism hold that an agent can have a reason to act of which that agent is currently unaware. What internalism does require is that the agent should be motivated to do that action if the agent should become aware of all the relevant facts and deliberates rationally about them.

50 An internalist account of how artefacts play their second role, how they influence what we actually do, would not be impossible to construct. However, internalism seems to be more suitable for explaining the first role of artefacts, that of instruments, due to its focus on agents and their desires and goals rather than on what we encounter in the external world (see e.g. Franssen: 2006).
the right thing to do in the situation. This gets practically relevant when we wish to advise engineers and designers, investigate why some technologies succeed and others fail, or evaluate design methodologies like value-sensitive design, where artefacts are designed for values that actual users may or may not hold. Consider for example the Ruton Robot, a universal household appliance developed in the Netherlands in the 1950s that could be used as a vacuum cleaner and as a mixer, among other uses (Lintsen: 2005, 264). While nothing was wrong with the technology itself, it never became a success. The reasons account offers us a number of places where we could start our investigation into why this technology failed. Were the right facts present? Were the relevant agents (the intended customers) aware of those facts? Were those facts enabled to be reasons for action (by their moral features, or by features of the agent or the context)? Were those reasons intensified to be good reasons for action, both instrumentally and morally? Were the good reasons actually motivating? In this specific case, it turned out that potential customers found the idea of cleaning the house and preparing food with the same appliance repulsive. In other words, they judged the value of flexibility to be less important than values of e.g. cleanliness and hygiene. In a reasons account, we could say that facts about the artefact were reasons for using it, and that the potential customers were aware of that. The potential customers, however, generally did not consider those reasons to be good reasons. In addition, the reasons against using the artefact were considered to be much stronger by the potential users than the designers had anticipated.

Note that even if an artefact provides a reason for action, that action need not be the (morally) best, or even a good, option: an obstacle in the road may be a reason for a car driver to take a detour over the bicycle path, but it would not necessarily justify that action. Furthermore, facts about artefacts can seem to be reasons for action, while they actually are not (e.g. when the actual facts do not favour that action). It may seem to me that I have a reason to press the button of a pedestrian traffic light if I want to cross the street, while in fact this may not be so, for example, if the button does not work.

Having gone into Dancy’s account of practical reasons, I will now examine how artefacts alter our reasons for action when they enable and facilitate actions in sections 5.4 and 5.5 and prescribe and invite actions in sections 5.6 and 5.7.
5.4. **Artefacts as enabling actions**

Artefacts can enable actions, that is, make it possible for us to do things that we would otherwise not be able to do: a knife enables us to cut, a plane enables us to fly. They thus let us achieve ends that we already had, but that were previously unattainable – humans fantasised about flying long before planes were invented, as well as ends that can be attained by artefact use, but which are discovered by users rather than designed for – using telephones for gossiping rather than for making the short business calls for which they were originally intended, or playing Pong on an oscilloscope.\(^5\) Likewise, artefacts can disable actions, that is, make it impossible for us to do things that we could have done without them: a wall disables our capacity to move in a certain direction, fetters disable our capacity to run. Artefacts can also enable actions that are disabled by other artefacts, as is the case with a key and a lock. To avoid confusion with reason-enablers and -disablers, I will henceforth speak of ‘artefacts making actions (im)possible’ rather than of ‘artefacts enabling/disabling actions.’ I described this before as the first role of artefacts, where artefacts change what we can do without necessarily changing our actual goals or behaviour.\(^5\)

Working with Dancy’s framework, we find that the fact that an artefact makes a certain action possible can never be a reason for action without taking into account further information. This is because, according to Dancy, there is only a reason to perform an action if some good can be obtained by that action, and we cannot tell without further information whether some good can be obtained by performing the action that the artefact makes possible. Note that you cannot have a reason to do what is impossible, though you can have a reason to try to do what might be possible (Streumer: 2007; see also Heuer: 2010; Streumer: 2010)

---

\(^5\) I go deeper into artefacts creating, replacing and eliminating ends and the moral significance thereof in section 6.6.

\(^5\) Whether the artefact actually makes an action possible for an agent depends on how fine-grained your description of the action is. A specific violin certainly makes it possible to ‘play that violin’, but it only makes ‘playing a violin’ possible if no other violin is present, and it only makes ‘enjoying oneself’ possible if no other ways to enjoy oneself are present. The important thing is that the artefact should be able to make an action possible if circumstances are such that there is no other way of performing the action under that description.
 Artefacts and reasons for action

– and you can have a reason to design and create an artefact that makes it possible to do that impossible thing.

Two kinds of facts are particularly important in enabling the fact that an artefact makes a certain action possible to be a reason for action. The first kind is the fact that performing the action with the artefact is itself valuable for the agent in some way (Dancy: 2006). This can for example be the case with playing the violin, bathing, or playing a computer game. The second kind is where the result or consequences of the action performed with the artefact are valuable for the agent in some way (Franssen: 2006). This is the classical case where an agent uses an artefact as an instrument to obtain some end: using a knife for cutting bread, using an electric shaver to make oneself presentable, and so on.

5.5. Artefacts as facilitating actions

Besides making actions possible, artefacts can also facilitate actions: they can make it easier for us to perform actions that we otherwise could have performed as well, either with natural or already available artificial means, but in more time, at a greater expenditure of energy, with the aid of more auxiliary items, at greater risk of failure or the production of unwanted side effects, etc. Running shoes facilitate my natural ability to run: ideally, they let me run faster while using less energy and placing less strain on my knees. A flyswatter facilitates my ability to swat flies because it is bigger and can be swung faster than my own hand (and with modern electric swatters, the fly can be killed with but a touch.) Other ways in which an artefact can facilitate an action is by increasing the chance of success of the action, or the variety of conditions under which an action can be performed. Compared to an antique arquebus, a modern rifle facilitates shooting, partly because the chance of it actually firing under normal

53 Note that this is a reason-externalist formulation. What is ‘valuable for’ an agent need not be ‘valued positively by’ that agent, or be attractive to the agent. For example, a foul-tasting medicine may improve the agent’s health (be valuable for the agent), but the agent might simply not care about such an improvement (value the medicine). A reason-internalist formulation would rather use ‘valued positively by’ here.
Acting with Artefacts

circumstances is better, partly because it is safer to fire for the user, and partly because the rifle will also work in rainy weather which would render the gun-powder in the arquebus, and thus the arquebus itself, unusable. Conversely, artefacts can impede actions as well: they can make them take more time, more energy, etc. A time-delay lock on a safe takes more time to open than a regular lock, wooden shoes impede your ability to run (but facilitate your ability to walk through mud), etc. Again, this is an example of the first role of artefacts.

Like with the fact that an artefact makes an action possible, the fact that an artefact facilitates an action cannot be a reason to use that artefact without taking into account further information: if there is no reason to perform a certain action with the means available, the fact that there are now better means available will not change that. The fact can be a reason-intensifier, though: you can have a stronger reason to use the artefact that better facilitates your actions. In other words: if you already have a reason to use an artefact, you may have a stronger reason to use an artefact that is better with respect to achieving your purposes. Of course, your reasons for not using the artefact may be stronger still. For example, though my car facilitates getting to the office, in the sense that driving is faster and less tiring than walking, I might have stronger reasons to walk as walking gives me exercise and time to think.

In this section I have investigated what artefacts do to our reasons for action when they enable or facilitate actions. In the next sections, I will go into how artefacts can prescribe (5.6) or invite (5.7) actions in terms of altering our reasons for action.

5.6. Artefacts as prescribing actions

The notion of artefacts prescribing actions originates in Akrich (1992) and was subsequently picked up by Latour (1992). Akrich argues that artefacts contain ‘scripts’ that ‘prescribe’ actions to users. These scripts do not get in the artefact by accident: they are ‘inscribed’ by engineers, based on the engineers’ view of who the users will be, in which context the artefact will be used, etc. Unfortunately, the effects of the artefact are context-dependent and the script that in one context prescribes moral behaviour may well prescribe immoral behaviour in
Artefacts and reasons for action

another. For example, a movable bollard in a road may prescribe cars to pass one at a time. If it breaks down, however, and does not sink back into the road, it suddenly prescribes car drivers to leave the road in order to pass and drive over the neighbouring footpath or bicycle path. In addition to the moving bollard, I will consider two well-known examples of Latour here: the speed bump in the road and the seat belt that, when not fastened, activates an alarm and a flashing light.

One characteristic is common to these examples: the artefacts mentioned ‘enforce’ actions to some degree. In Latour’s words: ‘I will call (...) the behavior imposed back onto the human by nonhuman delegates prescription’. (1992, 157, my italics). It seems that the prescription of actions comes in two degrees. The first is ‘soft’ prescription: this includes the prescription of actions by the speed bump and the seat belt. Here, there is no physical force, but not using the artefact (properly) will have certain negative consequences. Those who race over the speed bump damage their cars (and their backs), those who don’t fasten their seat belts have to put up with the alarm. The speed bump and the seat belt thus seem to provide their users with a reason against not following a specific course of action. The second form of prescription is ‘hard’ prescription: the movable bollard in the road prescribes stopping by making it impossible to drive on. Here, not only a reason against driving on is provided, the fact that doing so would wreck one’s car, but as driving on is in fact made impossible, the reasons for driving on are destroyed as well, as one cannot have reasons to do the impossible (Streumer: 2007).

---

54 I will make no distinction between ‘artefacts prescribing actions’ (e.g. driving) and ‘artefacts prescribing the agent to act in a certain way’ (e.g. driving slowly or carefully). This distinction seems to originate in us having a specific repertoire of action verbs rather than in artefacts influencing our actions in different ways.

55 This example is based on the case study in chapter 8.

56 In a footnote to this sentence, Latour weakens the meaning of ‘prescription’ by saying: ‘We call prescription whatever a scene presupposes from its transcribed actors and authors...’ (177n8, my italics). He then gives the example of a painting that is designed to be viewed from a specific angle of view. This meaning of prescription seems simply to be that artefacts have a specific use plan that users have to follow in order to successfully use the artefact (Houkes and Vermaas: 2004). I will come back to this form of prescription at the end of this section.
Before we examine more closely what is going on with artefacts prescribing actions, it should be noted that the artefacts in the above examples are not used because of their instrumental value: we do not drive over speed bumps because doing so serves a certain purpose. Rather, the artefact is part of a bigger system (the speed bump and the movable bollard are embedded in the road) or another artefact (the seat belt is part of the car), that is used because of its instrumental value. The speed bump, the movable bollard and the seat belt are artefacts we are forced to use on pain of us not being able to use the road or the car. If we choose not to use the road or the car, the speed bump, the movable bollard and the seat belt do not prescribe any actions to us.

It seems that we can understand the prescription of actions, at least in the examples given, as a specific form of altering reasons for action. Here, the presence of the artefact makes certain facts the case (‘if you do anything but this, X will happen’) that are enabled by other facts (‘X is bad for you’) to be reasons for action. I will spell out my previous suggestions of how this works for both soft prescription (the seat belt with the alarm) and hard prescription (the movable bollard). For soft prescription:

1. James’s car will get him quickly from A to B.
2. James has to get from A to B quickly (enables 1. to be a reason for James to take his car).
3. So James has a reason to take the car.

If James takes the car:

4. If James were to drive without fastening his seat belt, the alarm and a flashing light would be activated
5. James is annoyed by loud alarms and flashing lights (enables 4. to be a reason for James against not fastening his seat belt).
6. So James has a reason against not fastening his seat belt.

Of course, 4. is not the only reason for James against not fastening his seat belt. Not fastening his seat belt increases James’s risk of suffering harm in the case of an accident, and that is a reason for him against not using it even when there is no alarm. Nor does 4. ensure that James does indeed fasten his seat belt: maybe...
he dislikes wearing it so much that that fact gives him a stronger reason not to fasten it, and he decides to put up with the alarm – or abandons his plan of taking the car altogether.

For hard prescription, the first three steps will be the same:

(1) James’s car will get him quickly from A to B.
(2) James has to get from A to B quickly (enables 1. to be a reason for James to take his car).
(3) So James has a reason to take the car.

However, when James drives on and encounters the movable bollard:

(4) If James were to drive on, his car would crash.
(5) If James’s car were to crash, James would not be able to get from A to B quickly anymore (enables 4. to be a reason for James against driving on.)
(6) So James has a reason against driving on.

Again, this is in addition to other reasons James would have against driving on: that he would risk being injured, that he would have to spend time and money on buying a new car, etc. Moreover, with hard prescription something extra is going on simultaneously. With soft prescription, James can just grit his teeth, put up with the alarm and drive from A to B. With hard prescription, this is impossible: if James does drive on, the movable bollard will crash his car, and he will never get to B. In a reasons account, what happens here is that not only does James have a reason against driving on, but also facts that are reasons for doing alternative actions (like driving on) are destroyed. In the example, the following steps need to be added:

(7) If James wishes to get from A to B quickly, he has to drive on. (Is enabled by 2. to be a reason for James to drive on.)
However, when the movable bollard appears:

(8) Driving on is made impossible. (Destroys the fact that is a reason for driving on in 7. Note that this does not have to affect James’s initial reasons for wishing to get from A to B quickly.)

(9) So James has one less reason to drive on. (Furthermore, all other possible reasons for James to drive on will also be destroyed by 8.)

It seems that an artefact can prescribe an action if it can enforce (to some extent) certain actions on the agent. This is most likely to happen when the agent cannot avoid the artefact in the pursuit of certain purposes. This means that even instruments can prescribe actions: a corkscrew doesn’t prescribe opening wine bottles, but if I have to open a wine bottle, the corkscrew does prescribe how I should go about it. In other words: the corkscrew has a use plan I have to follow if I wish to open the bottle with it (Houkes and Vermaas: 2004).57

In terms of our reasons account, we can say that an artefact prescribes an action if it provides reasons against not doing that action (as with the seat belt); in addition, it may destroy or disable facts as reasons for doing any other relevant action (as with the movable bollard). An artefact may also prescribe an action if it intensifies a reason against not doing that action, or attenuates facts as reasons for doing any other relevant action, but these forms of prescription will usually be rather soft.

It might be tempting to say that the movable bollard prescribes stopping by providing you with a good reason to stop, rather than providing you with good reasons against driving on. However, I use the second formulation here to distinguish between cases where something positive can be gained (as is the case with invitation) versus cases where something negative can be avoided (as is the case with prescription). Moreover, it should be noted that this temptation arises because there is only one other relevant action here, driving on. If we for a moment assume that there are multiple relevant actions available (taking a side road, driving over the grass), it will be clear that the obstacle does not provide you with a good reason to stop if it does not simultaneously destroy your possi-

57 See chapter 8, especially 8.4, for a closer look at the use plan analysis.
Artefacts and reasons for action

...abilities or reasons for bypassing the obstacle, for you could simply take one of the alternatives. If there were multiple relevant actions available, the obstacle would provide you with a good reason against doing a particular action, namely, driving on. But this would not be prescription, as it would not 'enforce' particular behaviour. In this case, the movable bollard would merely block one possible action out of many.

Finally, in case of bad design it may seem that an artefact prescribes an action without the user being aware of it. For example, suppose that a particular movable bollard is hidden in the road so well that even attentive drivers do not notice it until it rises up under their cars. Here, there are good reasons for stopping, but they are not available to the drivers. This case, however, is not one of prescription: we can certainly not say that this artefact is influencing actions. It just crashes cars.

5.7. Artefacts as inviting actions

When an artefact prescribes an action, it does so with a certain force. When an artefact invites an action, there is no force, but it rather provides an opportunity for action and makes the agent aware that there is some reason to do that action.58 When a restaurant owner in a Greek tourist resort invites a passing tourist into his establishment, he is not forcing the invitee to come in, but rather suggesting that she could come in, and that there are good reasons for doing so: his wine is excellent, he plays authentic Greek music, etc. Of course, these facts might not be (good) reasons for the tourist: perhaps she detests wine and authentic Greek music. An invitation does not carry force, but neither is it always enticing.

The idea that artefacts can invite (and inhibit or discourage) actions is suggested by Verbeek (2000/2005), who extends the earlier work of Ihde (1990). Ihde is

58 Strictly speaking, it is perceiving the behaviour of an artefact that can make the user aware of certain facts that are reasons, e.g. perceiving a movable bollard rising out of the road, the display of an obstetric ultrasound machine showing a fetus in a womb, or a traffic sign simply being in place at a prominent spot.
mainly interested in how technologies affect perception. He claims that technologies mediate our perception of the world: they give shape to our perception, and thereby influence how we experience the world. This mediation can take several forms, for example embodiment, where an artefact used to perceive the world, like a pair of glasses, becomes ‘part of the agent’, or representation (the hermeneutic relation), where the agent perceives an artefact that represents the world in some way, like a thermometer.

Verbeek argues that artefacts do not only mediate and shape perception, but action as well. Like Ihde, he distinguishes several possible ways in which artefacts can do so. The most important way for our present purposes is translation, where the artefact changes our relation with the world by inviting certain actions and inhibiting others. Verbeek (2008) illustrates the translation relation with the example of obstetric ultrasound technology. On the one hand, this technology can be said to invite abortions, since it can make parents aware of inborn deficiencies or (risk factors for) hereditary diseases of the fetus. On the other hand, it can also inhibit abortions by confronting the parents with the fetus as a real, live human being, which can strengthen the emotional bond between the parents and the unborn child.

As I mentioned before, when an artefact invites an action it both provides an opportunity for action, and makes the agent aware that there is some reason to perform that action. Note that this awareness does not have to be conscious: the agent just needs to have access to those facts in some way. With simple artefacts, making an agent aware of an opportunity for action can be easy: humans can often quickly see what they can do with artefacts and other objects, though the actual perceived action opportunities may depend on the need of the observer (Gibson: 1979, chapter 8). With more specialized or complex artefacts the observer might need knowledge of a use plan to see them (Houkes and Vermaas: 59)

Apart from this, Verbeek’s notion of mediation is meant to do more work: he claims that ‘mediation consists in a mutual constitution of subject and object’ (2000/2005, 130). For artefacts, this means (among other things) that ‘artefacts coshape the use that is made of them (171). For elaboration on the mediation concept, see Waelbers (2009). For criticism, see Illies and Meijers (2009).
2004; see also chapter 4 for more on the perception of action opportunities, or affordances).

Inviting is not only about providing an opportunity for action, but also about providing a reason for doing that action and making the agent aware of that reason, or making the agent aware of an existing reason for action. This does not necessarily mean that any agent will consider the relevant fact to be a proper reason for action, let alone a good one. Artefacts are usually designed with typical (groups of) users in mind who are likely to respond to the invitation, that is, who would consider the provided reasons good reasons for action. The artefact would invite other agents as well, but they might just not consider the provided reason to be a reason. And even if they would, they might not consider it to be a good reason, or have other reasons not to respond to the invitation. For example: a comfortable chair can be said to invite sitting, even though the fact that it is comfortable might be a good reason for me to sit in it, insufficient reason for you to sit in it (perhaps you have more important reasons to hurry on) and no reason at all for a baby to sit in it, who might need firm support in order to sit upright at all. Whether a fact is a (good) reason for action depends on both observer and context.

Like prescription, invitation has a soft and a hard variant. In the soft variant, the balance of reasons is not altered by the artefact, but it only selectively makes the user aware of facts that count as reasons.\footnote{Though in a sense, making an agent aware of reasons is always altering the balance of reasons, as an agent can be said to have a reason (though not necessarily a good one) to act on information that is readily available: it is convenient.} For example: in cars, a prominent speedometer is designed to make you aware of a reason to drive faster or slower (that you are under or over the legal speed limit). A prominent air-fuel meter showing the efficiency of your engine is similarly designed to make you aware of a reason to drive faster or slower (that you can increase the efficiency of your engine). Here, facts about the speed of the car and the efficiency of the engine are not changed, but different facts are made available to the user that may constitute reasons for different actions.
Returning to Verbeek’s example of obstetric ultrasound technology, it seems that this technology invites agents to act by making them aware of facts that were already reasons for them. The fact that a certain fetus has an inborn deficiency might be a reason for abortion, the fact that it looks human already might be a reason against it, but such facts are not accessible without ultrasound technology. Ideally, such technology would make you aware of all the relevant facts for considering what to do. In practice, though, technology often leaves the user unaware of certain relevant facts, for example, because of the low resolution of the image. Or worse, it might distort the facts, for example, by presenting the fetus as larger than it actually is. Here, something is presented as a fact that might be a reason for action, while the ‘fact’ is actually not the case.

In the hard variant of invitation, the artefact does alter the balance of reasons: it creates or intensifies a reason to perform an action where there was no or insufficient reason to perform that action before. We can say that a bath looks inviting because it makes an activity possible which the agent may have a reason to perform (e.g. because bathing can be relaxing). In fact, it seems that every artefact that makes an action possible for which the agent can have a reason to do it also invites that action, assuming it makes the agent aware of the fact that the artefact makes the action possible.

An artefact may also replace any action which one has no (good) reason to perform with an action that one has a (good) reason to perform with the same result. The agent who just wishes to achieve the result can then perform the action with the artefact rather than the one without it. This is for example the case with the photographer who invites a child to look at the camera by showing a colourful little bird behind it.

A second form of strong invitation is where artefacts destroy, disable, or attenuate facts as reasons against doing those actions, where there are also reasons for performing those actions. For example: reckless driving may get you to your destination faster, but it increases the risks of you getting involved in an accident where you will suffer bodily harm. SUVs diminish the risk of suffering bodily harm.

---

61 Of course, neither of these facts might be a good reason for or against abortion, and dependent on the values of the agent, neither fact may be enabled to be a reason at all.
Artefacts and reasons for action

harm for their drivers when involved in an accident relative to other cars, due to their sheer size and weight. The fact that they do so destroys a fact that is a reason against reckless driving. In that way, an SUV may be said to invite reckless driving. The fact that the car is safer may thus not be an reason for reckless driving, but it certainly removes a reason against it, and the agent might have other reasons for reckless driving that then tip the scales. To give an example of how this works:

(1) Reckless driving increases James’s risk of suffering bodily harm
(2) Suffering bodily harm is bad for James (enables 1. to be a reason for James to not drive recklessly).
(3) So James has a reason to not drive recklessly.

But:

(4) In an SUV, reckless driving does not increase James’s risk of suffering bodily harm (destroys the fact in 1. that is a reason for James not to drive recklessly).
(5) So James has one less reason to not drive recklessly while driving an SUV.

Generally speaking, we can say that an artefact can invite an action in two ways. The first is when it makes the agent aware of existing reasons for action. These reasons can be false reasons if the ‘facts’ provided by the artefact are not the case, but would have been reasons if they had been the case: here, the invitation is also a deception. The second possibility for an artefact to invite an action is when the artefact provides the agent with reasons for that action or intensifies existing reasons, and makes the agent aware of that. Alternately, the artefact may invite an action by destroying, disabling or attenuating facts as reasons against

---

62 This might not be true for all kinds of accidents, e.g. rollover accidents. It could be argued that only the perception of safety in an SUV is greater than that in an ordinary car. In that case, the fact that is a reason not to drive recklessly would seem to be destroyed, while this is actually not the case. Hence the design recommendation to make cars seem less safe than they actually are (Horswill and Coster: 2002).
performing that action and making the agent aware of that, provided that there are also reasons for performing that action.

5.8. Two counterarguments

In this section I will consider two possible counterarguments against my account of artefacts influencing human action by providing reasons, and argue that they do not threaten my account.

The first counterargument is that the notions of prescribing and inviting were not meant to deal with individual actions, but rather with human behaviour, conceived of as pattern or way of organisation of individual actions. Latour seems to use the notion of prescribing for both cases, and Verbeek (2000/2005) regularly gives examples such as pens and word processors that evoke or promote a distinct way of writing (114-115), and PDAs that can invite working during train rides (197). Here, what is invited is not a single instance of artefact use, or a particular quality of an action. Rather, the way the agent uses the artefact is affected, perhaps even the agent’s working practices and lifestyle, if only because artefacts like PDAs are so versatile that they enable, facilitate, prescribe and invite a host of different actions.

However, it seems that an account of how artefacts alter reasons for action is able to deal with artefacts influencing behaviour, as long as the agent changes her behaviour for a reason (or several reasons). The case of the PDA inviting working during train rides can be drawn in parallel to the case of the SUV inviting reckless driving. The agent may have reasons for working during train rides (maybe she needs the time to get her work done), and she may have reasons against it (she would need to carry a heavy bag full of files and notes in the train home and back the next morning). The PDA could then destroy this last fact that is a reason against working during train rides, for the lightweight device could store all the information in the files and notes. In short, a reasons approach can account for both artefacts inviting single actions and artefacts inviting more complex or major changes in the agent’s overall behaviour, as long as the agent alters her behaviour for a reason.
The second counterargument is that artefacts may influence agents without providing them with reasons to act. Humans show all kinds of (unconscious) biases in their behaviour that may be irrational. One way in which artefacts can change our behaviour is by exploiting those biases. For example, TV screens in buses or trams may grab our attention without showing us anything worthwhile, and in the supermarket products sell better when placed at eye level than on the bottom shelf. All things considered, it does not seem we have any special reason to watch the screen or buy the product at eye level. Yet it seems we can say that the screen prescribes watching it, and the product at eye level invites us to buy it. These would then be cases where actions are prescribed and invited where our reasons for action are not altered.

I think it is undeniable that artefacts may change our behaviour by exploiting our psychological biases. This, however, need not be a problem for my account. We have already seen that artefacts can provide us with reasons to act while leaving the choice of actions to us, but this is compatible with them exerting inescapable causal force. The speed bump does not make it impossible for us to drive over it at high speed: it leaves the choice to us. Yet if we choose to speed on, we do not get to choose whether we want to suffer the negative effects or not. We do, and that fact is a reason for us to slow down. Here, what generates the reason is a physical effect that is causally necessitated, but such effects might also be psychological. The TV screen might prescribe looking at it by impeding other actions, in which case you would be able to look away by exerting your willpower, or it could practically unavoidably force you to look at it, in which case you would need to resort to a strategy to avoid looking at it, like turning your back to it. The reason against looking away is then simply the fact that it requires more effort to look away than to look at the screen. Again, this may not be a very good reason to look at the screen, but it is a reason nonetheless: people may still act on it and cite it to explain their behaviour.

Of course, it is possible that our psychological biases have evolved as mechanisms to help us respond to reasons, at least in the context in which they have

---

63 I will return to the relation between willpower and strategies, and the role technical artefacts can play in developing strategies to maintain commitment, in chapter 7.
evolved. A possible evolutionary explanation of those biases would then be the following. A certain class of facts might once have been enabled by other facts about their survival value to be reasons for action. For example, the fact that you see a sudden movement is a good reason to look that way, since there might be a friend, foe or predator approaching. Likewise, the fact that there is food to be found at eye level is a good reason to take it: the fact that there might be better food high up or low to the ground is no better reason to take that instead, due to the costs involved in adopting vulnerable positions like reaching out or crouching. Over time, our behavioural make-up adapted to this so that our responses to those reasons became ‘automatic’. This came at the cost of sometimes responding ‘for no good reason’, e.g. looking at a movement which turned out only to be the wind in the leaves, or an uninformative TV commercial. As long as the ‘benefits’ outweigh the ‘costs’, however, the bias can be expected to remain. So artefacts can present facts to us to which we respond as if they are good reasons for action, since we are biased to treat them that way.

I will not go here into what the relation could be between our capacity for practical reasoning, our behavioural biases and the way in which those biases are shaped, since that would take us too far afield. I would like to stress again, though, that practical reasoning does not have to be a conscious deliberation involving propositions. In Dancy’s account, reasons are facts that (dis)favour certain courses of actions. The agent should be aware of those facts in some way, be able to ‘weigh’ them against each other, and the agent’s actions should be understandable as being done for a reason, but this is not necessarily done through explicit or propositional reasoning. This ensures that the scope of facts about artefacts that could become reasons or enablers is very wide, wide enough for reasons to play a meaningful role in explicating what it is about artefacts that lets them prescribe or invite certain actions.

5.9. **Could artefacts have agency?**

I have considered two counterarguments against my reasons account of the role artefacts play in influencing human actions, and argued why they can both be accounted for. I will now close this chapter by wrapping up two loose ends. First, I will raise an argument against the systems perspective as an explanatory
account of agency and actions by broadening the scope of my reasons account (as well as that of the analysed concepts of prescription and mediation) from artefacts to objects in general. Second, I will examine what it would take to ascribe at least a form of agency to artefacts, as opposed to humans or human-artefact systems.

To get the systems perspective in clear view again, let me illustrate it with a quote by Verbeek: “...in technologically mediated situations, intentions come about in a complex interplay between humans and nonhumans [including artefacts], resulting in a form of agency that is distributed over humans and nonhumans.” (2009, 255).

Verbeek claims here that agency is distributed over humans and ‘nonhumans’, and thus, should be ascribed to human-nonhuman systems. However, this claim seems to be either trivial or nonsensical, at least with respect to the translation aspect of mediation, where an artefact can mediate our actions by inviting some and inhibiting others. To see why, let me assert that no matter whether we talked about making possible, facilitating, prescribing or inviting, those influences on agents cannot only be exerted by artefacts, but also by natural objects. For example, the weather can prescribe wearing a raincoat, a lake can invite swimming, etc. It seems that the nonhumans in Verbeek’s quote would have to include almost all aspects of our physical and social world, and that there is no good reason to distinguish between technologically mediated situations and mediated situations in general. Verbeek’s claim can then say two things. The first would be that we cannot understand agency and intentions without any reference to the world in which the agent lives, which would be trivial. The second would be that agency is distributed over everything (though maybe not in equal measure) – but it is not at all clear what this would mean, or what consequences this would have. Artefacts can prescribe and invite actions, but so can the natural world, and unless artefacts can mediate our perception and action in additional ways that the natural world cannot, mediation will not be able to get any specific explanatory concept of agency or artefacts off the ground. To be fair to Verbeek, though, the concept of mediation (as well as my reasons account) can certainly be useful to support normative theories of good
design. After all, both artefacts and natural objects can mediate actions or alter reasons for action, but artefacts are explicitly designed to do so. It is the responsibility of the designer to make sure that artefacts influence agents in an ethically acceptable way, and mediation (as well as my reasons account) can help here by investigating how this should be achieved and what the pitfalls are.

With respect to ascribing agency to artefacts, it has been argued that in order to be an agent, one has to have a certain degree of reasons-responsiveness (Fischer and Ravizza: 1998). Let us suppose we consider the ability to influence the behaviour of others by responding to their specific reasons for action a (necessary but not sufficient) feature of agency. Say, we consider it necessary to ascribe social agency, the capacity to engage in socially meaningful interaction with (other) agents. Suppose also that we do not want to formulate this ability in terms of the ability to invite or prescribe actions simpliciter, for we have seen that this is something both natural objects and ordinary artefacts can do. What could be a stronger form of influencing behaviour by responding to an agent’s reasons for action that, if found in an artefact, say, an artificially intelligent robot, could let us consider ascribing social agency to it?

My suggestion involves the notions of prescribing or inviting, but also goes beyond them. Consider the Greek restaurant owner again. There are three things that distinguish him from a simple sign advertising the quality of his establishment, or from a robot-waiter mindlessly beckoning people in and repeating welcoming phrases, that make the owner’s invitations (presumably) more successful. The first is his ability to select potential customers, for example, those tourists who are walking slowly, reading the menus, etc. In other words, he selects targets to address that look most susceptible to the reasons he can offer for eating at his restaurant. The second is his ability to not just present all reasons for eating at his restaurant, or a pre-made selection of them, but to present those facts he thinks most likely to count as reasons for his potential customers to come in. For example, a couple with young children will regard different facts as reasons for entering the restaurant, like the availability of high chairs, than a lone backpacker, who might just be looking for the cheapest meal. The third is his ability to create and destroy facts that are reasons, or to enable
facts to be (or disable facts as) reasons, for or against eating in his restaurant. If a potential customer is deterred by the loud music, he might turn it down; if the potential customer is looking for a local wine he does not currently have in stock, he might arrange to have some brought over. Important in all this is the restaurant owner’s ability to find out what the potential customers value and to respond appropriately to that, either by making them aware of relevant facts that might count as reasons, or by realising those facts that would count as reasons for them.

Again, these abilities do not seem to be sufficient to ascribe full-fledged agency to anyone or anything that exhibits them, but it might make it more likely that we would ascribe some form of social agency. Indeed, the criterion for ascribing thought to a machine, according to Turing (1950), could as well be adapted to ascribe some form of social agency. According to Turing, we should treat a machine as if it has thought if it can win an ‘imitation game’: if we cannot distinguish it from a human by asking it questions and assessing the answers. This criterion could be rephrased in terms of reasons (though for belief rather than for action): the machine should be able to give you reasons to believe it is human, and those reasons should be stronger than, or as strong as, those the real human is able to provide. This then, the machine has to do by recognizing what facts would count as reasons for the arbiter to believe the machine is really a human, and presenting those facts, or enabling facts present to be reasons for this belief.

Ordinary artefacts like corkscrews and SUVs do not have these abilities of finding out and speaking to reasons of their users, but having artefacts with these abilities is not science fiction. To give one example: online booksellers like Amazon suggest titles to their customers on statistical considerations, and so present a selection of facts (‘customers who bought this book also bought book X’) that they hope will be reasons for the customer to also buy these books. In a way, this is an attempt at influencing behaviour by selectively responding to potential reasons for action of the buyer. While this doesn’t mean that Amazon’s algorithm has social agency or would be able to pass the Turing test, it is certainly a step in that direction.
5.10. Conclusion

Artefacts do not only make actions possible and facilitate them, they also influence what we actually do. Analytic philosophers of technology usually focus on the former, while continental philosophers of technology tend to focus on the latter. In this chapter I have constructed an analytic account of how artefacts influence what we actually do, in terms of the way artefacts alter our reasons for action. This account describes how artefacts can influence agents and thus, offers a partial answer to the second part of the main question: How do artefacts influence agents and agency? Not only can a reasons approach offer an analytic explanation of how artefacts can influence agents, it can also help us to analyse what goes wrong when artefacts fail to influence agents, contrary to designer expectations or intentions.

This chapter has set out to show that it is at least possible to construct an account of how artefacts influence what we actually do in an analytic framework, how such an account could work and how it connects to existing research on practical reasoning within the wider tradition of analytic philosophy. Both analytic and continental philosophy of technology arguably have strengths and weaknesses, but unless they try to account for the same phenomena, no real comparison is possible. As such, I hope not only to have explained an interesting phenomenon already explored by continental philosophers of technology in analytic terms, but also to have contributed to facilitating this comparison.
6. How enhancement technologies affect us as bounded practical reasoners\textsuperscript{64}

6.1. Chapter abstract

In the last chapter I have argued that artefacts can influence agents by altering their reasons for action. There is also a more radical influence that artefacts can exert on us: they can influence agency, or the way in which we work as agents. Prototypical artefacts that do this are those that figure regularly in the human enhancement debate, like brain implants or certain types of medication. Thus, we will investigate this influence in the context of the enhancement debate, with the dual aim of mapping out the ways in which artefacts can influence agency, and discovering whether this influence is in any way unique for or restricted to human enhancement technologies.

The human enhancement debate typically centres on moral issues regarding changes in human nature, not on the means for these changes. In this chapter we argue that one cannot grasp what is morally salient about human enhancement without understanding how technologies affect human action and practical reasoning. We present a minimalist conception of human agents as bounded practical reasoners. Then, we categorise different effects of technologies on our possibilities for action and our evaluation of these possibilities. For each, we discuss whether enhancement technologies have morally salient effects; which technologies show these effects; and whether these differ significantly from

\textsuperscript{64} This chapter was co-authored by dr. Wybo Houkes. An earlier version of this chapter has been published as: Pols, A.J.K. and Houkes, W. (2011), ‘What is morally salient about enhancement technologies?’, in: \textit{Journal of Medical Ethics} \textbf{37} (2): 84-87. Pols wrote the first drafts of the complete chapter, and the largest parts of sections 6.2, 6.3, and 6.7. Houkes has extended the scope of the chapter from effects of enhancement technologies on our action scheme (Illies and Meijers: 2009) to their effects on us as bounded practical reasoners by rewriting sections 6.4 through 6.6.
Acting with Artefacts

those of other, non-enhancement technologies. We conclude that enhancement technologies are morally salient in several respects, that not all enhancement technologies share all those morally salient respects, and that continuities with traditional technologies may be found in all morally salient respects.

6.2. Introduction

Whatever else the parties in the human-enhancement debate disagree about, they share the conviction that there is something morally salient about human enhancement technologies (HET) in general. ‘Transhumanists’ argue that HET should be promoted; ‘bioconservatives’ maintain that they should be curtailed or at least approached with caution. This essay examines this shared presupposition by analysing the (technological) means used for human enhancement. We argue that neither party in the debate offers a clear analysis of what moral issues are exclusively or especially raised by HET as opposed to other technologies or to non-technological enhancement methods: both parties appeal to implicit and unsustainable views about human nature or about technology in general. This has, rightly, raised doubts that anything informative can be said about the moral acceptability of human enhancement.

We propose an alternative, more fine-grained approach. To identify what is morally salient about enhancement technologies, we start from an explicit, minimalist conception of human agents as bounded practical reasoners and of technology as instruments. This conception allows us to categorise different effects of technologies on our possibilities for action and our evaluation of these possibilities. For each effect that HET can have, we discuss whether it is morally salient and whether it can also be brought about by non-enhancement technologies. We conclude that HET are morally salient in several respects, that not all HET share all those morally salient respects, and that continuities with traditional technologies may be found in all morally salient respects.

6.3. The nature of the debate

Human enhancement technologies (HET) are often loosely specified as technologies that enhance human capacities past their normal level. Commonly used
How enhancement technologies affect us as bounded practical reasoners

examples are gene therapy, brain implants and smart drugs. The controversy surrounding HET rages on in the popular and academic press. Yet it turns out to be surprisingly difficult to determine what is specifically morally salient about these technologies.

Bioconservatives like Kass (2003) argue against enhancement technologies by claiming that there is something intrinsically valuable about human nature that will be lost once we enhance ourselves. By enhancing our natural capacities, we may become ‘better’, but not necessarily ‘better humans’. Leaving aside the problem with spelling out the ‘intrinsic value’ appealed to (e.g. Fukuyama: 2002), the strength of this argument cannot be judged independently of the underlying conception of human nature and its relation to technology. Three kinds of conceptions come to mind, none of which supports a special case against HET. Firstly, one could accept that designing and using technologies is part of human nature. Then, there is no ground for morally questioning any technology based on an appeal to human nature alone. Alternatively, one could maintain that human nature is technology-free. This leads to a general conservatism concerning technology. That attitude has precedents in philosophy (e.g. Borgmann: 1987) and is hardly limited to HET. However, it seems hard to sustain on a closer look at the history of technology. Many technologies have drastically altered human behaviour, yet are now part and parcel of our everyday lives (Levy: 2007; Lewens: 2009): interventions in schooling, exercise and nutrition are prime examples. If there originally were moral qualms concerning these interventions, they have vanished. A third possibility is to characterise human nature by our current capacities, technology-supported or otherwise. Then, the bioconservatives have helped themselves to their conclusion; and moral problems cannot be created by stipulation. Summing up, no view that appeals to human nature supports the conclusion that, morally speaking, there is anything specifically salient about HET.

It would be an overreaction to immediately set aside all intuitive moral qualms regarding HET. Yet arguments to the contrary, given by transhumanists, do little to clarify what is at stake in the debate. Take Bostrom’s defence of HET by the claim that there is ‘no deep moral difference between technological and other means of enhancing human lives.’ (Bostrom: 2005). This suggests that
opponents of HET must draw a morally relevant contrast between technological and non-technological ways of enhancement – which would imply a general distrust of technology and an underlying conception of human nature as technology-free. However, Bostrom’s claim implicitly assumes that there is no morally relevant difference between technological enhancements. This is counterintuitive. Domestic garlic presses enhance our capacity to crush garlic, improving the lives of cooks around the globe. Yet they leave people cold morally. By contrast, many people enhance their capacity to travel with SUVs or other fossil-fuel hogs, which others consider morally irresponsible because of the deleterious effects those cars have on the environment. By assuming that his opponents must make a technology-specific case against HET, Bostrom tries to saddle them with a burden of proof that they need not accept.

Summing up, it cannot be concluded that all technologies are morally unproblematic on the basis of the claim that technological enhancements of human lives are ubiquitous, or similar to non-technological enhancements. Intuitively, we can distinguish morally problematic and unproblematic technologies. What is at stake is whether this distinction can be spelled out so that HET come out as specifically problematic, even if the technologies turn out to be similar to non-technological enhancements in some respects. Appeals to human nature and the general role of technology in human life just muddy the waters.

6.4. Ends and instruments

A critical analysis of the current debate may lead one to doubt whether anything generic can be said about the moral impact of HET (Lewens: 2009). Given the state of the debate, this scepticism might be a result of the common appeal to human nature. We therefore go on to examine an alternative approach, starting from an explicit account of the human-technology relation, and deriving from this a variety of morally relevant effects of technologies. Only then do we consider whether (and, if so, which) HET have these effects and whether these effects are different from those of technologies in general. We start from the commonsensical (but disputable) idea that moral reasoning is a kind of practical reasoning. The latter is generally concerned with finding viable courses of action; moral reasoning is specifically about what one ought to do. This means
How enhancement technologies affect us as bounded practical reasoners

that a first, broad inventory of the morally relevant effects of technologies may be
given by studying their effects on practical reasoning.

We specifically consider practical reasoning as determining an end, i.e., a
desirable state of affairs; or determining the means to an end, i.e., the actions
that might bring about this state. To make practical reasoning a human affair,
instead of the concern of some idealised ratiocinator, we acknowledge that
human reasoning is limited by practical matters. Here, we consider three types
of limits. First, practical reasoning is situated: it arises from perceived or real
needs, evoked by practical situations – although one may deliberate about what
to do before, during or after the need arises. Second, practical reasoning is
embodied: it requires physical and mental skills of human agents. Third, practi-
cal reasoning has a limited capacity: humans always deliberate about a finite set
of ways to achieve their ends, often on the basis of incomplete and possibly
incorrect information.

To this, we add a lightweight, ‘minimal-instrumental’ conception of technol-
yogy, as a set of objects or processes that are instrumental to realising practical
purposes. Most philosophies of technology start from the assumption that
technologies are not neutral instruments, but shape human lives and societies
(see chapter 5). These perspectives share the idea that, whatever else they might
be, technologies are at least also instruments. In this sense, minimal instrument-
talism provides a minimal (but not a ‘basic’ or ‘core’) conception of technology.
It allows us to study the impact of new technologies, by reviewing their possible
effects on practical reasoning. These effects are not exhausted by examining
merely which means become available on technological change. Technologies
also have potential effects on (choice of) ends and the above-mentioned bounds
of rationality.

This approach to technology intersects the commonly used therapy-
enhancement distinction, that is, both therapeutical and enhancement technolo-
gies can have the same morally relevant effects – or have no morally relevant
effects at all. This is only natural, as therapeutical technologies like Ritalin can

65 Alternatively, technologies may be conceived of epistemically, as (systems of) knowledge for
achieving practical goals (Mitcham: 1978).
Acting with Artefacts

also be used for enhancement purposes and vice versa. What will differ is the moral permissibility of using technologies for therapeutical or enhancement purposes. While this discussion is best left to normative ethics, our approach helps by showing which effects should be evaluated.

We do not defend the evaluative utility or descriptive accuracy of this view of agency and artefacts here. It is merely used as a basis for distinguishing a number of ways in which technologies might affect practical reasoning and, more specifically, moral reasoning – without assuming from the outset that these effects are exclusive to technology or a particular type of technology. In the remainder of this chapter, we review these effects specifically for HET, and thus examine their moral salience.

6.5. Enabling and facilitating actions

An obvious effect of technology on practical reasoning is that it makes available material means for achieving ends. To disambiguate ‘means’, we reserve that term for actions and express this first role of technology as ‘materially supporting actions’. The material support provided by technology may take two different forms. In the first, technologies are ‘enabling’: they support the achievement of ends that were previously unattainable, either practically or in principle. A second, more quotidian form of technological support is ‘facilitating’: a technology makes the process of achieving a particular end more effective or efficient, in terms of cost, labour, etc.; or it makes possible the achievement of other ends with other technologies, e.g., by eliminating negative side effects.66 Both enabling and facilitating technologies can be used in a context of enhancing humans, as well as alleviating the impact of disabilities, for which the term ‘enabling technology’ has been used before (Hansson: 2007).

Enabling technologies appeal to the imagination but may be relatively rare. Arguably, examples are the first submarines and printing presses. Most technologies, from water mills to vacuum cleaners, are at least facilitating since they

66 See also sections 5.4 and 5.5 on how technological artefacts can enable and facilitate actions.
How enhancement technologies affect us as bounded practical reasoners

improve the capabilities of some human beings to realise practical goals. Moreover, enabling and facilitating technologies are morally salient. Weapons of mass destruction are morally reprehensible because they enable destruction on an unprecedented scale; and technology to capture and store CO2 may be morally problematic because it facilitates fossil-fuel technologies. Moral debates can and should be held about the adoption and diffusion of these technologies. There is no reason to expect specific moral salience of HET in this respect, nor to assume that the enabling or facilitating role of HET can be discussed in general. If moral problems can be found along these lines, they must be identified through the ends, behavioural patterns or other technologies that are enabled or facilitated. If there is a devil, it is in the details – of which goal is enabled or facilitated, and how. While all HET serve the general goal of human enhancement, they do so in many specific ways: smart drugs facilitate other ends than gene therapy.

6.6. Ending ends

Focusing on practical reasoning also reveals less obvious relations between enhancement technology and human actions. One is that HET might make reasoning about particular ends redundant, because they significantly change or even eliminate some practical needs. For example, Fluoxetine (Prozac) may remove (possibly perfectly reasonable) feelings of unsettledness or alienation and therefore the need to ‘find one’s place in the world’ (Elliott: 2000). This could make redundant reasoning about changing jobs, how to find new friends, et cetera.

Though this relation might be interesting, it is neither exclusive to HET nor necessarily morally salient. To start with the former: many technologies eliminate previously urgent needs; virtually all technologies do if these needs include

67 The distinction between enabling and facilitating technologies depends on the specification of the goal state: if this includes detailed ‘technical’ specifications, virtually every innovation is an enabling technology. A laundry ball, for instance, facilitates the broad goal of cleaning clothes; but it enables the specific goal of machine-washing clothes without the use of laundry detergent.
Acting with Artefacts

those subsidiary to using earlier technologies. Central heating, for instance, makes redundant the need to have one’s chimney cleaned; and – to give a trivial example – breeding new varieties of aubergines has eliminated the need to cover slices with salt in order to extract bitter juices. The moral salience of this general phenomenon is not obvious. For one thing, needs do not simply disappear but are typically replaced by other needs – not by moving up Maslow’s pyramid, but because of concerns like obtaining and repairing the new technology or its subsidiary items. Central heating creates a need for having one’s heating system checked regularly or perhaps for raising funds to have one’s house outfitted with a heating system. Even a decrease of the ‘net amount’ of ends might be taken as a morally neutral phenomenon: if one no longer needs to deliberate about achieving an end, nor to actually try to achieve it, one is free to pursue other goals. These might be reprehensible or commendable, but those moral aspects are independent of the eliminated need.

On closer inspection, however, replacing and eliminating ends might turn out to be morally non-neutral. Consider, for instance, the effects of new household appliances in the industrialised West in the first half of the 20th century. These, arguably, freed time and resources of household members, which could be spent on other tasks. Whether this led to more or less gender equality, e.g., on the labour market, remains the subject of debate (e.g. Cowan: 1983; Mokyr: 2000). What matters is that it might have, and that this is a morally relevant consequence. Expensive medical technologies and treatments, to give another example, tend to be accessible only locally, to the well-off in wealthy societies, thereby reducing global equality. The availability, accessibility and social impact of technologies make them morally relevant. The moral salience of HET has been discussed from this accessibility or cost perspective (Singer: 2009). This is certainly justified, insofar as deeper issues about equality are involved. However, as our examples make clear, these concerns are not exclusive to HET (see also Farah et al.: 2004).

6.7. Transcending our bounds

We have considered the effects of technologies on choosing means and on having needs. Some technologies that feature regularly in the enhancement
How enhancement technologies affect us as bounded practical reasoners

debate do not fit these categories. Examples are: deep brain stimulation (electrodes implanted in the brain) which can alter your personality (Leentjens et al.: 2004; Glannon: 2009), Transcranial Magnetic Stimulation (TMS) to improve motor learning (Bütefisch et al.: 2004), medication that reduces performance decrements due to sleep loss, like Modafinil (Gill, Haerich and Westcott: 2006), and the ‘concentration drug’ Methylphenidate (Ritalin) (Butcher: 2003).

Whereas most technologies are made for episodic use, some HET are made to become ‘part of the agent’. Often, they are not only physically closely coupled to the agent, but also difficult or impossible to remove from the agent once the end is achieved for which they are used. Brain implants and genetic enhancements provide forceful examples; so does medication that, once it has entered the bloodstream, has to run its course through the system before its effects diminish. Of course, some implants can be turned off, and TMS is used in episodes. Still, those technologies can profoundly affect the agent as long as they are in use. Apart from that, turning on a brain implant could alter the agent so that the agent would not want or be able to turn it off again (Leentjens et al.: 2004).

This permanency of use is one respect in which HET affect the ‘boundary’ aspects of practical reasoning, namely its embodiment. Episodic use can be preceded by deliberation; someone might decide against using the technology, either choosing an alternative or choosing not to pursue the end. Deliberating about technologies that are used permanently is more restricted. One might deliberate about having brain implants, but once they are in place not using

---

It should be noted that ‘permanent use’ and ‘becoming part of the agent’ are not necessarily restricted to technologies that are physically implanted or ingested. There seems to be no good reason not to include in this class of technologies those that are ‘permanently available’ to the agent, like watches and mobile phones for many inhabitants of Western countries, especially if we adopt Clark and Chalmers’s idea that the mind can extend beyond the boundary of skin and skull (Clark and Chalmers: 1998; Levy: 2007). As these technologies are used on a much larger scale than implant technologies, it seems certainly worthwhile to investigate their effects as human enhancement technologies.

Note that the permanent availability of technologies does not imply that one can always perform basic actions with them (see chapter 3), though this is certainly possible, e.g. in the case of a prosthesis that the agent is skilled in using.
them is no longer an option, supposing they operate permanently and cannot be removed.

Permanent use is, from a moral perspective, not the same as a chain of episodic use. Users bear responsibility for the episodic use of classical technologies, since they have a measure of control over this use: designers transfer responsibility and control by communicating how users may interact with the technology in order to realise some practical ends (Pols: 2010; see also chapter 8). Some HET, such as prostheses, may allow for the same control; others, such as gene therapy, do not. Thus, the responsibility for correct operation of the technology cannot be shared between designers and users in the same way as it is for classical technologies. This creates a morally salient difference between permanently used HET and other technologies. The problem of permanent use and loss of control is not unique to HET, though: it also applies to e.g. geoengineering technologies used for intentional climate change (Jamieson: 1996).

Two other possible effects of HET on the bounds of practical reasoning concern the set of actions considered by an agent in deliberation. Some enhancement technologies target capacities that determine how actions in this set are considered; this is one way of interpreting Bostrom and Sandberg’s (2009) statement that cognitive enhancements improve our ‘core cognitive capacities’. Ritalin, for example, improves our capacity to concentrate, thereby changing deliberation processes. Alternatively, technologies might help us to consider more practical options, e.g., by enhancing our creative or improvisational capacities or by allowing us to look beyond routine solutions. The two effects are independent: conceivably, one could deliberate in depth about a small set of options; or one might know many options and select one at random.

Many or all technologies affect the set of considered actions, namely by enabling or facilitating actions. However, enhancing mental capacities in the ways indicated above does not affect deliberation about ‘local’ practical problems; rather, these enhancements affect all practical reasoning. Therefore, their moral

---

69 Illies and Meijers (2009) have proposed the concept of ‘action scheme’, ‘the set of possible actions with different attractiveness that is available to an agent or group of agents in a given situation’ (427). Our analysis just features the set of actions considered by an agent, undifferentiated with regard to attractiveness.
implications may be far-reaching. Assuming that these deliberation-enhancing technologies have no negative side effects, such as emotional detachment or apraxia, they would also be largely beneficial. After all, enhancing our capacities to consider solutions to practical problems makes us better practical reasoners and thus, potentially, better moral reasoners: tendencies to overlook options or to choose familiar but morally questionable or unsustainable actions could be decreased; or we could concentrate on deliberating about the morally ideal option, without prestige or conformity bias (which provide fast but often not so frugal heuristics for solving practical problems). In other words, in these cases cognitive enhancement goes hand in hand with moral enhancement (Persson and Savulescu: 2008).

There is a clear continuity between these HET and other technologies, broadly conceived. A large variety of support systems, none of which involves brain implants or smart drugs, is available to support reasoning and decision-making processes, especially in groups. Closer to our disciplinary home, researchers have come up with rules for practical reasoning, decision theory and rational choice theory. These tools are not without problems, e.g., that few people seem prepared to use them in everyday life. Yet that they exist shows that, also in this respect, HET are not uniquely morally salient. That existing support systems are hard to implement should raise scepticism regarding speculations that our reasoning capacities – let alone our moral reasoning capacities – could be quickly and drastically improved.

6.8. Conclusion

The debate about the moral acceptability of human enhancements has suffered from implicit appeals to views of human nature and its relation to technology. These views are overly general, often reducing to general distrust of technologies or equally sweeping optimism. To avoid the sceptical conclusion that nothing informative can be said about the moral salience of HET, we have proposed another approach. This starts from an explicit conception of human agency, centering on bounded practical reasoning; and a view of technology as instrumentation for practical purposes. With this minimalist conception, we have identified three categories of effects of technologies that may be morally salient:
enabling/facilitating actions, eliminating ends, and effects on the bounds of reasoning. Particularly through shifting the bounds of reasoning artefacts can influence how we work as agents, and thus, can influence agency. This then, is the partial answer to complement that from chapter 5 to the second part of the main question: How do artefacts influence agents and agency?

With regard to the enhancement debate, we have discussed whether and, if so, which enhancement technologies could show these effects; what would be the specific moral salience of these effects; and whether these effects would be qualitatively different from those of non-enhancement technologies. In none of the three categories of effects considered here, enhancement technologies are discontinuous with more traditional technologies. However, we have identified some categories in which specific moral discussion is warranted nevertheless: the enabling of actions and the elimination of ends, for instance, cannot be profitably discussed in general, so that enhancement technologies having these effects need to be evaluated separately.

Although our proposal is one way to avoid overgeneralised moral attitudes towards HET, we do not claim that it is the only one. Alternatives to our minimalist conception are not hard to come by, and neither are more detailed views of the relation between moral and practical reasoning. Our main claim is that the debate about enhancement should be refined by adding these conceptions and differentiating between morally relevant effects of technologies on their basis. This enhancement of the debate can be achieved even with minimal means, as we have argued here.
7 Robust joint commitment and the use of strategies\textsuperscript{70}

7.1. Chapter abstract

Many authors writing in the field of joint action agree that joint action or collective intentionality needs some kind of joint commitment to support it. All agree that, contrary to individual commitments, it should not be possible to unilaterally rescind a joint commitment. Interestingly, those authors who do develop a notion of joint commitment tend to stress its use for giving participants grounds for demanding compliance with the commitment, or demanding explanations or compensations in case an individual walks out on a joint action committed to.

In this chapter I propose a more ambitious notion of joint commitment, robust joint commitment. This notion assumes that the purpose of making a joint commitment is to minimise the risk of individuals walking out on a joint action: blaming them if they do is one way to do this, but there might be other, and depending on circumstances, better methods available. In particular, I argue that commitments can be strengthened by implementing various strategies, primarily intentional, social and technical strategies. I examine the strengths and weaknesses of each kind of strategy before showing how they can be incorporated in a notion of robust joint commitment. Implementing and using artefacts can be part of a technical strategy due to the capacity of artefacts to make actions (im)possible and alter our reasons for action (see chapter 5). Thus, artefacts can not only influence individual but also collective agents and agency.\textsuperscript{71}

Adopting a notion of robust joint commitment has several advantages, including enabling commitments between parties that may not be connected

\textsuperscript{70} An earlier draft of this chapter has been presented at the Conference on Collective Intentionality VII.

\textsuperscript{71} The notion of collective agency is admittedly more controversial than that of collective or joint action. My claims in this chapter hinge only on the existence of the latter, though I briefly mention issues related to collective agency in section 7.3 and 7.5.
emotionally or fully trust each other, as is often the case in large groups or
organisations. I will end by considering how a robust joint commitment could
be made to stop and counter human-induced climate change, and I evaluate differ-
ent strategies that have been proposed to strengthen such a commitment.

7.2. Introduction

Suppose that in an early draft of the Odyssey, Odysseus had decided to pass the
sirens in a different way. Rather than having himself tied to the mast and
plugging his men’s ears with wax, he and his shipmates had communicated
among themselves that they intended not to jump overboard at the siren’s call.
As a crew, they would thus have made a joint commitment to stay on the ship
and sail past the sirens. Had that been the case, Odysseus’s voyage would likely
have ended abruptly.

Joint action is a relatively young topic inspired by the observation that peo-
ple can act together in ways that do not seem to be adequately analysable in
terms of individual intentions and actions. Among the concepts involved in joint
action, commitment seems to be a central one. Without at least a commitment
by all participants to perform a certain action or achieve a certain goal together, it seems there can be no joint action (Bratman: 1992; Castelfranchi: 1995; Roth:

Unfortunately, the notion of joint commitment has received relatively little
specific attention in literature on joint action, and those who investigate it
primarily see joint commitment as stemming from individual commitments,
conceived of as acts of the will (Gilbert: 2006). While joint commitments can in
principle be based on individual commitments without being reducible to them,

72 While I use the label ‘joint action’, different labels are in use to denote the field, depending on
the particular interests of the author and what the author takes to be joint or shared, like
‘collective intentionality’ (Searle: 1990), ‘shared cooperative activity’ (Bratman: 1992), ‘shared
agency’ (Roth: 2004) or ‘plural agency’ (Helm: 2008).

73 Achieving a certain goal together can include performing a certain action together, executing a
plan, refraining from interfering, et cetera.

74 Pettit and Schweikard (2006) think that joint action precedes joint commitment, but they do
not offer a detailed account of how this would work.
and conceiving of individual commitments as acts of the will does not seem to be problematic either, basing joint commitments on individual acts of the will alone may yield us a definition that is somewhat restricted in its application. In particular, this notion of joint commitment would be vulnerable to individuals cheating on or walking out on the joint action committed to. Current definitions of joint commitment state that joint commitments cannot be unilaterally rescinded. They also explain when and why we can blame those who walk out on a joint action they are committed to and demand compliance from them, which is definitely useful. However, it seems that the reason why joint commitments are made is not to have grounds to blame those who walk out on the joint actions committed to, but rather to have people work together as a group and keep working together, in order to do things that require coordinated and sustained effort like sailing safely past the sirens. Creating grounds to blame those who walk out on a joint action they are committed to is one way to do so, but there are other, and depending on circumstances, possibly better options available. For example, current definitions cannot properly handle cases where enough individuals walk out on a joint action to render blaming ineffective, whether those individuals do so intentionally or unintentionally. Even worse, cases like the multiperson prisoner’s dilemma can further erode joint commitment by introducing a temptation to cheat that grows stronger as more participants cooperate, where individual commitments would not be affected as individuals cannot cheat on themselves.

In this chapter I propose a more robust notion of joint commitment based on that of Gilbert,\textsuperscript{75} that takes minimising the risk of individuals walking out on a joint action to be the main purpose of a joint commitment, rather than creating grounds for demanding explanations or repairs from those individuals if they do. I understand minimising risk here as either making it less likely that an individual will walk out on a joint action, or reducing the costs of an individual actually doing so. Walking out on a joint action can be done intentionally (cheating) or unintentionally (by losing agency, suffering from akrasia, etc.). In particular, I will argue that to get from Gilbert’s notion of joint commitment to a notion of

\textsuperscript{75} Whenever I refer to Gilbert, I refer to Gilbert (2006) unless stated otherwise.
robust joint commitment an extra joint commitment has to be made: the com-
mitment to implement an appropriate strategy together to minimise the risk of
individuals walking out on the joint action previously committed to.

Current notions of joint commitment suffice to address many practical prob-
lems and deal with many well-known examples of joint action: taking a walk
together, singing a duet, having a smoke and a chat after work. As such, the
notion of robust joint commitment is not intended to replace it, but rather to
extend the scope of joint commitments to cases where regular joint commit-
ments would become powerless, like the case of Odysseus and his men.

I will work towards a notion of robust joint commitment in the following
way: In section 7.3 I argue why basing joint commitments on individual acts of
the will alone is for practical reasons more problematic than basing individual
commitments on individual acts of the will alone. In section 7.4 I address the
need for implementing strategies in order to maintain commitment. I explore
three different kinds of strategies: intentional strategies that only require mental
activity on the part of the agent to be implemented, social strategies that pre-
dominantly depend on involving other agents, and technical strategies that
predominantly depend on involving technical artefacts or other physical means.
In section 7.5 I propose how a notion of robust joint commitment can be estab-
lished that focuses on minimising the risk of individuals walking out on the
joint action, rather than on how we can create grounds for demanding explana-
tions or repairs from those individuals. In section 7.6, I end with a practical
example by considering how a robust joint commitment could be made to stop
and counter human-induced climate change, and I evaluate different strategies
that have been proposed to strengthen such a commitment.

7.3. Joint commitment

Of the authors writing on joint action, only a few explicitly analyse the concept of
joint or mutual commitment (Castelfranchi: 1995; Roth: 2004; Gilbert: 2006).
This may be because, for individual actions, intentions and commitment usually
go hand in hand (Bratman: 1987; Velleman: 1997; but see Gilbert: 2006 for a
different view). It could then be assumed that an analysis of collective intention-
ality would make a separate analysis of joint commitment unnecessary. This
would be assuming too much, however: several proposals have been made to get from individual to collective intentions, but all seem to require an extra notion of joint commitment. I will briefly consider three such proposals here, those of Bratman (1992), Velleman (1997) and Gilbert (2006).

Bratman (1992) has pointed out that for individuals to have collective intentions, those intentions have to much more resemble plans than individual intentions do, if collective intentions are to go beyond what one individual can attempt to do, control or exercise discretion over. There might not be much differences between intending and planning as such, but one of the key differences between ‘having an intention’ and ‘having a plan’ seems to be that one can have a plan without commitment, but not an intention. If a decision theorist draws up a plan for maximising profits in a game of roulette this night, that does not commit her to execute it, while if she forms the intention to do so this night, it seems to commit her to do at least something: act on the intention this night, or abandon it (Cohen and Levesque: 1990). As collective intentions resemble plans, they do require additional commitment in order to make them binding.

Velleman (1997) argues that a collective intention can be created by forming individual conditional intentions that are conditional upon each other. For example: if I intend to take a walk only if you do, and you intend to take a walk only if I do, and we communicate this to each other, then we can be said to together intend to take a walk. Such a construction, however, could fall prey to the same problem conditional commitments would have: if one party walks out on the other, the other has no ground to demand apologies or repairs, since such collective intentions and commitments could be unilaterally rescinded (Roth: 2004).

The last alternative I consider here to come to a collective intention is to form a collective or group agent that can have intentions (Gilbert: 2006; Helm: 2008). While such notions have been considered too implausible to even start investigating by some (Searle: 1990; Kutz: 2000), much depends on the criteria a

---

76 For clarification, this holds for both direct and indirect control. While we may coordinate our efforts when doing something together, I do not control you like I (directly) control my body parts or (indirectly) control my model plane. See also chapter 3.
collective agent would have to meet in order to be called as such, and whether its intentions would have to be unique mental states that are not in the head of any individual (Pettit and Schweikard: 2006). The more plausible notions of collective agency depend on the work of individuals to bring such an agent into being and maintain it. But without any commitment to do this, collective agents and their intentions could become quite unstable and vulnerable to individuals simply walking out on them.

Not only is joint commitment needed for collective intentions and joint action, this commitment has to be more than a collection of individual commitments towards a particular goal, just like a collective intention is more than a sum of individual intentions (e.g. Searle: 1990). Castelfranchi (1995) gives the example of two university professors committed to finding a cure against AIDS, yet who work independently and consider each other rivals rather than allies. Clearly, an extra or stronger form of commitment is needed than individual commitment to properly support collective intentions and joint action. Several proposals have been put forward: here, I will examine and extend that of Gilbert (2006).

Gilbert’s solution to get from individual commitments to joint commitments is to extend the notion of individual commitments. According to her, an individual commitment is a decision, an act of the will. Individual commitments give one a reason to act on them, though it need not be an overriding reason: if there are better reasons to abandon the commitment, the decision can be made to do so. In this respect, Gilbert argues, commitments differ from intentions that can just fade out or disappear from existence (see also Gilbert: 2005).

A joint commitment can be created if each group member makes an individual commitment to emulate together with the others a single body or person that intends to perform a certain action. If these commitments are made and mutually expressed, and this is common knowledge among the participants, the

77 Other notable accounts are those of Castelfranchi (1995) and Roth (2004). Both propose an extra form of commitment to make the transition from individual to joint commitments, namely, commitments to someone (possibly yourself) to perform a certain action. While I work with Gilbert’s account, my extension could be adapted to those other accounts of joint commitment as well.
Robust joint commitment and the use of strategies

joint commitment is created. This can be as easy as two persons agreeing to meet at a certain time. It may even happen implicitly: Gilbert gives the example of two factory workers who acquire the habit of meeting each other after work for a smoke and a chat, thus forming a joint commitment without ever formally arranging their meetings. Similarly, commitments can be jointly rescinded like they are created, or simply fade away over time. A joint commitment cannot be unilaterally rescinded, however: if one party walks out on the joint action it is committed to, the disadvantaged individuals have a right to demand an explanation or compensation.

While Castelfranchi, Roth and Gilbert have all proposed different accounts of joint commitment, they all stress the fact that joint commitments cannot be unilaterally rescinded, and thus that there are reasonable grounds to criticise the individual who walks out on the joint action to which that individual is committed. This, of course, is important: an individual should not be able to unilaterally rescind a group’s commitment, leaving the group with empty hands and no grounds to complain. However, it seems that an account of joint commitment could be more ambitious. Compare the case of commitment to that of morality. If Plato’s Gyges or Hume’s sensible knave ponders the possibility of immoral deeds, the challenge is not to point out to them that those deeds are immoral or that they will be responsible for what they do: they know already. Rather, the (much harder) challenge is to tell them why they should be moral in such a way that no rational defiance is possible. Likewise, in the case of joint commitment, the harder challenge is not to find grounds to blame individuals for walking out on a joint action, but to minimise the risk of individuals actually walking out on it. Creating grounds for blaming those who walk out on a joint action and demanding compliance is one way to do this – but there are more and, depending on the situation, possibly better methods available.

The harder challenge supposedly holds for both individual and joint commitments. For example, akrasia or loss of agency in an individual can threaten the stability of both individual and joint commitments, and in both cases, strategies are available to help strengthen the commitment. Two reasons, however, make it particularly pressing to take on the harder challenge when dealing with joint
commitments as opposed to individual commitments. The first reason has to do with epistemic uncertainty: with individual commitments, it is never in doubt whether all those involved know of any desires or plans to rescind commitment and agree to do so, as there is only one person involved. It is not possible to decide to walk out on an individual commitment, since deciding to walk out on an individual commitment is rescinding it. With joint commitments, this is not the case, and this uncertainty about individual attitudes may well hinder cooperation. In a worst case scenario, I could mistakenly interpret your compliant behaviour as cheating or defecting, motivating me to actually defect, motivating you to actually defect. Thus, a joint commitment can end through a misunderstanding even if no one intends to walk out on the joint action first. In a more robust joint commitment, this would be less likely to happen: if I know that the risk of you walking out on our joint action is minimised in some way, I would be more surprised at your apparent cheating and more likely to check whether my interpretation of your behaviour is indeed correct. Also, I would be less likely to retaliate for your perceived defection by defecting myself, because the strategies implemented to minimise the risk of any of us defecting would work against me as well.

The second reason for taking on the harder challenge is the potential incentive for cheating on a commitment. Cheating on your own individual commitment is impossible, and there are many cases of joint commitments where there is little incentive for cheating, as in the cases of two people going for a walk, singing a duet, or meeting after work for a smoke and a chat. However, there are also plenty of cases where there are strong incentives for individuals to cheat, even knowing that they would be held answerable for it. In the worst cases, like in prisoner’s dilemma cases or the tragedy of the commons (Hardin: 1968), individuals would be tempted to cheat even though keeping to the joint commitment would overall be better for them, even if they can communicate freely with the other participants. In those cases, the rewards of cheating often

---

78 Different kinds of strategies for maintaining individual commitments or, more specifically, avoiding procrastination, can be found in Heath and Anderson (2010). Heath and Anderson include social strategies as ways for individuals to keep procrastination in check.
rise as more people cooperate and invest in the cooperation. In such situations people might refrain from entering joint commitments at all since the risk of failure is too high, even while acknowledging that it would be best if all jointly committed to a certain course of action.

One practical consequence of addressing the harder challenge would be that it could help tie together two separate strands in the field of joint action. Some authors argue that joint action requires a great degree of trust (Roth: 2004), emotional connection (Helm: 2008) or a sense of togetherness (Seemann: 2009). This implies that joint action will usually be restricted to small groups of agents who know each other well. Other authors have suggested that the field of joint action is particularly suited to investigate joint actions by large groups or within organisations (Castelfranchi: 1995; Kutz: 2000; Tuomela: 2007). This has led to minimalist accounts of joint action that place less requirements on the connection between participants. Both kinds of accounts seem incompatible with each other as the individuals within an organisation rarely forge a strong emotional connection or sense of togetherness. Thus, it seems that we can either require the strong connection but can then only explain cooperation in small groups of friends, or aim to investigate joint actions within organisations but are thereby forced to adopt a minimalist account.

One reason for this incompatibility could be the perceived risk of individuals within large groups or organisations of investing too much (emotionally or otherwise) in co-workers whom one hardly knows, and who might turn out to be unreliable. Strengthening a joint commitment by minimising the risk of individuals walking out on a joint action could then be a way to facilitate cooperation by limiting the possibilities of failure, or the costs of misplaced trust, thereby removing a barrier for individuals in organisations to trust or depend on each other. This would in turn allow those who require strong connections between individuals to extend their accounts to joint actions of larger groups of people than they can currently handle.

To extend Gilbert’s account of joint commitment by including strategies, in the next section, section 7.4, I go into when and why committing yourself by an act of the will alone cannot always be successful. I then show what kind of strategies could be adopted to strengthen a joint commitment. In section 7.5 I
argue how those strategies can be included in a definition of robust joint commitment, and what kind of strategies would be useful for what kind of situations.

7.4. Self-control and strategies

An individual commitment may be a decision that binds the agent, but unless it has a certain degree of persistence, it is of no use whatsoever. Constantly changing commitments would not be irrational, but it would at least be a sign of indecisiveness, just like constantly changing plans (Raz: 2005). I assume that, as commitments are considered to be acts of the will, this persistence is initially maintained by willpower or self-control, that allows the agent to work towards fulfilling the commitment and maintaining it in the face of temptations or frustrations.79

Unfortunately but not unexpectedly, psychologists who have investigated self-control argue that it is hardly perfect. According to one well-tested model, self-control resembles a muscle: it can be used repeatedly, and training strengthens it, but like a muscle, self-control tires after repeated attempts and exercising it becomes progressively more difficult. This happens independently of the task for which self-control is exercised, whether mood regulation, vigilance, or coping with stress, for example (Muraven and Baumeister: 2000; Baumeister, Vohs and Tice: 2007). The analogy can be drawn further: it seems that our self-control mechanism, like a muscle, relies on glucose to function. Apparently, low levels of blood glucose impair self-control (Gailliot et al.: 2007). This means that keeping a commitment by exercising self-control alone can become problematic in the face of multiple temptations or frustrations, or enduring temptations or frustrations that require the agent to remain vigilant and keep pressing on. Exercising self-control becomes even more difficult if the agent's attention is directed towards the rewards involved in abandoning commitment, at least for young children (Mischel and Ayduk: 2002).

79 While not identical, willpower and self-control are very similar, especially with respect to resisting temptations and overcoming frustrations. In this chapter I will only refer to self-control for the sake of simplicity.
Fortunately, keeping a commitment can be made easier by using strategies. I consider a strategy to be a measure that can be taken to enhance the probability that an agent will achieve a certain goal, like keeping a commitment. This entails that making an individual commitment itself is not a strategy: the decision to adopt a goal does not necessarily enhance the probability that the agent will achieve it, and it does not seem to be irrational for an agent to have inconsistent goals (Velleman: 1997). Making a joint commitment relies on a strategy, though: communicating your individual commitment to emulate together an intending person to the other participants. This works as a promise and thus gives you a reason to keep to your individual commitment (and gives the others grounds for demanding compliance). Exercising self-control can in this light be seen as implementing a basic commitment-keeping strategy. While some strategies can be useful for both maintaining commitment and achieving the goal committed to, the two are conceptually distinct: not all strategies for maintaining commitment can be used to achieve the goal committed to, and vice versa.

Both groups and individuals can use strategies, though some strategies, like intentional strategies, are restricted to individuals, while others, like social strategies, are restricted to groups. Strategies can be combined in order to achieve better results, and indeed, often are – see e.g. Pettit (1996) on institutional sanctions, or Brennan and Pettit (1993, section 3.2) on different strategies employed in keeping jurors in a jury system committed to their task. In the remainder of this section, I will sketch three classes of strategies, intentional, social and technical strategies, and assess their value for supporting joint commitments. Interestingly, psychologists have particularly been interested in intentional strategies (e.g. Mischel and Ayduk: 2002), while philosophers, especially ethicists, have mostly examined social strategies (e.g. Searle: 1969, 1996), and sociologists of science and technology have typically studied technical strategies (e.g. Latour: 1992). This chapter, therefore, is also an attempt to direct the attention of philosophers towards intentional and technical strategies and the role they could play in supporting joint actions.

I consider an intentional strategy to be any strategy that merely requires mental activity on the part of an individual to implement it. Examples would be: count-
Acting with Artefacts

ing to ten to prevent yourself from getting angry, imagining yourself in a peaceful place in order to fall asleep and trying not to think of white bears – a sure strategy to make yourself think of white bears (Wegner et al.: 1987). To work out an example in more detail, Mischel and Ayduk (2002) argue that directing attention away from certain characteristics of the situation can be useful to overcome temptations and frustrations, and thus help to maintain commitment. This has successfully been tested in delayed gratification paradigms, of which the ‘marshmallow test’ is particularly well-known. In this test, a researcher gives children a marshmallow and two choices before leaving: either they wait for an unknown period until the researcher returns and gives them two marshmallows, or they ring a bell which will bring the researcher back immediately, but then they only get the one marshmallow. Delaying gratification becomes much easier for children when they focus on abstract qualities of the situation, like the resemblance of the marshmallow to a white cloud, or think pleasant thoughts unrelated to the situation. In contrast, delaying gratification becomes harder for them when they focus on tempting aspects of the immediate reward, like the sweetness and taste of the marshmallow that is right in front of them. Similar tests have been performed on adults as well (e.g. Ayduk, Mischel and Downey: 2002) with similar results.

While useful, intentional strategies have their limitations. Strategies in general always need to be adopted by making a decision, the decision to adopt that strategy, but intentional strategies are the only strategies that can immediately be abandoned by making the decision to do so as well. Of course, a good strategy ensures its own success, in this case by drawing the agent’s attention away from the reasons for deciding differently and giving in to temptation. Most problematic for us, however, is that intentional strategies are not easily extendable to joint commitment. Groups can advise participants to adopt certain intentional strategies, or encourage them to do so, but it is always the individual who has the last say over the application of an intentional strategy. This might be enough if all participating individuals are well-intending but weak-willed or prone to temptation, but not if individuals actively seek out opportunities for cheating. In such cases, there might not even have been a proper joint commitment to start with – but this would be little consolation for the deceived parties.
I consider a **social strategy** to be any strategy that predominantly depends on involving other agents to reach a goal. Those other agents should be involved as agents, not merely as instruments, otherwise the strategy is technical rather than social. Social strategies include communicating commitments by signing contracts and making promises. They may also go beyond this, as when agents participate in institutional activities, assist others with doing their parts or motivate themselves to do something by promising others that they will do that thing (Velleman: 1997). As many authors writing on joint action have argued (e.g. Searle: 1969; Castelfranchi: 1995; Gilbert: 2006), making a promise or commitment to someone to achieve a certain goal entails all kinds of things: the other is entitled to be informed when the goal is reached or has to be abandoned, is entitled to require that the goal be reached, can demand repairs in case of failure, et cetera. As such, those strategies can certainly be effective in the sense that they do generate reasons for action, and have understandably received much attention in investigations of joint action. Without invoking any social strategy at all, joint action seems impossible. According to Gilbert, when you communicate your individual commitment in order to form a joint commitment, you do not just motivate yourself to comply: you fulfil a necessary condition for joint commitment. Again, this can be done explicitly, e.g. by making a promise, but also implicitly, as in the case of the factory workers.

The strength of social strategies, moral obligations and written contracts is that they cannot be unilaterally rescinded like individual commitments and strategies. Moreover, walking out on a joint action that is ‘protected’ by social strategies can not only lead to a loss of reputation, but also have physical consequences: if I walk out on my contractual obligations, I may get a demotion, which entails loss of income. If I fail to repay a loan in time, this can be registered at a credit registration office, resulting in a lower credit rating for me, making it more difficult for me to loan money in the future. However, while social strategies do motivate and protect commitments to some degree (Goodin: 1996), and combining social strategies can help to do so, the joint actions they protect can all be walked out on by an individual suffering from loss of agency or akrasia, or simply deciding that the rewards for walking out on a joint action or breaking a promise outweigh the costs. In other words, it seems that implement-
Acting with Artefacts

ing a social strategy is necessary, but not always sufficient, for jointly making and keeping a commitment.

I consider a technical strategy to be a strategy that predominantly depends on technical artefacts or other physical means or instruments to enable an agent or group of agents to reach a goal. Examples are: Odysseus having himself tied to the mast of his ship with a sturdy rope, his men plugging their ears with wax, setting an alarm clock (for yourself or another) to wake up at a certain time, et cetera. The suggestion of physical force inherent in technical strategies may remind one of the ‘Mafia sense’ of joint action (Bratman: 1992), in which I ensure that we drive to the docks together by clubbing you unconscious and locking you in my car trunk. Here, however, there is clearly no joint action: the difference is that the implementation of technical strategies has to be committed to by the group in order to count as such.

Whether technical strategies can be abandoned differs per strategy. Some can be abandoned by the action of any individual: the alarm clock can be activated and deactivated by anyone who has access to it. Some can only be abandoned by other group members or a decision of the group itself, like tying Odysseus to the mast with a sturdy rope. Some technical strategies cannot be abandoned at all until the goal is reached. A prime example of this is the Russian Doomsday Device from the Kubrick film Doctor Strangelove that will destroy the world if any of a number of Russian military bases is destroyed, and that cannot be turned off, in order not to weaken its effect as a deterrent. Real-life examples would be pulling the pin from a grenade to prime it or administering a vaccine to provide disease immunity.

As is clear, the advantage of technical strategies is that they can not only minimise the risk of individuals walking out on a joint action, they can make it almost impossible for them to do so, or prevent their walking out from having any effect on goal achievement, in ways that other strategies cannot. Where social strategies can generate reasons for action, technical strategies can likewise do so, but they can also physically constrain what any individual or group can
Robust joint commitment and the use of strategies

actually do: the rope can constrain Odysseus where his own decisions or promises would have given way to the siren’s luring song.

On the other hand, technical strategies have two important disadvantages. The first is exactly their advantage: they can prevent individuals from walking out on a joint action, or ensure the achievement of a goal, even if all group members agree that the goal should be abandoned and the commitment to the joint action rescinded. In Doctor Strangelove, a renegade general who does not know about the Doomsday Device shuts down all communications and orders his men to attack one of the Russian bases. Though both Americans and Russians work together to prevent the activation of the Doomsday Device and the destruction of the world, they fail to do so. A more realistic example of an irreversible technical strategy is that of conquerors who destroy the ships of their invasion fleet upon arrival. They thereby make retreat impossible and thus assure their soldiers’ commitment to the cause.

The second disadvantage of technical strategies is that they cannot enable joint commitment in any way. While they can support and strengthen existing joint commitments, they cannot be used to make them. Unleashing a technical strategy upon an unsuspecting individual is not creating joint commitment: it is enforcing compliance without commitment, or acting together in the Mafia sense. Browbeating someone into making a joint commitment seems possible, and this may depend on the initiator being able to implement technical strategies to make good on the threat: threats that cannot be followed up on are empty, just like promises. However, such a joint commitment would be unstable at best, and walking out on the enforced joint action could well be justifiable by external standards, irrespective of the moral value of the goal committed to.

80 The artefacts used in technical strategies can not only alter our reasons for action, but also enable, disable, facilitate or hinder actions: see chapter 5. Theoretically, enhancement technologies could even increase our capacity for practical reasoning or maintaining self-control, thereby making it easier for us to maintain a commitment by using certain intentional strategies: see chapter 6.

81 This strategy is most famously ascribed to the Spanish conquistador Hernán Cortés, who conquered the Aztec Empire in the 16th century, but also to the Muslim Umayyad general Tariq ibn Ziyad, who conquered Visigothic Hispania in the 8th century.
A final note on strategies: the distinction between intentional, social and technical strategies is not clear-cut, nor is it meant to be. Rather, the distinction is made to draw attention to the various aspects of making and keeping to joint commitments that are used in practice, and to examine their relationships. In the next section I will examine these relationships and propose a notion of robust joint commitment that takes on the harder challenge proposed in section 7.3: explicating how we can minimise the risk of individuals walking out on a joint action.

7.5. Making a robust joint commitment

Philosophers working on joint action have so far primarily been concerned with social strategies to ensure joint commitment. This focus is understandable because social strategies are necessary for commitment, as I have argued in the last section, and also sufficient, provided one does not wish to address the harder challenge. Indeed, in many examples from the field of joint action, like two persons going for a walk, singing a duet or meeting for a smoke and a chat at the factory gates, there seems to be little need for minimising the risk of cheating or defecting. A notion of robust joint commitment can deal with these cases too, if it leaves open which strategies are appropriate for which activity, but it is also applicable in circumstances where existing notions might not adequately describe the situation, like in the case of Odysseus and his crew or in large organisations where cheating or individual failure might be more likely.

In short, we need to do something to our joint commitment to minimise the risk of individuals walking out on the joint action committed to; to prevent individual cheating and failure if possible, and to prevent individual cheating or failure from having significant costs if it occurs. Of course, a group should remain free to abandon their commitment as a group. A group may want to prevent itself from doing so, like the Russians did when constructing their Doomsday Device, but this goes beyond our definition: it rather seems that the Russians in the end gave up their joint commitment to activating the Doomsday Device, but were unable to thereby prevent its activation. Likewise, a group may fail to achieve a goal without losing commitment to that goal. This is why robust
Robust joint commitment and the use of strategies

joint commitment applies to situations where the risk of individuals walking out is minimised, rather than where goal achievement is ensured.

Let us revisit Gilbert’s conditions for joint commitment. First, according to Gilbert, joint commitment to a goal means that each involved party is committed to emulate together with the others, as far as is possible, a single person who intends to achieve that goal. For this to be successful, the following conditions must be fulfilled:

1. Each involved party must make the individual commitment to emulate together with the others, as far as is possible, a single person who intends to achieve a specific goal.
2. These commitments must be mutually expressed.
3. These commitments must be common knowledge between the parties.

To make this joint commitment into a robust joint commitment, a second joint commitment needs to be made, namely a joint commitment where the specific goal is to implement appropriate strategies to minimise the risk of individuals walking out on the joint action previously committed to. By making the second joint commitment, the first joint commitment is turned into a robust joint commitment. Note that this does not make the original joint commitment redundant: it is possible to commit to implementing a strategy appropriate for minimising the risk of individuals walking out on a joint action, without committing to the joint action itself. For example, I may sign up for an exercise program that is intended as a strategy to keep individuals to their commitment to exercise together regularly, not because I want to exercise together regularly, but because I think it is a good way to meet new people. Of course, I could in time experience the benefits of exercising together regularly, and then adopt that as a goal.

For a group, implementing appropriate strategies has benefits over and above minimising the risk of group members walking out on the joint action they have committed themselves to. I will mention two here. First, the group has to decide

---

Gilbert speaks of ‘performing an action’ rather than ‘achieving a goal’, but I prefer ‘achieving a goal’ as it is broader: it can also include executing a plan, refraining from interfering, et cetera.
how important the goal is and what measures would be appropriate to ensure commitment, and this could give individuals in the group a stronger sense of group membership and working together, thus strengthening the original commitment. Second, by implementing strategies like explicating the ‘rules of the game’, establishing proper decision procedures and deciding when an individual can be relieved from a commitment, the group works towards fulfilling the conditions for becoming a group agent or collective individual, allowing it to really act as one (e.g. Pettit and Schweikard: 2006).

The part of the second joint commitment that requires elaboration is, of course, what an appropriate strategy is. This depends on circumstances, but based on our overview in section 7.4 we can make some suggestions. First, if it is reasonable to assume that there will be no significant temptations or frustrations for individuals, and the costs of individual cheating and failure are low, simple social strategies like making a promise, even an implicit one, will suffice to both create and maintain a joint commitment. This is the case with many small-scale examples from the literature, like two persons going for a walk, singing a duet, or meeting outside the factory gates. As temptations and frustrations for individuals increase, or the costs of individuals cheating or failing rise, stronger social strategies might come into play, like signing contracts or swearing oaths. This would give individuals additional reasons to adhere to the joint commitment. Unfortunately, this might work best in cases where individuals know each other well, and risk of cheating is already low. Walking out on a promise may hurt one’s reputation or standing with other group members, but if those group members are all strangers and will likely not be encountered again, this would not be a high price to pay. Indeed, in prisoner’s dilemma cases people tend to adopt more cooperative strategies if they know that they will have to play the game again with the same players, in other words, if they know that reputation is important (Goodin: 1996, 14).

If the participants are well-disposed but likely to be weak-willed or sensitive to temptation, additional intentional strategies might be implemented. For example: when a therapist and a patient work together to relieve the patient’s arachnophobia, the therapist could suggest thinking of the ultimate aim of the therapy and how good relief from the phobia would be, every time the patient
would be tempted to skip a session due to fretting about the large, hairy spiders used in the later stages of the desensitisation therapy.

Finally, the more temptations and frustrations arise, the larger and more diffuse the group grows, the higher the odds of individual cheating and failure become, and the stronger the need for technical strategies becomes. As noted, technical strategies can not only provide reasons for action, but also constrain what people can actually do. Technical strategies are themselves not foolproof: one has to commit to implementing any strategy, including technical strategies, and this implementation could again be walked out on. However, as long as the costs of this are relatively low, it will usually not be necessary to pose an extra commitment to a strategy to commit to the strategy: the point of requiring commitment to a strategy is to weed out undecided individuals or cheaters at an early stage, before participants start investing resources into jointly trying to achieve their goal in earnest, when the costs of cheating can rise much higher.

Depending on the stakes and the situation, it may be the case that only other group members can release the individual from the joint commitment, like Odysseus could only be untied by one of his men. It may also be the case that only the group (the single person emulated) can release an individual from the joint commitment, e.g. by a voting procedure or by all members signing a document. This amounts to the first possibility if the group consists of only two persons. Individual failure or cheating could also be allowed, but at a price. Suppose that we intend to go to the theatre together, we can commit ourselves to doing so by buying the tickets in advance. If I then fail to do my part in our plan and fail to show up, I will have lost my money.

7.6. Robust joint commitment and human-induced climate change

To end this chapter with an example of how a robust joint commitment could be made to address a practical problem, I will give an overview of several kinds of

\[ \text{83 It should be noted that making a group decision can be tricky: Pettit and Schweikard (2006) show that any group of three or more people can fall prey to the discursive dilemma, that is, cannot guarantee that they can generate a common, consistent body of judgements by a regular voting procedure.} \]
strategies suggested to maintain commitment to stop and counter human-induced climate change. I will assume that this commitment is shared by many individuals, as is illustrated by international events such as the 2009 Copenhagen summit on climate change that drew delegates from all over the world, yet that it is by itself too weak to get individuals to work and keep working together to achieve their goal, as is illustrated by the relatively weak agreement that was the result of the Copenhagen summit. The overview will show how the division between intentional, social and technical strategies can be used to classify suggested strategies and search for new ones.

Many environmental problems associated with human-induced climate change have the structure of a tragedy of the commons (Hardin: 1968) or a multiperson prisoner's dilemma (Gardiner: 2001), which by their structure discourage cooperation and encourage cheating. Gardiner has argued that the intergenerational problem of human-induced climate change is even worse than a tragedy of the commons: we can negotiate with our peers about use of the commons, but we can by definition not negotiate with future generations, they cannot offer us anything that we cannot just take ourselves and, by the time they have replaced us, they have just as much incentive to pollute as we do. Yet several strategies have been proposed to deal with environmental problems with such structures, and I will discuss them here.

**Intentional strategies** were not considered useful by Hardin (1968) when he first wrote about the tragedy of the commons. He argued that humans were simply too much focused on self-interest for a widespread change in mentality to be successful. Even worse, it would work best with conscientious people, but not with the cheaters, leading to great advantages for the cheaters over the cooperators. However, Fox (1985) argues that while other kinds of solutions might treat the symptoms, only a change in mentality can treat the disease, the root of the problem. Moreover, if we manage to change our mentality, other kinds of solutions will work that much better. This has led authors like Nussbaum (1994)

---

84 The Copenhagen Accord was not passed unanimously, nor is it legally binding for those who have accepted it. See e.g. the BBC News of 19 December 2009 at http://news.bbc.co.uk/2/hi/8422133.stm. Accessed 20 January 2011.
Robust joint commitment and the use of strategies

and Dower (2003) to advocate a notion of world or global citizenship, describing the kind of mindset that we can and should adopt to keep to our commitments and deal with the ‘tragic’ structure of many environmental problems. The primary ways to achieve this change in mindset would seem to be education and information, and moral, religious and ethical appeals to prosocial behaviour (Gardner and Stern: 2002). This could work even if humans are fundamentally self-interested, as Hardin argues. For example, cooperation might not always serve one’s self-interest in a narrow sense, but it might do so in a broader sense: it might be better in the long term, for one’s children, community, country, etc. (Gardiner: 2001).

Social strategies were Hardin’s solution of choice, particularly privatisation of the commons and strict laws, or mutually agreed upon coercion. Privatisation of the commons is not always possible, however: pollution in one part of the air or the sea will spread out and affect the whole system, rather than just the part assigned to the polluter (Nussbaum: 1994; Gardiner: 2001). Strict laws might work on a local or national level, but are problematic to implement on an international level because there is no global government. As a consequence, current international treatises tend not to be strict laws. For example, while one cannot legally withdraw from national laws, the 2009 Copenhagen Accord is not legally binding and even the stricter 1997 Kyoto Protocol allows (with some conditions) withdrawal from the Protocol without sanction. 85

Another social strategy that can minimise the risk of individuals walking out on a joint action has apparently been overlooked by Hardin, but has been suggested by both Gardiner (2001) and Gardner and Stern (2002): relying on small-scale social mechanisms like a sense of obligation towards the community, a sense of honour, reciprocity, fair play and the punishment of cheaters. The disadvantage of this strategy is that it works best at the level of small communities where individuals know each other well and interact with each other repeatedly. Supposedly, a global citizenship mentality would enlarge our sense of

community and thus, make this strategy more effective on a larger scale, but this mentality would have to be implemented first.

Gardiner (2001) has argued that all of these social strategies are in trouble when confronting the intergenerational, rather than the intragenerational problem. Privatisation and strict laws can easily be undone by future generations, and they cannot reciprocate or detect and punish cheating on our part: there is no incentive for us to cooperate and deal with the problem now, nor will there be for any future generation once it becomes the present generation. Though I think Gardiner is right to warn us for this problem and its tragic structure, his account seems to be too pessimistic for two reasons. First, while we can by definition not negotiate with future generations, current individuals and organisations can (and indeed do) represent them, just like people can choose to represent endangered species and nature reserves, and can thus apply strategies like blaming, praising or commenting on behalf of their charges on the behaviour of the current generation. Second, all but the narrowest forms of self-interest will include interest in the well-being of one’s children, if only as a biological disposition. This means that the step from self-interest to interest in the well-being of at least some particular members of a future generation should be relatively easy to make. This, then, could serve as a basis for broadening self-interest to include interest in the well-being of future generations, and as a way to empower small-scale social strategies, where our behaviour can be praised, blamed or commented on by (representatives of) our children.86

Hardin explicitly claims that there are no technical strategies to deal with the tragedy of the commons due to the problem’s structure. Gardiner does not talk about them, though Gardner and Stern mention them in their list of solutions to the tragedy of the commons under the broader heading of ‘changing incentives’. It does not help that technical solutions to the problem of intentional climate change, like geoengineering, have generally been regarded sceptically (e.g.

86 This happens implicitly in the well-known proverb: “We have not inherited the Earth from our fathers. We are borrowing it from our children.” The proverb expresses an intentional strategy in that it advocates a change of mindset. This is strengthened by appealing to two separate social norms, neglect of which generally elicits social sanctions: that you ought to take care of your children, and that you ought to return something borrowed in good condition.
Robust joint commitment and the use of strategies

Jamieson: 1996). Yet there is at least one clear way in which technical strategies could help maintain a commitment to stop and counter human-induced climate change.

If we look at the intergenerational problem again, there is another reason why Gardiner’s account seems too pessimistic. Gardiner argues that part of the problem is that each new generation will have the same incentives to pollute, rather than to preserve the earth for future generations. This is not entirely true: our supply of fossil fuels is finite, and the more we consume, the less is available for future generations to consume. In addition, the costs of extracting fossil fuels will likely rise as those that are easy to obtain are used first, which means that future generations will likely have less incentives to obtain and use a lot of fossil fuels. This, in turn, will make it easier for future generations to successfully commit to stopping and countering human-induced climate change, assuming that their alternatives to burning fossil fuels are not more polluting. In this sense, consuming fossil fuels now is a technical strategy that will help future generations to maintain their commitment, though both an unintentional and an unethical one.

The more general observation we can draw here is that the material world we find ourselves in determines our incentives and thereby, our behaviour (e.g. Latour: 1992; Verbeek: 2000/2005. See also chapter 5). As we currently live in a hydrocarbon economy that mostly uses energy from fossil fuels, it is easier for us to continue living this way than, say, switching to a hydrogen economy would be. This aspect of the material world can be turned to our advantage, however: if we were to switch to a hydrogen economy, future generations would find it easier to maintain it than to switch back to the old hydrocarbon economy. Similarly, any sustainable technology we develop and implement can serve future generations as well as us. In this respect, the development and implementation of new technologies is similar to strict laws, in that implementation of both requires a sacrifice from us, after which both change our incentives and make it in our narrow self-interest to behave differently. However, technical strategies are more robust than strict laws: a law can be repealed by a new government or collectively ignored, while a new technology or technical infrastructure, once in place, may require considerable time and energy to dislodge.
Acting with Artefacts

Finally, let me note that we need not and should not ask which kind of strategies would be the ‘best’ for making a robust joint commitment here. Rather, given the scope and the complexity of the problem of human-induced climate change, it seems that a combination of all kinds of strategies would be needed in order to empower us to make a robust joint commitment to stop and counter it (see Gardner and Stern: 2002).

7.7. Conclusion

Current notions of joint commitment aim at defining a joint commitment that cannot be unilaterally rescinded. Rather, if participants walk out on a joint action they are committed to, the others would be justified in demanding an explanation or compensation from the responsible individual.

While useful and sufficient for many situations, this does not seem to be the primary aim of joint commitment, which is rather getting individuals to actually work and keep working together, or minimising the risk of individuals walking out on the joint action. Enabling the group to demand explanations or compensations from individuals who do so is one way to do this, but there are other ways that might work better, depending on circumstances.

In this chapter I have examined intentional, social and technical strategies, and argued how they can be (and indeed are) used to minimise this risk. I have incorporated these strategies in an extended notion of joint commitment, robust joint commitment. Incorporating strategies to come to a notion of robust joint commitment has other advantages, such as enabling commitments between parties that may not be connected emotionally or fully trust each other, as is often the case in large groups or organisations. As technical strategies might include the use and deployment of technical artefacts, this chapter has suggested an additional way in which artefacts may influence individual agents: by helping them to keep their commitments to certain actions, especially in the face of temptations and frustrations. Here, it is not so much the bounds of reasoning that are affected as well as the bounds of our capacity to act on what we regard (or regarded at some point) as our best reasons for action. Additionally, insofar as collective agency is a meaningful notion, artefacts can influence collective
agency by keeping individual participants committed to the collective action and thus, increase coherence and stability of the collective agent.

Interestingly, if we return to the comparison between commitment and morality, it seems that (especially technical) strategies might help us there as well to minimise the risk of immoral behaviour (Verbeek: 2000/2005, 2006). Indeed, as Latour (1992, 157) states: “...no human is as relentlessly moral as a machine...” Spring mechanisms ensure that we close doors behind us, speed bumps ensure that we do not drive too fast and alcohol locks in cars ensure that we do not drive after drinking too much. I have argued that especially social and technical strategies offer a powerful way of ensuring robust joint commitment; like philosophers working on joint action, it seems that ethicists can ignore these strategies as ways of ensuring moral or at least discouraging immoral behaviour only at their own disadvantage.
8 Transferring responsibility for artefacts through use plans

8.1. Chapter abstract

In chapters 2 to 4 I have addressed the first part of the main question: What is an action with an artefact? In chapters 5 to 7 I have addressed the second part of the main question: How do artefacts influence agents and agency? In this final chapter I will look at responsibility of engineers and users for actions with artefacts.

 Particularly I will look at the transfer of responsibility for artefact use from engineers to users. Engineers are initially responsible for the artefacts they produce, but at some point, part of the responsibility for the artefact shifts from the engineer to the user. This chapter will analyse how and when this transfer of responsibility takes place. Specifically, it will combine the theory of responsibility and control of Fischer and Ravizza (1998) and the use plan theory of rational design and use of artefacts by Houkes and Vermaas (2004, 2010) into a new theoretical framework which specifies the conditions under which this transfer can take place. After introducing both theories and combining them, I will give an example of how the combined theory works and apply it to a test case to show how it functions in practice.

---

87 Earlier drafts of this paper have been presented at the KIVI NIRIA / 3.TU Ethics PhD lecture series and the 2007 Workshop on Philosophy and Engineering. An earlier version of this chapter was published as: Pols, A.J.K. (2010), ‘Transferring responsibility through use plans’, in: Van de Poel, I., and Goldberg, D.E. (eds.), Philosophy and Engineering: an emerging agenda, Springer, Dordrecht: 189-203. The copyright to this Contribution is transferred to Springer Science + Business Media B.V., Dordrecht. The copyright transfer covers the sole right to print, publish, distribute and sell throughout the world said Contribution and parts thereof, including all revisions or versions and future editions thereof and in any medium, such as in its electronic form (offline, online), as well as to translate, print, publish, distribute and sell the Contribution in any foreign languages and throughout the world. The original publication is available at www.springerlink.com. DOI: 10.1007/978-90-481-2804-4_16.
Acting with Artefacts

8.2. Introduction

Engineers are responsible for the technical artefacts they produce. The claim seems straightforward enough, yet how far does this responsibility go? Both on a theoretical and a practical level, issues are entangled.

On a theoretical level, responsibility can be approached in two different ways. The merit-based approach ascribes responsibility to agents on the basis of their actions, focusing on what it means for an agent to be responsible: whoever performs a certain action merits, or deserves, a certain reaction. The consequentialist approach ascribes responsibility to agents so that it will lead to the desired effects, focusing on when an agent should be held responsible, namely: if he or she is in the best position to make those desired effects happen, or avoid undesired effects (Eshleman: 2004). Both views can affect engineers: on the one hand, they are causally responsible for the technical artefacts they produce. This means they can be held morally responsible and be praised or blamed for those artefacts and the effects they produce. On the other hand, they are in a good position to improve aspects of both single technical artefacts and extensive technical systems, so it makes sense to ascribe certain responsibilities to them in advance. Sweden, for example, has introduced a policy for road transport systems where system designers are designated ‘ultimately responsible’ for traffic safety. This does not mean that responsibility for traffic safety is taken away from individual road users, but rather that the system designers are encouraged to take measures ‘so that the mistakes and errors of some individuals, regardless of who is considered to be responsible, do not have fatal consequences and that such mistakes and errors will not be committed with the same frequency.’ (Fahlquist: 2006, 1118).

On a more practical level, just as meaning and application of the term ‘responsibility’ has shifted over time (Mitcham: 1987), ideas about the responsibilities of engineers have changed as well. A century ago, for example, ethical codes for engineers did not mention responsibility for the welfare of the public. While this has lead Mitcham and Von Schomberg (2000) to claim that this responsibility was considered less important than that of loyalty to the firm and customer, Davis (2001) makes a good case against this interpretation, though only later responsibility for the welfare of the public was explicitly made
of paramount importance. The American National Society of Professional Engineers’ Code of Ethics, for example, states that the engineer should ‘hold paramount the safety, health and welfare of the public’ (NSPE: 2007), while the British Royal Academy of Engineering has issued a statement of ethical principles which mentions that engineers ‘work to enhance the welfare, health and safety of all whilst paying due regard to the environment and the sustainability of resources.’ (Royal Academy of Engineering: 2007). While laudable, these statements have their drawbacks. It is not always clear how these requirements translate to engineering practice and Broome (1989) has argued that certain risks in engineering are unavoidable.

This chapter will give a merit-based account of responsibility, focusing on who is responsible for a technical artefact, the engineer or the user, and under what circumstances this responsibility is transferred. While the responsibilities of the engineer have increased over time, responsibilities for what the artefact can or cannot do and for the consequences of using the artefact remain ‘core responsibilities’: these are the responsibilities that will be analysed here.

This does mean that because of its specific scope, my framework allows for an engineer to make a torture device and successfully transfer responsibility for that device to a user. Isn’t this letting the ‘evil’ engineer off too easily? It is important here to keep in mind that not all responsibilities of an engineer are transferrable. Next to the responsibility for specific artefacts, engineers also take on more general, non-transferrable responsibilities when they enter the profession like those established in ethical codes. These allow us to say that any engineer who would knowingly and willingly build torture devices would always behave irresponsibly, as this is a violation of the engineers’ responsibility for the safety, health and welfare of the public. Likewise, being responsible for the safety of a product might entail recalling it when tests bring unnoticed defects to light, and responsibility for the environment might translate into offering opportunities for recycling used products. For these aspects, the engineer remains responsible for the artefact during its complete lifecycle.\(^{88}\)

\(^{88}\) See Whitbeck (1998) for a good overview of non-transferrable professional responsibilities of engineers.
Acting with Artefacts

This chapter will clarify the conditions of responsibility transfer between engineer and user by combining two existing theories into a new theoretical framework. The first theory concerns control over and responsibility for actions by Fischer and Ravizza (1998). The second theory is the use plan theory for technical artefacts by Houkes and Vermaas (2004). Fischer and Ravizza are not interested in applying their theories to artefacts and engineers, while Houkes and Vermaas are not concerned with responsibility. I will argue that (the communication of) use plans can account for the transfer of control over an artefact, and can thereby affect a transfer of responsibility for the artefact from the engineer to the user.89

Four caveats are in place here. First, it is not the purport of this chapter to analyse the distribution of responsibility within a specific case in engineering. The combined theoretical framework will be its main focus, though I will apply it to a test case to show how it could function in practice.

Second, this chapter is about moral responsibility for artefacts, not legal responsibility or liability. However, assuming that legal responsibility at least overlaps with moral responsibility, the findings of this chapter would be relevant for legal responsibility as well.

The third remark concerns the role of the individual engineer. My account uses a simplified model of engineering where one engineer creates a product for one user. This does not fit well with real practice, where teams of engineers often work together, embedded in institutional frameworks, and users might also be groups or institutions. As my framework is concerned with the transfer of control and responsibility between two parties, however, I will adhere to the simplified model of one engineer / one user for the sake of practicality. My analysis will show on which side the responsibility lies. A theory describing the distribution of responsibility within organizations could then be used to more accurately pinpoint the person(s) responsible.90 The test case will show how the theoretical framework works in a more complex situation.

90 See for example Royakkers, Grossi and Dignum (2006).
Finally, I mentioned that engineers are ‘causally responsible for the artefacts they produce’. While this may be trivially true – without the engineer, there is no artefact, engineers may not be responsible for all aspects of the artefacts they create, as they are themselves dependent on raw materials, machinery with which to build the artefacts, etc., which may all be delivered by third parties. This, however, does not affect my framework, which only determines whether responsibility lies on the engineer’s or the user’s side. If in a certain case responsibility is shown to be on the engineer’s side, it could well be that responsibility turns out to lie with the supplier: maybe the raw materials delivered lacked a promised quality. It would be interesting to see if the framework could be adapted to work on different levels, for example to distinguish between the responsibilities of engineer and supplier, but such a project lies outside the scope of this dissertation.

I will begin this chapter by constructing the theoretical framework. I will summarize the theory of control and responsibility by Fischer and Ravizza and the use plan theory by Houkes and Vermaas. Then, I will show that two assumptions have to be made to connect both theories, and that a strong case can be made for both. After that, I will apply the framework to a test case: the Abcoude dosing lock.

8.3. Responsibility and control

In Responsibility and Control, Fischer and Ravizza aim to ‘explore and develop systematically the conditions of application of the concept of moral responsibility.’ More specifically, they examine responsibility for actions, omissions and the consequences of those actions and omissions. While I will apply their account primarily to cases where actions have negative consequences, I agree with Fischer and Ravizza that responsibility can also elicit praiseworthiness, e.g.

91 I am grateful to Carl Mitcham for bringing this point to my attention.
92 According to Fischer and Ravizza, you are morally responsible when you are an appropriate candidate for the reactive attitudes (gratefulness, resentment, etc.). In this, they follow Strawson (1962).
when you are responsible for doing good (2). Also, the book of Fischer and Ravizza is primarily situated in the determinism/free will debate: I will extend their theory in another direction by applying their concept of responsibility to artefacts.93

Fischer and Ravizza start by examining the two conditions first formulated by Aristotle94 under which agents have no moral responsibility for what they do. The first condition is ignorance: if you are unaware of what you are doing or what the consequences can be, you are not responsible for those doings and their consequences. Of course, you can be responsible for being ignorant, reckless or failing to investigate what the possible consequences of your actions could be. The second condition is force: you are not responsible when you can not act freely, or more specifically, when you can not control your behaviour (13). It is this condition Fischer and Ravizza focus on when developing their account of responsibility.95

It is important to note that Fischer and Ravizza distinguish two kinds of force: resistible force like threats or coercions and irresistible force. Only the latter is strong enough to exempt someone from moral responsibility. This can happen either when the mechanism issuing in the action is not ‘the agent’s own’, for example when a brain implant controls my behaviour, or when I fail to be at least moderately responsive to reasons, for example when I am hypnotised.96 If I am ‘merely’ threatened, say, a robber puts a gun to my head and

---

93 For a critical examination of Responsibility and Control, see the book symposium in the journal Philosophical Explorations 8(2). For a good overview of the topics in the contemporary debate on responsibility and free will, see Fischer (1999).
94 Nicomachean Ethics (1985), 1109b30 – 1111b5.
95 These conditions are not only recognized by the Anglo-American analytical tradition in which Fischer and Ravizza write. Jonas (1979/1984) adopts them too as he writes: ‘The first and most general condition of responsibility is causal power, that is, that acting makes an impact on the world; the second, that such acting is under the agent’s control; and third, that he can foresee its consequences to some extent’ (90). While Fischer and Ravizza take the causal power of actions for granted, they view the recognition of an agent that her actions have causal power as a crucial part of her becoming a responsible agent.
96 Chapters 2 and 3 of Fischer and Ravizza’s book deal in more detail with what they mean by ‘moderate reasons-responsiveness’. Roughly, they claim that agents are moderately reasons-responsive when they acknowledge that at least in some situations there might be reasons to
Transferring responsibility for artefacts through use plans

demands my money, none of these conditions are met: I decide to give him my money because the fact that I will remain alive by doing so makes for a good reason to do so. I thus exercise control over my action of giving the robber my money, and am responsible for that action. Fischer and Ravizza mitigate this conclusion by stating that while I am responsible, given the circumstances I will probably not be considered blameworthy.

As control seems to be an instrumental notion to understand moral responsibility, Fischer and Ravizza set out to examine it. They distinguish two different kinds of control over actions: guidance control, which involves an agent’s freely performing an action and regulative control, which involves the power to exercise guidance control over an action and the power to exercise guidance control over another action instead (31).\(^7\) Usually, both forms of control are not clearly separated. Fischer and Ravizza give an example where they are: that of Sally taking driver’s lessons in a car with dual controls, one for her and one for the driving instructor (32). Imagine that the car approaches a right turn. Sally steers the car to the right. In so far as she guides the car, freely performing the action of steering to the right, she can be said to have guidance control over her action. However, if she would have turned left, or not turned at all, the instructor would have intervened and steered the car to the right instead. This means Sally has no regulative control, as she could not have caused the car to do anything else than go to the right. Only the instructor has regulative control. This does not mean that Sally is not responsible for her action: she is, because she freely performed it. Furthermore, both bear full responsibility for the consequences: Sally has guidance control over (and is responsible for) the consequences because she has guidance control over her action, and in this case it is reasonable to expect her to know what the consequences of that action will be (121). The driving instructor is responsible for the consequences because he is responsible for the omission of

choose another course of action than in the current situation, and that there should be a pattern to those reasons, e.g. ‘I will not buy this car if the price is over $5000.’

\(^7\) Guidance control could be further divided into direct and indirect control: see chapter 3. Fischer and Ravizza argue that no one actually ever has regulative control, as this would be incompatible with their deterministic view of the universe, but that this does not matter for our practices of ascribing responsibility.
an action: if he does not intervene, that should only be because he approves of Sally’s actions. Since he could have intervened and prevented the consequences, he is also fully responsible for their occurrence.

One more question needs to be answered here: where does responsibility come from? Fischer and Ravizza claim that initially, you have to take responsibility. They view the taking of responsibility as a vital step in the life of a human being. By recognizing yourself as an agent, realizing that the mechanism that issues in actions is ‘your own’, you take responsibility for those actions (210). Whoever does not do that will not be recognized as a person, but rather as ‘a distasteful object or a dangerous (or annoying) animal.’ (213). Dennett (1984) uses an engineering metaphor to illustrate the concept of taking responsibility: ‘I take responsibility for any thing I make and then inflict upon the general public... (...) I have created and unleashed an agent who is myself; if its acts produce harm, the manufacturer is held responsible.’ (85)

Fischer and Ravizza and Dennett talk about taking responsibility as a human being. Engineers, however, have special responsibilities over and above their basic responsibilities as human beings; this is emphasized by the concept of ‘role responsibilities’. According to the idea of role responsibilities, we each play many different roles in our society: we are not only humans, but also colleagues, parents, supervisors, customers, etc. Each role is accompanied by specific responsibilities which you have to take and internalize in order to properly fulfil that social role. In this light, ‘being an engineer’ can be seen as adopting a certain social role which requires specific training and commitments.

---

98 For a critical examination of this claim, see Judisch (2005).
99 For an elaboration of the concept of role responsibilities, see May (1992), chapter 9, or Mitcham and Von Schomberg (2000). Mitcham and Von Schomberg, however, find that the concept focuses too much on the individual to be useful for engineers. They propose an alternative called ‘collective co-responsibility’ to better fit engineering practice.
8.4. Use plans

While Fischer and Ravizza give a nice example of how both forms of control work, they take it for granted that the car works as it should and that Sally and the instructor have adequate knowledge of how it works.\textsuperscript{100} Also, there is no mention of the skills needed to drive a car: it seems that guidance control involves at least a basic skill in the action, but this is not explicated.\textsuperscript{101} This might be unproblematic for everyday actions like walking and pressing buttons, but especially in the operation of complex technical artefacts skills are important. Lastly, there is no mention of cases in which the engineer might be (partly) responsible, for example when the car would malfunction. How do engineers transfer knowledge about artefact functions to users? Which skills can they assume, and which should they mention explicitly? To deal with these questions I will now turn to the use plan theory (Houkes and Vermaas: 2004, 2010; Houkes: 2006).

The use plan theory is a theory of when design and use is rational. In order to rationally use an artefact, one has to have a certain kind of knowledge of the function of that artefact. Houkes and Vermaas argue that it is better to speak of knowledge of artefact use, and that this knowledge is different from ‘classical’ declarative or procedural knowledge. They call this knowledge ‘use know-how’, which has two components: ‘knowledge that a sequence of actions leads to the realisation of a goal, and the skills needed to take these actions.’ (Houkes: 2006, 105). The first component is called knowledge of a use plan.

Artefacts are designed for embedding in use plans (Houkes et al.: 2002; Houkes and Vermaas: 2004). Moreover, successful design requires the construction and communication of at least one rational use plan. This does not mean that such an artefact would only have one use plan: if a sequence of actions with that artefact will lead to the realisation of a certain goal, then that sequence is a use plan, whether it is explicated by the engineer or not. What it does mean is

\textsuperscript{100} Indeed, Yaffe (2000) notes that Fischer and Ravizza never directly address the issue of how physical constraints in general might affect moral responsibility. This chapter should at least help to fill the lacuna concerning the constraints those physical objects known as artefacts place upon us.

\textsuperscript{101} See the discussion on skills and basic actions in chapter 3, especially 3.6.
Acting with Artefacts

that the use plan constructed and communicated by the engineer defines the ‘standard use’ of that artefact (Houkes and Vermaas: 2004).

Use plans can be communicated from engineer to user in a variety of ways: textually (via user’s manuals or written instructions), through pictures or icons, hardwired in the design itself, et cetera. The engineer needs to communicate more information together with the use plan, however. According to Houkes and Vermaas, ‘in a rational plan, the user believes that the selected objects are available for use – present and in working order – that the physical circumstances afford the use of the object, that auxiliary items are available for use, and that the user herself has the skills necessary for and is physically capable of using the object’ (59). If any of these factors might not be a matter of course, the user should be alerted to it. When communicating the use plan of a car with a manual transmission in a country where automatic transmissions are dominant, for example, it should be mentioned that operating this kind of transmission requires a specific skill. That some countries do not allow you to drive a car with a manual transmission when you have taken your licensing test using an automatic transmission is information important for the social and legal context of driving, but it does not have to be communicated together with the use plan, as it is not necessary for ‘using’ the car itself: it is physically perfectly possible to drive a car with a manual transmission without a corresponding license – as long as you possess a minimal skill in operating manual transmissions.

While reasonable for practical use, Houkes and Vermaas’ requirements can become problematic. After all, there is no clear boundary between what can be regarded as ‘common knowledge’ and what is important or specific enough to mention together with the use plan. To avoid claims of legal responsibility, engineers prefer to include too much over too little. On the other hand, large manuals full of warning signs may deter rather than invite potential users, and the list of possible conditions influencing the operation of the artefact is endless: at some point, even the thorough engineer has to rely on the ‘common knowledge’ of the user. It might also be argued that the user has a certain (role) responsibility to acquaint herself with the intended use plan of the artefact and actively seek out how it is supposed to work. However, this does not diminish the responsibility of the engineer to communicate the use plan, for she still
Transferring responsibility for artefacts through use plans

needs to make the information accessible to the user in some way. In practice, what should be mentioned in a use plan seems to depend on a number of factors, for example the risk involved in improper use of the artefact.

Some knowledge might not be necessary to use an artefact, but might help to enhance its lifespan or efficiency. Knowledge of how efficient your engine is at specific speeds is not necessary to drive a car, but it can help you drive more efficiently, using less fuel, saving money and reducing emissions. This ‘supere-rogatory knowledge’ does not have to be communicated with any use plan as long as it does not significantly affect the artefact’s functioning. However, in so far as engineers have to hold paramount the safety, health and welfare of the public, they might be said to be responsible for communicating these aspects of artefact use as well.

What the use plan theory does not do is investigate what this transfer of knowledge about artefacts means for the moral responsibility of both engineers and users. This issue will be addressed in the next section.

8.5. Combining approaches

Here, I will argue that the theories of Fischer and Ravizza, and Houkes and Vermaas can be combined, and that communication of use plans can transfer responsibility from engineer to user by transferring guidance control over an artefact. For this combination, however, both theories need to be extended. The theory of Fischer and Ravizza deals with control over actions; in order to be relevant for engineers, it should include control over (and thereby, responsibility for) artefacts as well. The use plan theory needs to be extended to show that use plans not only transfer a procedure, but control as well to enable the transfer of responsibility. In this section, I will show that both extensions follow naturally from the existing theories.

What do we mean by ‘controlling an artefact’? Fischer and Ravizza seem to assume this can have two different meanings when they state about their example: ‘Sally controls the car, but she does not have control over the car (or the car’s movements)’ (32). Dennett (1984) makes a similar distinction in an example about an airplane: ‘...The pilot not only strives to control the plane at all times; he
acting with artefacts also engages in meta-level control planning and activity – taking steps to improve his position for controlling the plane by avoiding circumstances where, he can foresee, he will be forced (given his goals) to thread the needle between some Scylla and Charybdis.’ (62-63). I will argue that both forms of control correspond with both forms of control over action.

The first meaning of ‘controlling an artefact’ is really a subclass of control over actions, namely control over actions performed with artefacts. I control my car in so far as I exercise guidance control over the actions I perform with it.

The second meaning of ‘controlling an artefact’ is more similar to exercising regulative control over actions. Artefacts are not passive recipients of human action, but can exhibit behaviour of their own. ‘Controlling an artefact’ can then be seen as ensuring that the behaviour of the artefact does not interfere with the agent’s goals and keeps within certain limits of for example safety and sustainability. More specifically, by exercising regulative control in such a way, the agent will not end up in a situation in which the exercise of guidance control over actions with that artefact becomes impossible. For example, if I drive in a car and see an icy road ahead, on which I know I might go into a skid and lose control over the car, I can take measures to prevent this from happening such as putting snow chains on the tires. By (exercising guidance control over the action of) putting snow chains on the tires, I ensure that I will remain able to exercise regulative control over the car. Apparently, both forms of control over artefacts can be reduced to control over actions.102

This elaboration enables us to make more explicit what is meant with ‘being responsible for an artefact’. I will continue my parallel between actions and

---

102 At first glance, this might seem to be a departure from Fischer and Ravizza’s claim that we never have regulative control. However, the phenomenon underlying this meaning of ‘controlling an artefact’ can also be explained without regulative control: it is that, by exercising guidance control over an action now, we change what we can do and what can happen in the future (or according to Fischer and Ravizza: we sometimes change what we will do and what will happen, relative to the wholly counterfactual case in which we would not have exercised guidance control over that action now) (see also Illies and Meijers: 2009). Thus, I use ‘exercising regulative control’ here as similar to ‘choosing’ and ‘deciding’: useful to make sense of our behaviour in everyday practice, but if determinism is true, possibly an incorrect description of what actually takes place.
artefacts here. Fischer and Ravizza distinguish between three main forms of responsibility: for actions, for omissions and for the consequences of those actions and omissions. 103 I will treat responsibility for artefacts as being of a similar threefold nature, including responsibility for actions performed with artefacts, omissions of actions with artefacts, which may either be not performing an action with the artefact, or not intervening in the behaviour of the artefact, and the consequences of those actions and omissions of actions with artefacts.

Now to the second question: do use plans transfer control? To be more precise, I will rewrite this question to: does the communication of a use plan (for an artefact) transfer guidance control (over that artefact) to an agent? If we write the question out with the definitions of use plan and guidance control, we get the question: does communication of a sequence of actions that leads to the realisation of a goal transfer the ability to freely perform those actions to an agent?

The problematic part of this definition is the word ‘freely’. It would be bizarre to assume that use plans enable people to act freely if they couldn’t do that in the first place. To provide for this, I will assume that Houkes and Vermaas make the implicit assumption that the agents to whom the use plan is communicated enjoy freedom of action. Indeed, by definition an agent structurally without freedom of action could not be considered an ‘agent’ at all. 104 This assumption will account for the most problematic part of the question.

Apart from the freedom condition, there seems to be a second point of attention, namely that communication of an action does not necessarily lead to the physical or mental ability to perform that action. 105 I might be told how to drive a car, or ride a bike, or play golf, but that in itself is not sufficient. I also need to practice, or in other words, to gain procedural as well as declarative knowledge.

---

103 These are not the only forms of responsibility possible. Dennett (1984), for example, also deals with ‘responsibility for the self’.

104 An agent does not need to have freedom of the will to execute a use plan: the use plan might already have been constructed specifically with the goals of this agent in mind. The distinction between freedom of the will and freedom of action is elaborated upon in Frankfurt (1971).

105 I assume that all technical components needed to perform those actions are already included in the artefact (i.e. batteries) or readily available. After all, the use plan is explicitly made for a functional technical artefact.
This problem is removed by Houkes and Vermaas’ requirement that it should be mentioned whether specific skills or abilities are needed for the execution of a use plan, like the driving skills necessary to operate a car. In short, if freely performing an action, and thereby exercising guidance control over it, might be difficult or require training for an agent, it should be mentioned together with the use plan. The ‘ability’ in the question is thus reduced to not so much a physical or mental ability, as well as the ability to perform actions ‘under a certain description.’ For example, if I know nothing about cars, my actions with them are limited to ‘turning the wheel’ and ‘pressing the pedal on the right’. Knowledge of the use plan also allows me to intentionally perform the actions of ‘steering to the right’ and ‘accelerating’.

The theoretical framework thus obtained can be formalized as follows:

An engineer $E$ transfers moral responsibility for an artefact $A$ to a user $U$ if:

1. $E$ is morally responsible for $A$.
2. $E$ successfully communicates at least one rational use plan $P$ for $A$ to $U$.
3. $P$ can (under normal conditions) physically be executed with $A$.
4. $U$ is able to execute $P$.
5. $U$ has access to $A$.

Conditions (1) – (3) need to be met by the engineer to enable the transfer of responsibility. Conditions (4) – (5) need to be met by the user in order to accept that responsibility. In other words, when an engineer transfers responsibility to a user, the engineer of an artefact enables the user to take responsibility for that artefact. The user thus gains a (forward-looking) responsibility for that artefact,

106 See chapters 2 and 3 for an explication of the idea of actions under a description.
107 Specifically, this condition is about forward-looking moral responsibility.
108 Houkes and Vermaas see the construction and communication of at least one rational use plan as a prerequisite for ‘good design’.
109 While an engineer cannot exempt herself from responsibility by fulfilling conditions 1 – 3 without an able user present, users can take responsibility for an artefact (or any natural object) by fulfilling conditions 4 – 5, assuming they have acquired a use plan $P$ in some way and
Transferring responsibility for artefacts through use plans

so that if her actions or omissions of actions with that artefact cause harm, she can be held (backward-looking) responsible. This does not automatically mean that the engineer loses responsibility for the artefact: if conditions (4) and (5) still also hold for the engineer, the engineer will be in a similar position with regard to the user as the driver’s instructor is to the student, and there may be an expansion rather than a transfer of responsibility. If at least one of these conditions does not hold anymore, say, the engineer produces a car and the user drives away in it, leaving the engineer without access to the car, the engineer cannot be held responsible anymore for not acting with that artefact and the consequences thereof. Again, the engineer could still be held responsible for various other things like producing the car, selling it to that particular user, etc.

I will illustrate this framework with an example. Suppose an engineer E wants to make a car A for user U and also wants to transfer responsibility for this car to user U.\textsuperscript{110} First, E has to take responsibility for A. Within our society she has already (implicitly) done so by becoming an engineer, designing and constructing the car and providing it to U. By accepting her job, her ‘role in society’, she has accepted her role responsibilities, which include in this case being responsible for the artefacts she designs and constructs. Because of this, E is morally responsible for the car. Condition (1) has been met.

Second, E needs to communicate at least one rational use plan P for the car to U. E has several ways to do this and she will probably use more than one of them. A user’s manual might highlight specific features of the car, the car itself is made to prescribe or invite certain actions\textsuperscript{111}: pedals are made to be pushed, the gearstick can only be moved in such directions as to switch to certain gears and so on. Part of the use plan will probably be communicated via intermediaries such as driver’s schools, and commercials and advertisements can also alert the actually try to execute it (experimenting). Houkes and Vermaas (2004) analyse several ways in which users may acquire use plans.

\textsuperscript{110} I will leave out middlemen like car dealers to focus on the transfer itself: the section on the Abcoude dosing lock will deal with a more complex case. Alternatively, you could read ‘engineer’ as ‘the institution or organization that employs the engineer’. The ‘user’ can similarly be an institution or organization.

\textsuperscript{111} See chapter 5 for an elaboration on what these terms mean.
user to aspects of the use plan. Because different users might already know different parts of the use plan, some information might be redundant to some users, but this is less problematic than an incomplete communication of the use plan in which case condition (2) might not be met. By these different channels, a successful communication of at least one rational use plan $P$ for the car can take place.

Third, it must be physically possible under normal circumstances to execute $P$ with $A$. What normal circumstances are depends on what the artefact is made for: for a space shuttle, for example, they are quite different than for a submarine. This condition precludes the transfer of responsibility when the artefact doesn’t function as stated in the use plan, for example, when it would malfunction due to a construction failure that could reasonably have been prevented. When constructing the car, $E$ probably performs various tests and safety checks, which ensure condition (3) has been met, before she gives $U$ access to the car. In so far as these tests and checks are required by law, she would not only risk accountability by skipping them but also liability for defects in the car and the consequences thereof.

There are two more conditions to be met. (4) states that the user needs to be able to execute $P$. $P$ assumes a certain repertoire of basic actions, but its communication should include the mention that minimal driving skills are required to execute $P$. $U$ is thus warned that she cannot take responsibility for (driving with) the car unless she has the driving skills needed in order to execute $P$. Assuming that $U$ is rational and has no overriding reasons to try and use the car unskilled, she will heed this warning and not try to use the car until she has learned how to drive and is thus able to execute $P$.

Condition (5), finally, states that $U$ has to have access to the car: she should be physically able to get to the car and drive it onto the road in order to become responsible for it.

In this section I have argued that control over actions can be extended to control over artefacts as well, and that use plans can transfer control and thereby re-

112 See chapter 3.
transferring responsibility for artefacts through use plans

I have illustrated this with an idealized example of how the transfer of responsibility for a car can occur. In the next section I will bring up a real-life case to further test the theoretical framework, clarifying the distribution of responsibility and identifying issues where further research might be necessary.

8.6. The Abcoude dosing lock: a test case

Responsibility for traffic safety is a complex issue. In this section I will examine a test case with multiple users, where one of the users is a municipality, which has installed a traffic safety system affecting the behaviour of road users. In this test case, control is transferred to the municipality but not to the road users. For them, the use plan does not transfer control; rather it communicates the fact that road users do not have control over certain actions.

In 2006, the municipality of Abcoude, The Netherlands, installed a ‘dosing lock’ or movable bollard (doseersluis) on one of the exit roads. The goal was to reduce cut-through traffic and thereby increase traffic safety in the village centre. The dosing lock consisted of a narrowing of the road, a traffic light, warning signs that only one car could pass every time the light turned green, and a moveable obstacle that would block the road when the light turned red, to sink back into the road when the light turned green. The dosing lock was activated only during rush hours (Abcoude: 2006).

The dosing lock was very successful in reducing cut-through traffic. It had an undesirable side effect, however: within half a year, over forty cars had crashed on the obstacle, leading to leakage of oil and dangerous chemicals, many traffic jams and drivers bypassing them by driving over the bicycle path.113 The lock was disabled for some time while the municipality took extra measures to alert drivers to the obstacle, which apparently led to a decrease in the number of accidents.

113 The main cause of these accidents was people driving through the red light (project leader public works Abcoude, private correspondence).
Who would be responsible, according to the theoretical framework? First, it seems that the engineers have successfully transferred responsibility to the user, the municipality. The engineers have taken responsibility for building the dosing lock (1), and communicated its use plan to the municipality (2), thereby giving the municipality control over the dosing lock – and responsibility for it as well. This use plan could physically be executed with the dosing lock (3). The municipality was able to execute the use plan to realize its goals (4). It also had access to the dosing lock (5).

Now, the municipality took on the role of ‘engineer’ by having the artefact implemented in a road system used by other users. The municipality thus had a dual role as user (of the dosing lock) and engineer (of the public infrastructure, in this case, the road-with-dosing lock): I will regard their situation as comparable to an engineer who uses existing components to construct a new artefact, commonly called ‘off-the-shelf engineering’, or as Houkes et al. (2002) put it, ‘brochure engineering’.

The municipality did construct the road-with-dosing lock for the users, mainly car drivers, of that particular road. Their intention was also to transfer responsibility for the road-with-dosing lock to the road users.

Condition (1) was met in so far that the municipality had taken responsibility for the road-with-dosing lock as part of its role responsibility for maintaining traffic safety. The communicated use plan for the road-with-dosing lock could physically be executed with it (3). The road users had the ability to use it, in this case, to pass the dosing lock safely (4). Also, the road-with-dosing lock was accessible to them, indeed, the logical choice to reach some particular destinations (5).

The main complaint of the road users concerned condition (2): they felt that the municipality had failed to communicate certain aspects of the (rational) use plan to them. One particular problem seemed to be that the use plan for the traffic light of the dosing lock differed from that of regular traffic lights. Driving through a regular traffic light is punishable by law, but physically possible. In this case, many road users mistakenly thought that they could exercise guidance control over the action of driving on, as would be the case with a regular traffic light. However, those who tried to drive on found out that in fact the obstacle
forcefully prevented them from doing so.\textsuperscript{114} As the road users claimed they hadn’t realize that, they held the municipality responsible for the results of their actions, in particular the damage done to their cars. The municipality didn’t agree, but they did close the lock down for some time in order to increase the salience of the obstacle by several measures. In other words: they sought to better communicate a rational use plan for the road-with-dosing lock to the road users. In 2010, however, due to continuing problems it was decided to completely remove the dosing lock and take other measures to deal with cut-through traffic.\textsuperscript{115}

In this test case, the framework has shown that the fault lay not so much with the engineers who designed the dosing lock as such, as well as with the municipality who placed it in a context, designing the road-with-dosing lock, making it clear on which level the success of the transfer of responsibility was in question. In particular, the framework has given a possible reason for why the traffic light of the dosing lock wasn’t as effective as it should have been, namely that its use plan differed from that of ‘ordinary’ traffic lights, indicating that special attention would have been needed for communication of the differences.

\section*{8.7. Conclusion}

In this chapter I have argued that certain responsibilities of engineers for artefacts can be transferred to users. I have also shown how this transfer can take place. I have done this by first summarizing the relevant parts of the theory of responsibility and control by Fischer and Ravizza and the use plan theory by Houkes and Vermaas. I have then shown how both approaches could be combined to support my thesis. I have given an example of how the framework functions and demonstrated its practical merit in helping to indicate the responsible party in the Abcoude dosing lock case.

\textsuperscript{114} The municipality \textit{did} provide emergency services and regional buses with a remote control, with which they could deactivate the dosing lock if they had to pass quickly.

\textsuperscript{115} Financial reasons weighed heavily in on this decision: the repair and/or replacement of the dosing lock cost a lot of time and money after each crash, as did two cases of suspected arson. Furthermore, in January 2009, the magistrate of Utrecht found the municipality partially liable for two out of three examined accidents, as the safety measures taken at the time of the accidents were judged insufficient.
Acting with Artefacts

At the beginning of this chapter, I mentioned the two ‘Aristotelian conditions’ which exempt an agent from responsibility: being ignorant and being forced. Fischer and Ravizza started out working on the second, but the use plan approach seems to have brought us to the first. After all, the communication of a use plan is a transfer of knowledge, relieving the agent of ignorance about an artefact. Is this shift strange? Not if we notice the overlap between the force and ignorance conditions: if an agent has no clue about how an artefact works, he is ‘forced’ to submit to its behaviour just as much as he is ‘forced’ to submit to the laws of nature. However, where knowledge about the laws of nature does not give an agent the ability to control them, knowledge about the workings of an artefact can give the agent control over that artefact, the ability to perform actions under various descriptions with it, and thereby, responsibility for it.
9 Conclusion

The question I have answered in this dissertation is: What is an action with an artefact, and how do artefacts influence agents and agency? I have defended the action identification thesis in chapter 2 and derived a ‘standard view’ on what actions with artefacts would be from it. Particularly, I have claimed that an action with an artefact is a doing, that is done by doing something else or done directly, that is intentional under at least one description that describes the action in terms of artefact use. According to Davidson, the description of an action in terms of artefact use would always be a description of the action in terms of its consequences, as all actions are bodily movements. I have argued against that view and proposed an alternative in chapter 3. Particularly, I have argued that a description of the action in terms of artefact use may also be the teleologically most basic description: this is the case if the artefact is intentionally used directly, not by doing something else (say, hammering a nail into a board or prodding an obstacle with a blind man’s cane.) For this, the artefact has to become transparent in use, which seems to require incorporation of the artefact in the body schema.

Chapters 5 to 7 have analysed how artefacts can influence individual (chapters 5 and 6) and collective (chapter 7) agents and agency. In chapter 5 I have argued that artefacts can influence agents by altering their reasons for action, while in chapter 6 I have argued that they can influence agency by altering agents in their capacity as bounded practical reasoners. In chapter 7 I have suggested an additional way in which artefacts may influence individual agents: by helping them to keep their commitments to certain actions. Here, it is not so much the bounds of reasoning that are affected as well as the bounds of our capacity to act on what we regard (or regarded at some point) as our best reasons for action. Additionally, artefacts could influence collective agency by keeping individual participants committed to the collective action and thus, increase coherence and stability of the collective agent.

While acceptance of my answer regarding the first part of the main question does not commit one to accept my answer regarding the second part of the main
question or vice versa, both answers offer independent reasons for action theory to further investigate the role of technology and technical artefacts in what we do and why we do it. We are not ‘merely’ embodied agents: we are intrinsically tool-using animals who influence our own and each other’s actions and agency by technologically shaping our environment and, to some degree, even ourselves. Any theory of action which seeks to have practical application would do wise to recognise and incorporate this fact.

Besides addressing the main question, this dissertation has given rise to two interesting trends that I wish to highlight in this section. I will end this section by suggesting some questions for further research for action theory that this dissertation has raised.

The first trend is a move away from the physical body and bodily movements as the central focus for action theory: this trend is most notable in Chapters 3 and 6. This will come as no surprise as the starting point of my investigation lies at the intersection of the philosophy of action and the philosophy of artefacts. The philosophy of artefacts is primarily concerned with functionality, and only secondarily with structure – the structure is constrained by the functionality the artefact needs to have. Our bodies are not artefacts, nor are they designed, but they do have a certain functionality by virtue of their structure, and we use them continually in the same way and for the same reason as we use tools: to attain particular ends (which may simply be engaging in pleasant activities). Granted, we have a kind of special authority (or privileged use) over our bodies in the sense that we can do some things ‘directly’ with them – but this special authority is not restricted to our bodies, as I argue in Chapter 3, nor is it something we have over all our body parts.

This blurring of the boundary between artefacts and the body is not new. Aristotle already called the hand the ‘tool of tools’. Clark (2003) goes so far as to say we are tools all the way down, while Merleau-Ponty (1945/2002) has argued in

\[\text{116} \text{ Of course, functionality is also constrained by structure, but designers choose the materials of artefacts and design their structure according to their intended function or use (e.g. Houkes et al.: 2002; Hilpinen: 1993).}\]
the other direction, that tools can extend our body. Affordance theory (Gibson: 1979) and its contested notion of effectivities, properties or capacities that the animal needs to have in order to act on certain affordances (Turvey and Shaw: 1979; Michaels: 2003), can easily accommodate the change of opportunities for action when an embodied being picks up a tool. The action scheme similarly does not distinguish between action opportunities with and without artefacts (Illies and Meijers: 2009). All these developments suggest that action theory might miss out on interesting opportunities to further the field if it maintains a strict focus on the body and bodily movements.

The second trend that has emerged in this dissertation is the minimalisation of the differences between natural objects and technical artefacts due to the focus on use rather than function. When function (roughly, what the artefact should do or be used for) is taken as a starting point for conceptual analysis into the nature of technical artefacts (Kroes and Meijers: 2006), there is an immediate difference between technical artefacts and natural objects: the first are designed to fulfil a certain function (and can thus malfunction, be used inappropriately, etc.), where the second are not.117 Some questions that then arise are whether the function of an artefact can change independently of designer intentions (Preston: 1998; Scheele: 2005; Vermaas and Houkes: 2006; Schyfter: 2009), and whether natural objects that are used intentionally have (gained) a function.

When actions with artefacts, or more specifically, artefact uses, are taken as a starting point for examining the nature of technical artefacts, no interesting difference is discernible anymore between artefacts and natural objects. Both can be acted upon and acted with. Both can figure in basic actions (chapter 3). Both have affordances (chapter 4): this is why the affordance concept is usable in the psychology of perception of both the natural and the artificial environment, as well as in design sciences. Both artefacts and natural objects can influence agents by altering their reasons for action (chapter 5) and can be used in main-

117 Biologists do speak of organs, body parts, etc. as having functions, but notions of biological (organ) functions differ from those of technological (artefact) functions (Vermaas and Houkes: 2003).
taining joint commitment (chapter 7; e.g. we might agree to work on our tax return on a rainy day, trusting that the rain will help us maintain our commit-
ment and not go out for a walk together instead). Of course, artefacts can enable us to do things like flying to the moon, and shift the bounds of agency in ways that are impossible to do with unaltered natural objects. But this is a practical rather than a conceptual issue: it is possible, though unlikely, that a freak natural process produces a space shuttle that could be used to fly to the moon with, and the active substances in enhancement medication are often naturally produced by particular kinds of plants. In other words: when designers turn raw materials into designed objects, they make use of processes that could occur spontaneously in nature as well (and how could it be otherwise?) The difficult questions for function theorists that I mentioned earlier are easy to answer from a use perspective: how an artefact is used can clearly differ from how its designer intended it to be used, and that natural objects can be intentionally used is no problematic claim either.

In light of these considerations, it might seem that I should have called my dissertation *Acting with Objects* rather than *Acting with Artefacts*. Yet, I have not chosen this title for alliteration purposes only. As artefacts are intentionally designed for a purpose, actions with artefacts have a normative dimension that actions with instruments and natural objects lack: artefacts can be used properly or improperly, according to function or not, irresponsibly or not, et cetera. This is the result of the intention of the designer and whether the designer has successfully communicated at least one rational use plan to the user. Some of these normative aspects have been investigated in Chapter 8 and also Chapters 4 and 5, and the notion of function seems very useful for capturing them. Never-
theless, the main focus of the dissertation has been on what actions with artefacts are, and how artefacts influence agents and agency, and with regard to those questions, there is no interesting difference between artefacts and natural objects.

Finally, let me consider some questions for further research that arise from this dissertation. Primarily, an important question arises from the claim that not all actions are bodily movements. I have argued in chapter 3 that it would be useful
to focus on the notion of (motor) skill to determine what we can do directly, and thus, determine our basic actions. This would open up new possibilities to connect with the psychology of skill and skilled behaviour, and possibly forge a closer connection between physical and mental actions: one can be skilled in actions of both kinds. Apart from that, a good notion of skill would be helpful in discussions on the role of luck in intentional action, which is often regarded as its counterpart (e.g. Mele: 2003; Knobe: 2006). Characteristically, skilled task performance is highly automated, smooth, precise and quick, while lucky task performance characteristically lacks these features, but is nevertheless successful. Unfortunately, skill as a philosophical concept seems under-investigated, especially when compared to luck. While it has received some attention in the phenomenological tradition (Merleau-Ponty: 1945/2002; Dreyfus and Dreyfus: 1986; 2005), it rarely seems to be explicitly investigated within the analytic tradition, except possibly under the broader Rylean header of ‘know-how’ (some exceptions are Dennett: 1984; Clarke: 2010; Montero: 2010). So the question what skill, a skilled action, or skill in a particular task is, would be relevant to address.

Related to and relevant for the notion of skill are notions of guidance and control (see e.g. Frankfurt: 1978; Dennett: 1984; Fischer and Ravizza: 1998). Agents may perform actions without being skilled in them, as long as they have some measure of control or can exert guidance over what they are doing. Not only is control or guidance seen as a prerequisite for actions, but it is also seen as a prerequisite for moral responsibility (Fischer and Ravizza: 1998): one cannot be held responsible for what happens ‘outside one’s control’, except when one is indirectly responsible, by relinquishing control earlier on.

Unfortunately, authors writing on control over actions tend to emphasize, again, the ‘special’ kind of control, direct control, one can have over one’s body parts, as opposed to the indirect control we supposedly have over artefacts (e.g. Vollmer: 2001; Mossel: 2005). There are two problems with this. The first is that, just as we can perform basic actions with artefacts, it seems that we can control some artefacts directly (see chapter 3). The second is that, even if we ignore the first problem for the moment, there seem to be at least two kinds of control we have over our bodies – apart from the indirect control we have e.g.
over our heart and intestines. The notion of skill helps to lay bare this distinction: becoming skilled in a task usually involves training and building up motor routines that one can execute without monitoring them, that let one focus on the task at hand rather than on the specific bodily movements that are involved in executing it. In other words, many bodily movements involved in skillfully executing a task are under automatic control. Our conscious control, meanwhile, operates at a higher level, monitoring progress on the task, keeping an eye out for possible distractions or complications, et cetera. It would be useful for any notion of skill, and any account of actions or responsibility based on control, to investigate how these different forms of control relate, whether they can both be properly called control, and whether there is a meaningful difference between direct control and indirect control that does not depend on a special status of the body.

Related to this is the question about what impact theories of direct perception should have on philosophical theories of action. Roughly, direct perception claims that there are links from perception to action that work without conscious interpretation of what is perceived or reasoning how the perceived object should be acted upon: affordance perception as treated in chapter 4 is a form of direct perception. Direct perception originates in the work of Gibson (1979) and was subsequently developed in work by Ungerleider and Mishkin (1982), Milner and Goodale (1995), Jeannerod (2006), and others. Clark (2001, 2007) has emphasized its importance for the philosophy of mind and perception. If the calibration of our bodily movements indeed happens via direct perception, without intervening conscious control, it might well impact theories of action in general and those based on the concept of control specifically. Moreover, if specific bodily movements are guided by the environment rather than intentionally controlled, it would be one more reason to claim that the explanation of bodily movements itself can only be a part of the explanation of intentional actions.

A last suggestion for further research concerns the systems perspective. While I have indicated why taking a systems perspective raises problems of its own, these problems could also be conceived of as challenges to overcome. Moreover,
Conclusion

the field of collective intentionality has long been interested in collective agents, and if the ‘unit of agency’ could reasonably be extended to include multiple humans, it might be extended along the same lines to human-artefact systems. Such an extension would probably be most successful if only restricted areas of application were considered, where some of the problems of the systems perspective would have less impact. For example, if one studies, say, drivers’ behaviour when approaching and navigating a traffic calming measure, one can assume that driver-car systems will usually remain stable and well-defined throughout the period of observation, which makes them good candidates to be units of agency. Taking a systems perspective here would be especially useful if it would turn out that individual drivers adopt different driving styles when driving different types of cars.

One interesting implication of this approach would be that artefacts would not only be subject to norms like effectivity and efficiency (Vaesen: 2008), but also to norms that govern the constitution of agency (Korsgaard: 2009). That is, if artefacts together with humans constitute agents, they should contribute to or at least not diminish e.g. the agent’s rationality, or ability to act according to the categorical and hypothetical imperative. This requirement already holds for enhancement technologies (chapter 6) that change humans who are agents, but if human-artefacts systems in general are regarded as agents, the scope of this requirement vastly increases.

If memory serves, it was science fiction writer Philip K. Dick who once wrote a short story in which the U.S. president truly acts on behalf of his people: a brain implant determines his behaviour according to what the people (as determined by the polls) want him to do or say. If president Truman had had such an implant when he signed the order to bomb Hiroshima and Nagasaki, it is obvious that ascriptions of intentionality and responsibility would have been very different, and Anscombe’s *Intention* would have been of much less help in sorting them out. While brain implants are still mostly the stuff of science fiction, this dissertation has shown that technical artefacts can influence what we do, how we do it, and affect the degree to which we are responsible for our actions and their consequences. It is therefore my expectation that, should the
U.S. president ever decide to take such a brain implant, this dissertation can help us work out the action-theoretical consequences.
References


Acting with Artefacts


References


Acting with Artefacts


Acting with Artefacts


Acting with Artefacts


Acting with Artefacts


References


Acting with Artefacts


Acting with Artefacts


## Index of Names

<table>
<thead>
<tr>
<th>Name</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akrich, M.</td>
<td>86</td>
</tr>
<tr>
<td>Anscombe, G.E.M.</td>
<td>1, 7, 15, 52, 60, 62</td>
</tr>
<tr>
<td>Bennett, J.</td>
<td>19</td>
</tr>
<tr>
<td>Bostrom, N.</td>
<td>105</td>
</tr>
<tr>
<td>Bratman, M.E.</td>
<td>119, 128</td>
</tr>
<tr>
<td>Clark, A.</td>
<td>37</td>
</tr>
<tr>
<td>Dancy, J.</td>
<td>81-85</td>
</tr>
<tr>
<td>Davidson, D.</td>
<td>6, 15-16, 19, 21, 27, 33-34, 38-39, 45-47, 62</td>
</tr>
<tr>
<td>Dennett, D.</td>
<td>148, 151-152</td>
</tr>
<tr>
<td>Dancy, J.</td>
<td>81-85</td>
</tr>
<tr>
<td>Gibson, J.J.</td>
<td>60-61, 63, 67, 70, 72</td>
</tr>
<tr>
<td>Gilbert, M.</td>
<td>120-121, 127, 131</td>
</tr>
<tr>
<td>Hardin, G.</td>
<td>134-136</td>
</tr>
<tr>
<td>Horneby, J.</td>
<td>19-20, 40-43</td>
</tr>
<tr>
<td>Houkes, W.N.</td>
<td>149-150</td>
</tr>
<tr>
<td>Kass, L.</td>
<td>105</td>
</tr>
<tr>
<td>Latour, B.</td>
<td>86-87, 139</td>
</tr>
<tr>
<td>Lombard, L.B.</td>
<td>17, 21, 30</td>
</tr>
<tr>
<td>Mackie, D.</td>
<td>19, 22-26</td>
</tr>
<tr>
<td>Maier, J.R.A.</td>
<td>69</td>
</tr>
<tr>
<td>Merleau-Ponty, M.</td>
<td>35, 49-50</td>
</tr>
<tr>
<td>Morreal, J.</td>
<td>28</td>
</tr>
<tr>
<td>Mossel, B.</td>
<td>21-26, 47-48</td>
</tr>
<tr>
<td>Norman, D.</td>
<td>60-61, 66, 71-72</td>
</tr>
<tr>
<td>Ravizza, M.</td>
<td>145-148</td>
</tr>
<tr>
<td>Sandis, C.</td>
<td>27</td>
</tr>
<tr>
<td>Thomson, J.J.</td>
<td>16, 22-23, 27, 29</td>
</tr>
<tr>
<td>Turing, A.M.</td>
<td>101</td>
</tr>
<tr>
<td>Velleman, J.D.</td>
<td>119, 125, 127</td>
</tr>
<tr>
<td>Verbeek, P.-P.</td>
<td>37, 52, 91-94, 96, 99</td>
</tr>
<tr>
<td>Vermaas, P.E.</td>
<td>149-150</td>
</tr>
<tr>
<td>Vollmer, F.</td>
<td>47-48</td>
</tr>
<tr>
<td>Weintraub, R.</td>
<td>27</td>
</tr>
<tr>
<td>Wierzbicka, A.</td>
<td>28-31</td>
</tr>
<tr>
<td>Wreen, M.</td>
<td>43</td>
</tr>
</tbody>
</table>
Summary (in English)

Technical artefacts are ubiquitous in modern society, and we act with them all the time. Philosophical action theory, however, has rarely examined the role technical artefacts play in the action process. This has left various interesting phenomena under-investigated, like the sensation that some artefacts extend our bodies while we act with them, and the influence artefacts can exert on both individual and collective agents. The main question I ask in this dissertation is therefore: What is an action with an artefact, and how do artefacts influence agents and agency?

The methodology I use in this dissertation is mostly conceptual analysis, though I draw heavily upon studies in the (neuro)psychology of action and tool use and studies in artefact design in order to generate results that are practically applicable. This dissertation also contains a more extensive case study where I examine the responsibility of engineers and users for the use of a traffic calming measure.

The dissertation is structured as follows. In the first part of this dissertation I address the first part of the main question: what is an action with an artefact? Specifically, I start by developing a ‘standard view’ of what an action with an artefact would be and defend it against objections based on the problem of the time of a killing. I then proceed to criticise this standard view and propose an alternative. In particular, I argue against Davidson’s claim that all actions are bodily movements and argue for the possibility of (basic) actions with artefacts. Finally, I show how a clear notion of actions with artefacts can unify the concepts of affordance used in ecological psychology and design science. I do this by building a model, the nested affordances model, that can ground and specify design recommendations.

In the second part of the dissertation I address the second part of the main question: how do artefacts influence agents and agency? Previous accounts of this influence, like those of Latour and Verbeek, explain this influence by extending concepts like ‘intention’ and ‘agency’ to artefacts or artefact-human systems. I offer an alternative account of this influence in terms of artefacts altering our reasons for action that does not depend on such an extension. I then
show that artefacts can also influence agency in a more radical way, namely, by influencing us in our capacity as bounded practical reasoners. I argue that the distinction between artefacts that can influence our capacity for practical reasoning and those that cannot allows us to distinguish artefacts that are especially morally problematic with regard to human enhancement from those that are not. Next, I apply these insights about the influence of artefacts on agents and agency to the field of collective, rather than individual, action. Particularly, I argue for a revised notion of joint commitment that is strengthened by including applying intentional, social and technical strategies to maintain this commitment. In this way, artefacts can influence not only individual but also group agents and agency. Finally, I zoom out from action theory and the artefacts themselves to address the issue of designer and user responsibility for actions done with artefacts. I do so by showing how responsibility can be transferred from designer to user by connecting Houkes and Vermaas’s use plan theory of artefact use and design to Fischer and Ravizza’s theory of responsibility.
Nederlandstalige samenvatting

Mensen zijn uniek onder de dieren vanwege het gemak en de inventiviteit waarmee ze gereedschap ontwerpen en inzetten om hun doelen te bereiken. We maken en gebruiken een enorme verscheidenheid aan technische artefacten, of technische gebruiksvoorwerpen, van knoflookpersen tot kernraketten. Toch heeft de filosofische handelingstheorie weinig aandacht voor de rol die artefacten (kunnen) spelen in het handelingsproces. Hierdoor zijn diverse interessante fenomenen filosofisch onderbelicht gebleven, zoals de ervaring dat we met sommige artefacten handelen alsof het verlengstukken van ons lichaam zijn: denk aan de vaardigheid waarmee blinden hun weg kunnen vinden met behulp van een blindenstok, of de beheersing van professionele tennissers over hun rackets. Een ander voorbeeld is de observatie dat artefacten niet alleen passieve instrumenten zijn die wij gebruiken om onze doelen te bereiken: ze ‘sturen’ ook ons gedrag, zoals de verkeersdrempel die te hard rijden ontmoedigt, of de sportwagen die er juist toe uitnodigt. De vraag die in dit boek centraal staat, is dan ook: Wat is een handeling met een artefact, en hoe oefenen artefacten invloed uit op handelaars en hun handelingsbekwaamheid?

De methodologie die ik gebruik in dit boek is voornamelijk conceptuele analyse, wat betekent dat ik naga wat er precies bedoeld wordt met woorden zoals ‘handeling’ en ‘sturing’ van ons gedrag door artefacten, en wat de implicaties daarvan zijn. Als alle handelingen bijvoorbeeld, zoals Davidson betoogt, lichaamsbewegingen zijn, betekent dat dat er geen handelingen met artefacten mogelijk zijn. Volgens Davidson zouden wij uitsluitend met ons lichaam handelen, en dit zou dan diverse gevolgen in de wereld hebben, waaronder het aan- of uitspringen en functioneren van onze technische artefacten.\footnote{Ik bekritiseer Davidsons claim in hoofdstuk 3.}

Daarnaast maak ik veel gebruik van onderzoek binnen de (neuro)psychologie van handelingen en ontwerpwetenschappen zoals ergonomie om mijn filosofi-
Acting with Artefacts

sche ideeën te toetsen aan inzichten uit de praktijk. Hiermee streef ik er ook naar tot resultaten te komen die praktisch bruikbaar zijn, zoals het nested affordances-model in hoofdstuk 4 dat bedoeld is om het begrip ‘affordance’, of ‘gelegenheid om te handelen’, te verduidelijken voor ontwerpwetenschappers. Dit boek bevat ook een meer uitgebreide case study, waarbij ik onderzoek in welke mate ingenieurs en gebruikers verantwoordelijk gehouden kunnen worden voor een serie ongelukken die hebben plaatsgevonden bij de doseersluis van Abcoude.

Het boek is als volgt opgebouwd. In het eerste deel van het boek behandel ik het eerste deel van de hoofdvraag: Wat is een handeling met een artefact? Hiervoor stel ik in hoofdstuk 2 allereerst vast wat een handeling met een artefact zou zijn volgens de handelingstheorie van Anscombe en Davidson. Centraal in de handelingstheorie van beide filosofen staat de idee dat we dingen kunnen doen door andere dingen te doen, zoals het licht aandoen door een knop in te drukken. Ik verdedig deze idee tegen kritiek gebaseerd op het probleem van het vaststellen van het tijdstip van een moord. Dit probleem is het volgende: stel, ik schiet mijn gehate rivaal neer op zaterdag; hij wordt naar het ziekenhuis gebracht, maar sterft op zondag. Critici van Anscombe en Davidson betogen dat hun theorie ons committeert aan de aanname dat ik mijn rivaal op zaterdag heb vermoord (door hem neer te schieten), een dag voor hij stierf, wat absurd zou zijn. Ik betoog dat een taalkundige analyse van ‘vermoorden’ ons helpt om te zien waarom we in dit geval hooguit kunnen zeggen dat ik mijn rivaal heb vermoord in het weekend.

In hoofdstuk 3 bekritiseer ik zelf de conceptie van handelingen van Anscombe en Davidson. Ik richt mij bij het eerste met name op Davids ons claim dat alle handelingen lichaamsbewegingen zijn, en beargumenteer dat handelingen met artefacten mogelijk zijn. Davidson baseert zijn claim op de aanname dat we alleen lichaamsbewegingen direct kunnen uitvoeren, terwijl we handelingen met artefacten indirect uitvoeren: we doen het licht aan door met onze vinger op de knop te drukken, maar we bewegen onze vingers en handen direct, niet door iets anders te doen. Ik laat echter zien dat er in sommige gevallen, zoals bij de
blinden met hun stok en de professionele tennisers, reden is om aan te nemen dat zij direct kunnen handelen met hun blindenstok of racket.

Tenslotte laat ik in hoofdstuk 4 zien hoe een heldere notie van handelingen met artefacten de verschillende concepten van affordances, gelegenheden om te handelen, kan verenigen, zoals ze gebruikt worden in de ecologische psychologie en de ontwerpwetenschappen, met name in affordance-gebaseerd ontwerp. Het is van cruciaal belang dat gebruikers weten wat ze met een artefact kunnen doen en hoe ze dat kunnen doen, met andere woorden, dat ze de affordances van het artefact eenvoudig kunnen waarnemen. Hoe deze affordances precies omschreven moeten worden is echter niet altijd even duidelijk: biedt een lichtknop de gelegenheid om hem in te drukken, het licht aan te doen, de kamer te verlichten, of alledrie tegelijk? Ik toon met een model, het nested affordances model, hoe deze beschrijvingen van affordances met elkaar verbonden zijn, wederom gebaseerd op de idee van Anscombe en Davidson dat we vaak dingen kunnen doen (met artefacten, in dit geval) door andere dingen te doen (met diezelfde artefacten). Ik laat zien hoe hiermee bestaande aanbevelingen uit affordance-gebaseerd ontwerp gerechtvaardigd en nader gespecificeerd kunnen worden, zoals de aanbeveling om feedback te geven aan de gebruiker.


119 Dit voorbeeld is bedoeld ter illustratie van mijn analyse en keert niet terug in de hoofdtekst.
Acting with Artefacts

geweren. Hoewel dit een relevante observatie lijkt, heeft zo kijken naar handelingen ook nadelen: het past bijvoorbeeld niet bij onze gewoonte om *mensen*, niet geweren of mensen-met-geweren, verantwoordelijk te houden en te straffen voor moorden.120 In hoofdstuk 5 bied ik een alternatieve verklaring voor de invloed die artefacten uitoefenen op handelaars, waarbij artefacten handelaars beïnvloeden door onze redenen om te handelen te veranderen. Een geweer kan mij bijvoorbeeld ‘uitnodigen’ om mijn gehate rivaal te vermoorden doordat het de moord mogelijk maakt (met mijn blote handen kan ik mijn rivaal niet aan), redenen tegen de moord wegneemt (bijv. het feit dat mijn rivaal onnodig zal lijden als ik hem op een andere manier ombreng), en/of me redenen geeft voor de moord (bijv. het feit dat het een manier is om mijn eigen leven te redden als mijn rivaal met zijn eigen geweer op me afkomt.) Ik laat zo zien dat de manieren waarop geweren het gedrag van mensen beïnvloeden prima te begrijpen zijn, zonder dat we ‘mensen-met-geweren’ moeten gaan beschouwen als aparte entiteiten die zich anders gedragen dan ‘mensen’. In mijn verklaring vermoorden geweren dus evenmin mensen – maar ze oefenen wel degelijk invloed uit op onze handelingen.

In hoofdstuk 6 laat ik zien dat artefacten ook op een meer ingrijpende manier onze handelingsbekwaamheid kunnen beïnvloeden, namelijk door te veranderen hoe wij binnen onze natuurlijke beperkingen, zoals een beperkte tijd en een beperkt redenatievermogen, tot handelen komen. Artefacten die dit doen zijn bijvoorbeeld de ‘concentratie-pil’ Ritalin, die ons in staat stelt om beter te concentreren, of implanteerbare electrodes in ons brein die bijv. ingezet kunnen worden om onze stemming te verbeteren. Veel van deze technologieën kunnen zowel gebruikt worden als therapeutisch middel, om ziektes of aandoeningen te behandelen, als als verbeterend middel, om de capaciteiten van gezonde mensen te verbeteren. Ik beargumenteer dat we een relevant onderscheid kunnen maken tussen artefacten die op deze ingrijpende manier veranderen hoe wij tot handelen komen, en artefacten die op andere wijze invloed uitoefenen op ons

120 Hierbij gaat het om de verantwoordelijkheid voor specifieke moorden. Dit is compatibel met bijv. wapenfabrikanten of de wapenwetgeving (deels) verantwoordelijk houden voor wapenge-relateerd geweld in de maatschappij. Zie ook hoofdstuk 8.
handelen (zoals het geweer). Dit onderscheid stelt ons in staat uit te vinden welke artefacten in het bijzonder moreel problematisch kunnen zijn met betrekking tot het verbeteren van onze fysieke en mentale capaciteiten, en het uitbreiden van ons repertoire van dingen die we kunnen doen.

In hoofdstuk 7 maak ik de stap van individuele naar collectieve handelingen, waarop ik voornoemde ideeën toepas over hoe artefacten invloed uitoefenen op wat we doen. Ik stel een nieuwe notie voor van betrokkenheid bij een gezamenlijke handeling, die hogere eisen stelt aan de betrokkenheid van handelaars dan eerdere noties dat doen. In het bijzonder vereist deze notie het toepassen van intentionele, sociale en technische strategieën om een blijvende betrokkenheid te garanderen. Intentionele strategieën vertrouwen op bepaalde gedachten of een bepaalde mentaliteit om de betrokkenheid te vergroten, bijvoorbeeld het denken aan je groep als een hecht team waarin de bijdrage van elk individu belangrijk is. Sociale strategieën vertrouwen op het inzetten van andere individuen om je eigen betrokkenheid te vergroten, bijvoorbeeld iemand iets beloven of een contract tekenen. Technische strategieën vertrouwen op het inzetten van technologie om betrokkenheid te vergroten, zoals het zetten van een wekker om op tijd op een afspraak te komen, of gaan werken aan je proefschrift op een plek zonder internetverbinding, zodat je niet afgeleid zult worden door nieuws of email.

Aangezien het gebruik van technische artefacten onderdeel kan zijn van technische strategieën, kunnen artefacten dus niet alleen invloed uitoefenen op onze individuele, maar ook op onze collectieve handelingen.

Tenslotte kijk ik in hoofdstuk 8 naar een aan handelingen verwant onderwerp: de verdeling van verantwoordelijkheid voor handelingen met artefacten tussen ingenieur en gebruiker. Een ervaren automobilist die door rood rijdt en daardoor een ongeluk veroorzaakt, is in principe verantwoordelijk voor dat ongeluk. Als deze automobilist echter door rood reed omdat de remmen niet werkten, of het rem systeem in deze auto heel anders werkte dan dat in soortgelijke auto’s, lijkt het echter redelijk om de ingenieur (ook deels) verantwoordelijk te houden. Ik laat zien onder welke voorwaarden verantwoordelijkheid voor artefactgebruik en de gevolgen daarvan overgedragen kan worden van ingenieur naar gebruiker, door het verbinden van de gebruiksplanaanalyse van Houkes en Vermaas, die voorwaarden opstelt voor rationeel ontwerp en gebruik van artefac-
ten, met de theorie van morele verantwoordelijkheid voor handelingen en de gevolgen daarvan van Fischer en Ravizza. Vervolgens toon ik hoe deze voorwaarden praktisch toegepast kunnen worden in een case study, waarbij ik onderzoek doe naar de morele verantwoordelijkheid van de gemeente Abcoude en weggebruikers voor ongelukken die gebeurd zijn met een doseersluis in die gemeente; een complex artefact dat geïnstalleerd was om de verkeersveiligheid binnen de gemeente te verbeteren. Ik betoog hierbij dat gebrekkige communicatie over de werking van de doseersluis van de gemeente naar de weggebruikers het grootste probleem was bij de overdracht van verantwoordelijkheid van gemeente naar weggebruiker.
Curriculum Vitae Auke Pols

Auke Pols (Assen, 1979) studied Cognitive Artificial Intelligence at Utrecht University from 1997 to 2002, for which he wrote a thesis on the phenomenon of insight in creative problem solving. From 2002 to 2004 he studied the Philosophy of Cognitive Artificial Intelligence at the same university, for which he wrote a thesis about the evolution of morality, ‘The Myth of Categorical Imperatives’. From 2007 to 2011 he worked on his PhD project in Eindhoven, as part of which he also visited Reading: this book is the result.

In between his studies and his PhD, Auke has worked on various contractor jobs, including database management at Rabobank International, and has made various contributions to Elf Fantasy Magazine as freelance writer. In his spare time he writes short stories, several of which have been published. Auke is married to Marijke Hobo and they have two sons: Leendert and Thijmen.
Simon Stevin Series in the Philosophy of Technology
Delft University of Technology & Eindhoven University of Technology
Editors: Peter Kroes and Anthonie Meijers

Books and Dissertations
Volume 2: Anke van Gorp, Ethical issues in engineering design, Safety and sustainability, 2005.
Volume 5: Melissa van Amerongen, The Interpretation of artifacts; A critique of Dennett’s design stance, 2008.

Research Documents
Peter Kroes and Anthonie Meijers (eds.), ‘Philosophy of Technical Artifacts’, 2005
Simon Stevin (1548-1620)

‘Wonder en is gheen Wonder’

This series in the philosophy 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.
Though humans are ‘the tool-using animals’, philosophical action theory has rarely investigated the role artefacts play in the action process. This book investigates this role, driven by its research question: What are actions with artefacts, and how do artefacts influence agents and agency?

*Acting with Artefacts* argues against Davidson’s claim that all actions are bodily movements, and uses the psychology of action to show that there can be actions with artefacts. The resulting picture of actions is used to clarify the concept of ‘affordance’ in design. Furthermore, this book argues that artefacts do not just enable or facilitate actions; they can also alter our reasons for action, and more radically, alter us in our capacity as bounded practical reasoners. Finally, this book shows how artefacts can support our commitment to collective actions, and how engineers can transfer responsibility for artefact use to users.

*Wonder en is gheen wonder*